THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INTEGRATED BRAKE CONTROL

DOWNHILL DRIVING STRATEGIES

PETER LINGMAN

Department of Applied Mechanics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2005

Integrated Brake Control
Downhill driving strategies
PETER LINGMAN
ISBN 91-7291-731-8

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 2413
ISSN 0346-718X

Department of Applied Mechanics
Chalmers University of Technology
SE-412 96 Göteborg, Sweden
Telephone +46 (0) 31 772 10 00

Printed by Chalmers Reproservice
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2005

Abstract

In this thesis, downhill driving strategies that co–ordinate the different brake actuators in heavy duty vehicles have been developed. As a starting point, the vehicle features affected by the retardation system are investigated. The main conclusions are that the retardation power demand will increase in the future and that, therefore, optimization of the brake systems will come to play a major role. In particular, strategies for the integration of foundation brakes, auxiliary brakes, and gear box have to be developed. Furthermore, these strategies must take component wear cost into consideration. Additionally, a thorough description of the current situation in terms of driver behaviour and existing systems for driver assistance is given.

Optimal control and nonlinear programming have been used for the generation of open–loop optimal driving strategies. Two different methods have been employed for the generation of implementable, closed–loop, driving strategies. First, a method that utilizes neural networks and genetic algorithms is presented. Second, in order to further enhance the controller transparency, and the possibility for robust implementation, the control problem is divided into two different modes of operation. Linear quadratic control using gain scheduling is then utilized for the controller design and generation of actuator reference values. Comparison with the open–loop optimal strategies is also made. It is shown that transport efficiency (i.e. mean speed) and retardation economy (i.e. component wear cost) can be improved significantly, even compared to what skilled drivers can achieve, by integrating the whole retardation system. It is furthermore shown that there is a trade–off between component wear cost and transport efficiency that must be balanced in order to achieve good brake performance. The main parameters that affect the longitudinal control of the vehicle are the level of vehicle utilization (mass) and road slope. Algorithms for estimation of vehicle mass and road slope are therefore developed and presented. Additionally, a downhill driving strategy is implemented and verified in a real truck.

Keywords: Auxiliary brakes, foundation brakes, downhill driving strategies, transport efficiency, retardation economy, brake heat fading, mass estimation, road slope estimation, cruise controller, component wear, heavy duty trucks.

Preface

The work in this thesis was carried out between 2000 and 2005 at Volvo trucks, department of chassis engineering and Chalmers University of Technology, department of Applied Mechanics in Göteborg, Sweden. The project was financed by Volvo trucks and, during the last year, also by Intelligent Vehicle Safety Systems (IVSS) which both are gratefully acknowledged.

Many people have contributed to this thesis. First and foremost, I thank my excellent supervisor duo consisting of Mattias Wahde (Chalmers) and Bengt Schmidtbauer (Chalmers). We have had many fruitful discussions during my studies and I'm mostly grateful for you giving me an opportunity to see things from an academic point–of–view. I also thank Mattias for the excellent help with debugging my prose. Furthermore, I thank Stefan Edlund (Volvo) for all the early morning meetings at your office, especially in the beginning of this project. Your ability to see things "from above" is really inspiring and I would like to think that some of it has rubbed of on me. I also thank my guidance group for valuable support.

When I started out at Volvo I was introduced to the "real world" by Mats Sabelström (Volvo). We went out driving lumber trucks in Norway and had meetings with brake system suppliers. This was very valuable for me even if it gave me a fare amount of sleepless nights – how could theories of control engineering be applied to improve the co–ordination of brake actuators, was everything already solved? Luckily (at least I thought so in year 2000), the answer is a big NO, and I believe this thesis defines and solves some interesting brake co–ordination problems.

I would also like to thank my present and former colleagues at Volvo and Chalmers for creating an inspiring and good working atmosphere.

During some periods of time, that sometimes seemed endless, the "production" of this thesis has affected my personal life. I would therefore like to give a sincere thanks to all my friends and family for still remembering me.

Sofia, my wife, thank you for your unconditional love.

I finally finished one school. Now it's time to enter a new one.

Peter Lingman
Göteborg, December 2005

Appended papers

- P. Lingman and B. Schmidtbauer. *Road Slope and Vehicle Mass Estimation using Kalman Filtering*. In Proceedings of the 17th International Symposium Dynamics of Vehicles on road and Tracks, IAVSD, Copenhagen, Denmark, 2001.

- P. Lingman and M. Wahde. *Transport and Maintenance Effective Retardation Control Using Neural Networks With Genetic Algorithms*. Vehicle System Dynamics, vol 42, pp 89–107. Taylor and Francis Ltd. 2004.

- P. Lingman and B. Schmidtbauer. *Utilizing Foundation Brakes in a Heavy Duty Vehicle Cruise Controller Application*. Submitted to Vehicle System Dynamics. 2004.

- P. Lingman, A. Duarte, and R. Oliveira. *Open Loop Optimal Downhill Driving Brake Strategies Using Nonlinear Programming*. Submitted to International Journal of Heavy Vehicle Systems. 2005.

Short glossary

4x2 truck *truck with a total of 4 wheels, 2 of which are driven*, page 6

adaptive cruise controller (**ACC**) *controls following distance*, page 14

automated manual transmission (**AMT**) *gearbox with a shifting robot*, page 77

auxiliary brakes (**ABs**) *engine or driveline retarder*, page 5

brake blending (**BB**) *distributes brake force between ABs and FBs*, page 16

brake factor (**BF**) *relates foundation brake pressure to torque*, page 6

brake gain *relates foundation brake pressure to torque*, page 6

chromosomes *built up by genes, define the property of an individual*, page 52

continuously variable transmission (**CVT**) *continuously variable gear ratio*, page 81

controller area network (**CAN**) *bus connecting all computer nodes*, page 1

crossover *random exchange of information between individuals*, page 54

cruise controller (**CC**) *system that obtains a constant speed*, page 26

electronic brake system (**EBS**) *electro–pneumatic brake system*, page 14

electronic pressure modulators (**EPM**) *pressure controling valve* , page 15

engine compression retarders *auxiliary brake integrated with the engine*, page 11

extended kalman filtering (**EKF**) *model–based nonlinear estimation method*, page 61

fading *loss of friction between brake pad and disc*, page 6

fitness function *defines the goal of the optimization*, page 54

foundation brake threshold pressure *pressure at which torque is generated*, page 28

foundation brakes (**FBs**) *wheel brakes or disc brakes*, page 5

gain scheduling *divide-and-conquer approach for controller design*, page 3

genes *fundamental functional units*, page 52

genetic algorithms (**GAs**) *non–gradient based optimization methods*, page 3

glazing *loss of friction between brake pad and disc*, page 6

hydrodynamic retarders *hydrodynamically working auxiliary brakes*, page 11

individuals *solution candidates*, page 52

kalman filtering *model–based estimation*, page 3

linear quadratic control (**LQ**) *model–based controller design method*, page 3

mutation probability *user–defined probability for mutation*, page 54

mutation *random modification of genes*, page 54

neural networks (**NNs**) *nonlinear mapping, grid of computational nodes*, page 3

non–ventilated *foundation brake without cooling channels*, page 5

non-wear brakes *auxiliary brakes*, page 11

nonlinear programming (**NLP**) *class of optimization problems*, page 3

overspeed *vehicle speed higher than driver set speed*, page 26

pareto optimality *used to study trade–off between objectives*, page 41

population *set of solution candidates*, page 52

primary retarder *brake mounted between the engine and the gearbox*, page 11

rapid control prototyping (**RCP**) *transfer of algorithms to real–time target*, page 63

secondary retarder *brake mounted between the gearbox and final gear*, page 11

selection *fitness–proportional selection of individuals*, page 54

sequential quadratic programming (**SQP**) *method to solve a NLP*, page 50

slip control *axle–wise brake force distribution in EBS*, page 14

transcription method *transformation to a finite–dimensional problem*, page 49

velocity–based linearization *linearization that is globally valid*, page 64

Contents

Chapter 1

Introduction

The theme of this thesis is **low–power retardation control in the longitudinal direction**. Low–power retardation is the most common brake manoeuvre performed in everyday truck driving. Adjusting speed to traffic and speed-keeping during downhill cruising are two typical situations where the retardation power is low (below 400 kW), compared to, for example, emergency braking (3000 kW).

Integrated retardation control

In order to increase traffic safety, transport efficiency, and customer satisfaction and to conform to environmental legislation, integration of different subsystems is becoming a key issue in the truck business. Steering systems are being integrated with brake systems in order to improve vehicle handling and safety. Engine and transmission have to work together to decrease fuel consumption and improve performance. The integration of different actuators in a retardation system is also becoming increasingly important, resulting in more complex systems.

With the development of automated gear boxes, continuously controlled auxiliary brakes, continuously controlled foundation brakes, and electronically controlled cooling fans, all communicating on the same **Controller area network** (**CAN**) bus, the possibilities to increase vehicle safety and performance are almost without limit. Transportation effectiveness can be increased without compromising safety, and by distributing the total brake force intelligently, component wear cost can be minimized. By also detecting vehicle behaviour, brake system performance, and the surrounding environment, the driver-vehicle interface can be improved to increase safety and to simplify the driver's daily work. One simple example of this is to provide information to the driver about the status of the retardation system (e.g. foundation brake condition). Another more complicated example is to use this information to guide actively the driver and vehicle to a

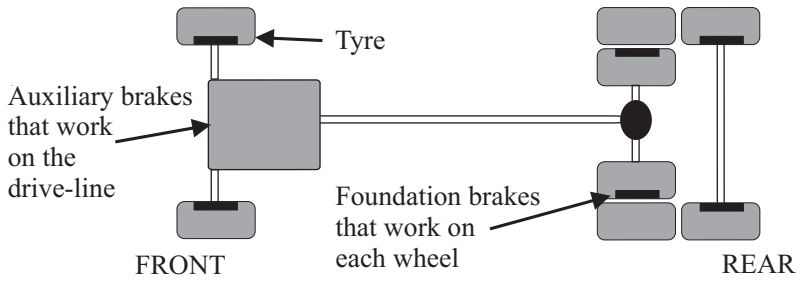


Figure 1.1: Auxiliary and foundation brakes.

more advantageous point of operation. The current approaches to the integration of auxiliary brakes and foundation brakes, as shown in Figure 1.1 and in the next section, which can be found in production vehicles are probably only the beginning of the integration of retardation subsystems to enhance traffic safety, transport efficiency, and customer satisfaction.

Actually, there are few systems on the market that target the downhill cruising situation, and those that can be found are mainly integrated into the infrastructure and are not on board the vehicle. Therefore, the emphasis of this thesis is on control strategies integrating chassis and driveline for continuous retardation situations (low–power retardation). Special attention is given to the problem of maximal usage of the brake system to improve mean speed on downhill slopes and to the problem of distributing a required retardation force between chassis and driveline in order to minimize the pad, disc, and tyre wear costs for the truck owner.

Also, the set of input signals required for retardation system control is discussed and special attention is given to the problem of estimating vehicle mass and road slope that, together with vehicle speed, are the major parameters in the energy balance of moving vehicles, setting the requirements on the retardation systems in the actual driving situation.

Outline of the thesis

In Chapter 2, the main features of a state–of–the–art brake system and the different components and subsystems forming the complete retardation system are described. Then, Chapter 3 proceeds with desired vehicle features (for example performance and economy) that are affected by the retardation system. The complexity of integrating different subsystems and the necessity to do so are put

into focus by pinpointing potential feature conflicts. Next, continuous retardation strategies used by drivers today and the benefits of using also foundation brakes to improve transport effectiveness are discussed in Chapter 4. In Chapter 4, a description of existing systems for driver assistance is given as well. Furthermore, retardation economy is defined and driver strategies are viewed also from this perspective. In Chapter 5, the derivation of open–loop optimal driving strategies using **nonlinear programming** (**NLP**) is presented. This is followed by the description of two different methods focusing on implementable, closed–loop strategies. The first method uses **neural networks** (**NNs**) optimized by **genetic algorithms** (**GAs**) and the second method uses **linear quadratic control** (**LQ**) and **gain scheduling**. Additionally, the selected methods are motivated and discussed in relation to each other. Next, the method of **Kalman filtering** to estimate important vehicle states is discussed. Finally, conclusions are drawn and a summary of appended papers is presented.

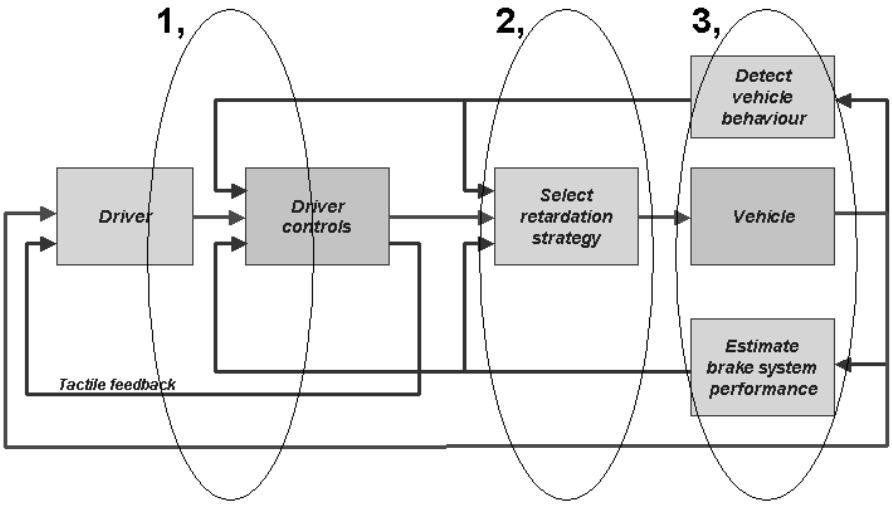


Figure 1.2: Research areas.

Research objective

The objective with the research presented here is *to develop a concept for integrated retardation control, i.e. simultaneous operation of different retardation actuators and gear, in order to increase system performance.*

The concept of an integrated retardation system is illustrated in Figure 1.2, where all feedback loops between vehicle, driver, and controllers are shown. Three areas of research, of interest for this thesis, can also be outlined from Figure 1.2:

1. Driver and vehicle interface satisfying demands on vehicle features,
2. Retardation strategies satisfying demands on vehicle features,
3. Measurement and estimation of vehicle and brake system performance necessary for 1 and 2 above.

Chapter 2

Retardation system

The complete retardation system in a heavy duty vehicle consists of several chassis, transmission, cab, and engine systems. In Figure 2.6, the principal brake system layout for a 6x2 (2 of the 6 wheels are driven) truck is shown, and in Figure 2.7 a more detailed view of the retardation system is illustrated. **Foundation brakes** (**FBs**), and **auxiliary brakes** (**ABs**), are the main systems.

2.1 Foundation brakes, FBs

Two different type of FBs dominate the market, drum brakes and disc brakes. In this thesis, the focus is on the disc brake type, since it is the one that is expected to dominate the truck market in the foreseeable future. In Figure 2.1, a disc brake used on the front axle of a Volvo truck is shown. During braking, the brake cylinder is filled with air creating a pushing force on the piston. This force transfers to the pads and creates a clamp force on the disc, and hence a brake torque is generated on the wheel. The kinetic energy of the vehicle is transformed into heat that leaves the disc by convection and radiation to the surrounding air, and by conduction to the hub, bearing, axle etc.There are also two different types of disc brakes on the market: **ventilated** and **non–ventilated** disc brakes. In Figure 2.2, a ventilated Volvo disc (or rotor) is shown. The disc mounts to the axle via the spline interface and the ventilation is achieved by the air–channels, as shown in Figure 2.2. One of the main benefits of using ventilated discs is that the weight per brake, due to the air channels, is approximately 7 kg lower than for non–ventilated discs. Therefore, the ventilated disc reduces the unsprung mass, which is very desirable from a payload perspective. The force characteristics of the two disc brakes are similar but the thermal properties are not. Due to the lower mass of the non–ventilated disc, it warms up faster than the ventilated one (even if the cooling is improved by the ventilation). Another drawback with the ventilated disc is that

it is more sensitive to long-time brake application when compared to the solid, non–ventilated disc. When the brakes are applied, the friction force between the pad and disc heats up the disc surface. However, the disc surface does not become evenly warm due to uneven pressure distribution in the contact area between pad and disc. Often only around 15–20 percentage of the total pad area is in contact with the disc, see for example Eriksson et al. (2002). Usually, when the brakes are applied, the disc surface becomes warm in a relatively narrow, circular, band (sometimes several bands appear). As the brake application continues, this band will travel radially over the disc surface, creating thermal stresses in the material. Such stresses also arise from the axial temperature gradients that are caused by the brake application. Due to the air channels, the ventilated disc is more sensitive to the thermal stresses than the non–ventilated disc, and in extreme situations cracks can develop on the disc surface. In Figure 2.3 the surface temperature development for a 45 s long brake application with low pressure and medium speed is shown using 6 snapshots distributed over time. Note that each picture shows only a sector of the ventilated disc. At the beginning of the brake phase, the surface temperature is relatively even, ranging from 80 °C to 110 °C . Slowly, the development of two warm bands can be seen. After 30 s the two bands are very narrow with temperatures of around 160 °C and after 45 s the temperature of the disc surface ranges from 100 °C up to 190 °C .

The characteristics of the FB can vary significantly due to high variations in the friction properties between pad and disc. Depending on e.g. disc temperature, pressure, velocity, temperature of the air, and unwanted coatings like road salt and rust, the friction can vary strongly. Friction variations and a method to estimate these in real time are reported in Granqvist and Rundqvist (2005). In Figure 2.4 and 2.5, friction measurements for a Volvo **4x2 truck** with ventilated discs are shown. The friction is in this case characterized by the **brake factor** (**BF**) (or **brake gain**). In both situations the initial vehicle speed is 50 km/h. In Figure 2.4 the brake pressure is 3 bar and the initial pad temperature is 460 °C and in Figure 2.5 the brake pressure is lowered to 2 bar and the pad temperature is lowered to 180 °C . It can clearly be seen that the brake gain varies rather much compared to the nominal value of 4400 Nm/bar/axle. In fact, the maximum deviation reaches 15 %.

Other extreme variations in the friction are created by two phenomena called **fading** and **glazing**. Fading is a loss of friction when the temperature exceeds a certain level. For disc brakes this begins around 600°C. Around 800°C there is only approximately 40% of the friction at 600°C left, and at 900°C the pads burn. Glazing, on the other hand, occurs when the FB is used too little. If the driver never uses the brakes for high power retardation, a low friction layer will develop between pad and disc, resulting in bad brake performance.

It is, in fact, a problem that drivers tend to use the disc brakes too little, re-

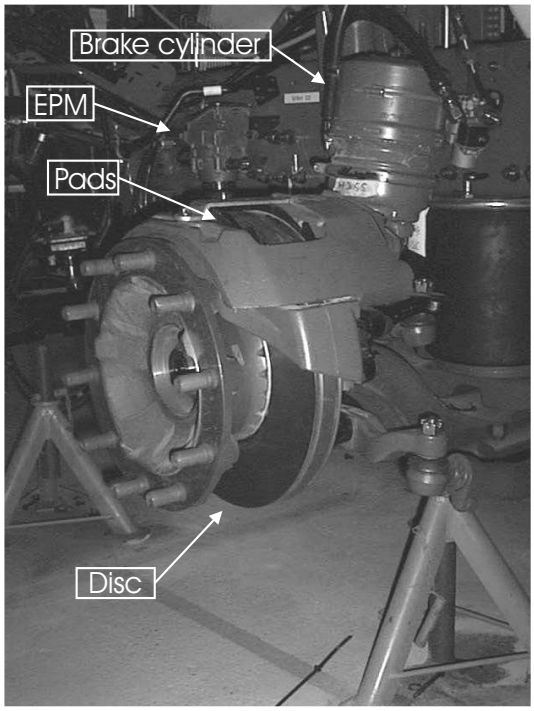


Figure 2.1: Front axle foundation brake. Disc, pads, modulator, and a radially mounted brake cylinder.

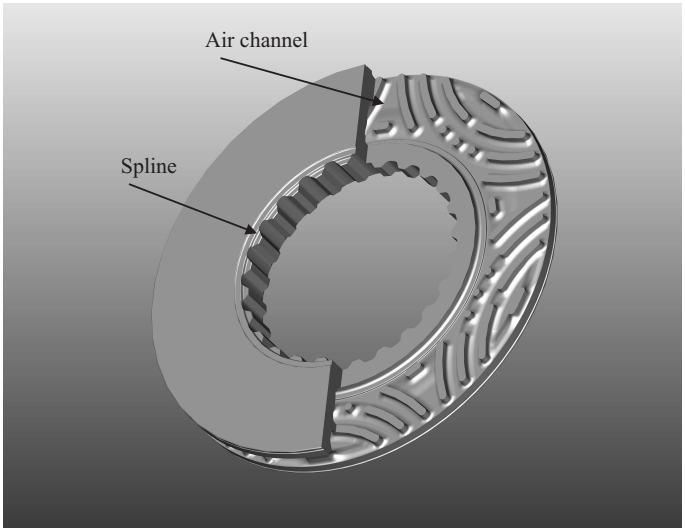
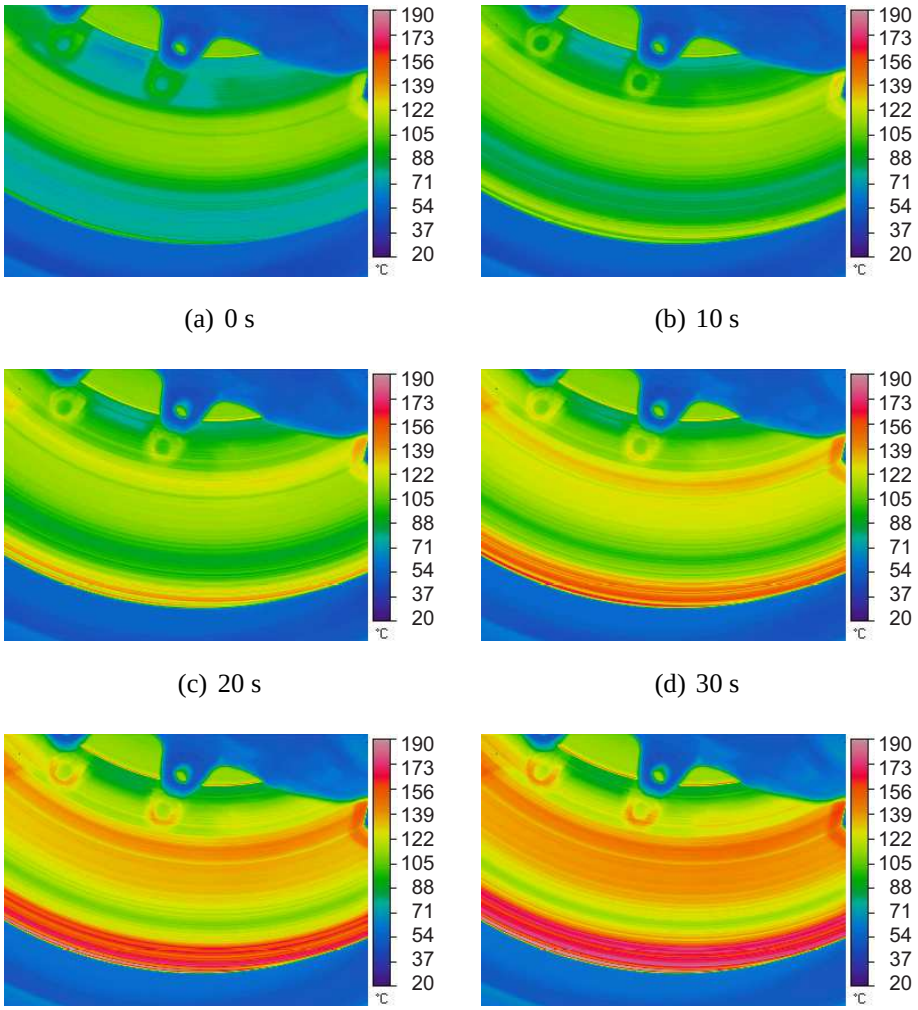


Figure 2.2: Volvo light–weight disc (or rotor) with ventilation. The solid disc does not have any air channels for ventilation.



(a) 0 s (b) 10 s

(c) 20 s (d) 30 s

(e) 40 s (f) 45 s

Figure 2.3: Six snapshots of disc surface temperature during braking with low pressure and medium speed. The brake duration was 45 s.

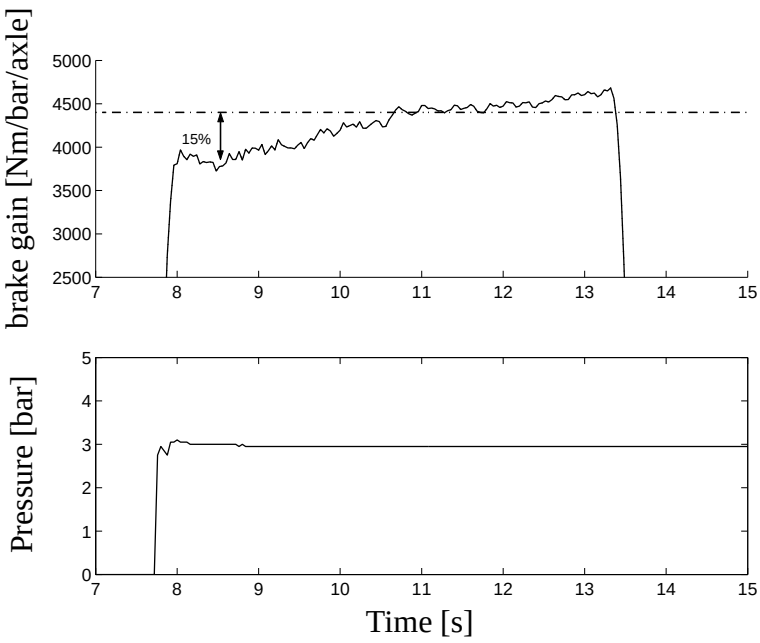


Figure 2.4: Measured brake gain for one axle. The maximum deviation from the nominal brake gain is approximately 15%. Initial velocity = 50 km/h, brake pressure = 3 bar, and initial pad temperature = 460 °C .

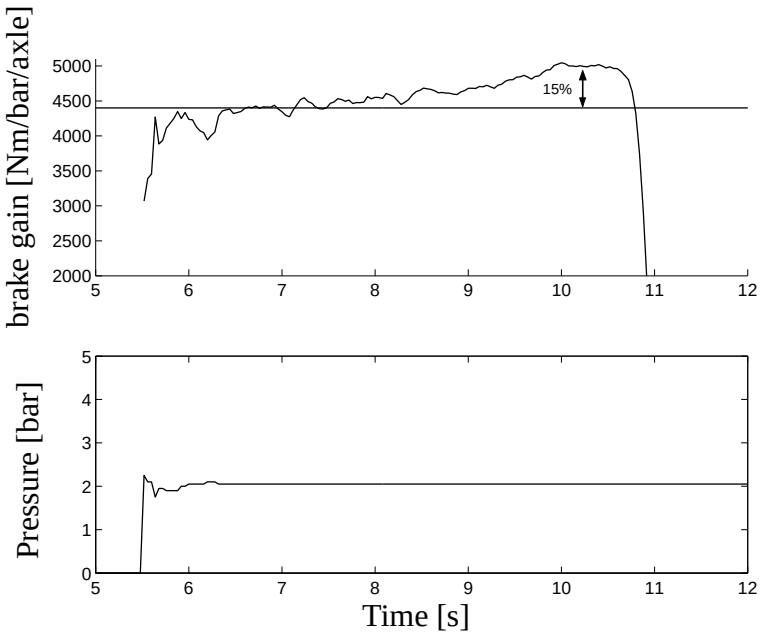


Figure 2.5: Measured brake gain for one axle. The maximum deviation from the nominal brake gain is approximately 15%. Initial velocity = 50 km/h, brake pressure = 2 bar, and initial pad temperature = 180 °C .

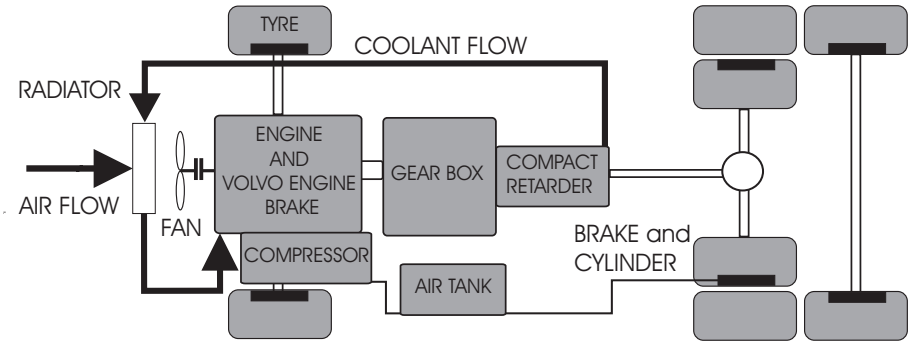


Figure 2.6: Volvo brake system.

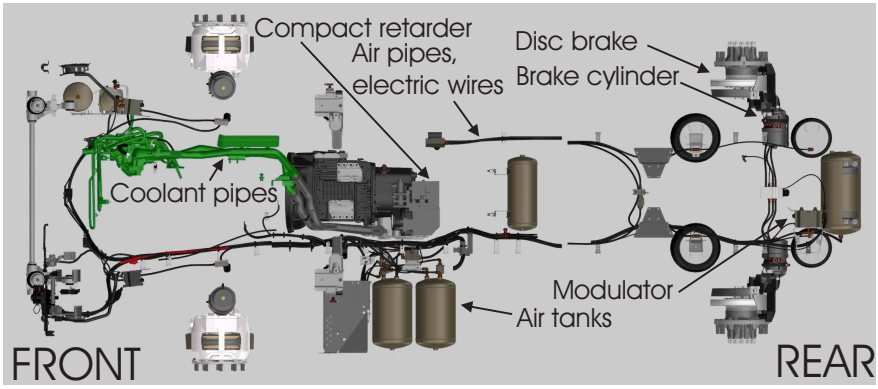


Figure 2.7: Detailed view of the Volvo brake system.

sulting in bad braking performance and low traffic safety. Low usage allows not only undesirable coatings to accumulate on the disc but also rust and dirt to be deposited on the brake mechanism. When it comes to glazing, it is not easy to quantify exactly what ”too little braking” is, since it depends on factors such as the surrounding environment, disc and pad material, and disc temperature when braking. A rule of thumb used is that one or two high power brake manoeuvres should be made each week to clean and condition the discs from unwanted surface coatings.

Not only safety, but also economical reasons, favor a reduction of the usage of FBs. Since pad and disc wear can be a relatively large expense for the truck owner, drivers traditionally utilize non–wear ABs more than FBs.

2.2 Auxiliary brakes, ABs

Several types of ABs (also called **non-wear brakes**) are on the market and they are all characterized by their non–wear properties. **Hydrodynamic retarders**, **engine compression retarders**, **engine exhaust retarders**, and **electromagnetic retarders (eddy-current brakes)** are some examples. In this thesis, hydrodynamic, compression, and exhaust type retarders are used to represent the ABs. The hydrodynamically working retarder is a **Volvo compact retarder (CR)** mounted between the gear box and final gear and, therefore, also called **secondary retarder** (behind the gear box, as shown in Figure 2.8). The principle of a hydrodynamically working retarder is that the friction in the oil between the stator (CR) and the rotor (propulsion shaft) generates a brake torque on the propulsion shaft. By controlling the amount of oil, using pressurized air, different brake torques can be achieved. Since all the brake energy is transformed into heat, the oil has to be cooled using the main engine cooler.

The Volvo Engine Brake (**VEB**) is a combination of a compression retarder and an exhaust retarder and is part of the engine. Since it is mounted in front of the gear box it is also called **primary retarder**, see Figure 2.6. The exhaust retarder is a controllable valve mounted on the exhaust manifold. When engaged, the valve prevents the exhaust gases from flowing out from the engine, thereby increasing the pressure in the exhaust manifold, and therefore generating a negative (braking) torque on the driveline. For the compression brake, two extra notches on each camshaft exhaust lobe are used to modify the intake and exhaust phases of the engine. This makes the normal exhaust stroke a compression stroke, whereby the exhaust gases are compressed. When the piston is almost at its top dead center, the exhaust valves are opened, thus releasing the compressed air, and a net retardation force is generated on the piston. A further description of the VEB can be found in

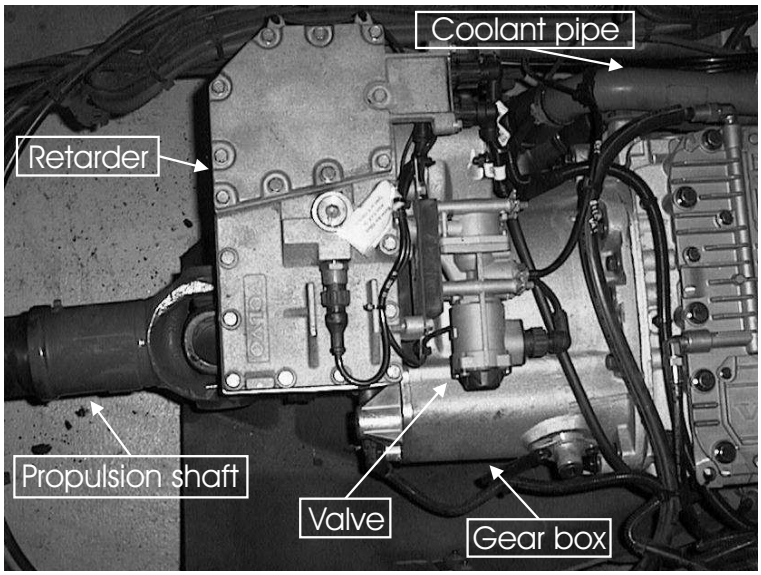


Figure 2.8: Compact retarder mounted on gear box.

Templin (2004). The CR and VEB transfer vehicle kinetic energy into heat which is then transferred to the engine cooling system of the truck. Since this system is primarily designed for cooling the engine in propulsion mode, it usually cannot handle the available brake power from the CR and VEB continuously. A typical cooling system can handle around 300 kW continuously, whereas a CR and VEB can brake around 600 kW together, of which around 400 kW is transferred to the coolant, and the remaining part is transferred to the exhaust gases. This means that the ABs have to be regulated in such a way that the coolant temperature does not exceed a certain limit. Approximately 40% and 100% of the brake power transfers to the coolant from the VEB and CR, respectively. Therefore, usually, the VEB is allowed to be fully engaged and the CR is down-regulated to keep the coolant temperature within limits, meaning that also the CR fades like the FB. In Figure 2.9 a principal fading curve for the CR is shown. The dip in the curve is mainly due to the dynamics of the thermostat. Note that the opening of the thermostat depends only on the temperature of the coolant at the inlet of the thermostat. Figure 2.10 presents an overview of the coolant flow in a heavy duty vehicle. The coolant transports the heat to the radiator. Usually the coolant flow is not directly controllable but rather proportional to the engine speed. Dissipation of heat in the radiator is controlled by a controllable fan and usually when the retarder is engaged, the fan speed is increased in order to provide maximal cooling performance.

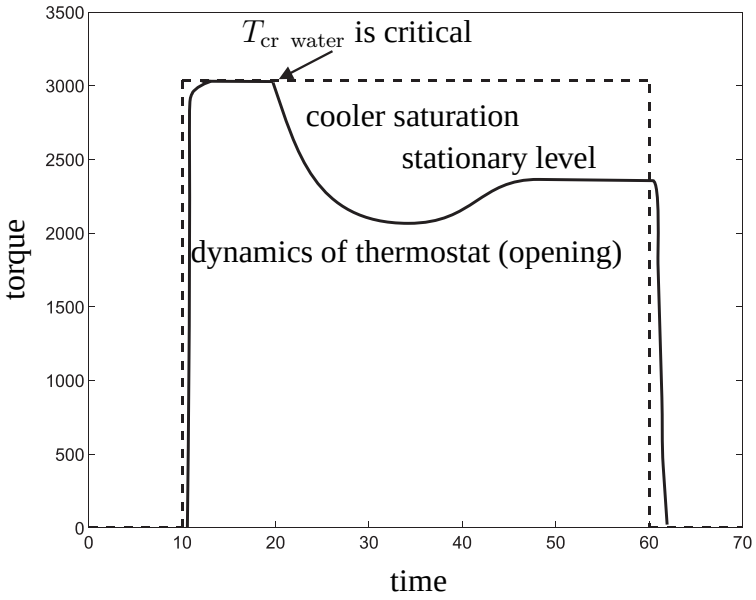


Figure 2.9: Typical fading curve for a retarder. Dashed line: request; Solid line: actual torque. The dip in the curve (also named the hammock) is mainly due to the dynamics of the thermostat.

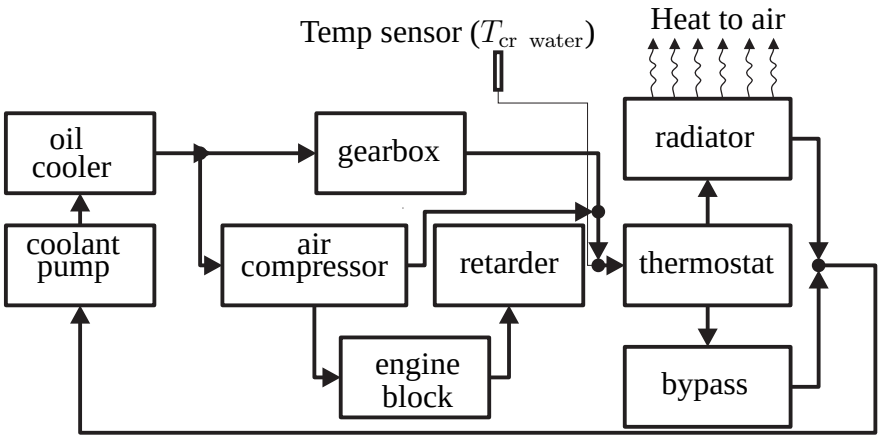


Figure 2.10: Coolant flow in a heavy duty vehicle. The coolant pump is driven by the engine.

2.3 Brake system: State–of–the–art

Modern heavy duty vehicles have an **electronic brake system** (**EBS**). This means that the energy source consists of pressurized air and that the brake control uses electronic transmission, i.e. the driver's demand is electronically transmitted from the controls to the valves that connect the air reservoirs to the actuators. By replacing pneumatic transmission with electronic transmission, system performance has been improved. Reduced stop distance and enhanced lateral vehicle stability are some of the benefits of using electronic transmission. Other benefits are fewer components and easier installation of the brake system on the vehicle chassis. Since there is no standard for the EBS system in the truck industry, several different systems can be found on the market and in the literature, all named EBS, see Palcovics and Fries (2001), Palcovics et al. (2001), and Lindemann et al. (1997).

Several functions have been introduced into the concept of EBS systems over the years, starting from the mid–eighties in the literature and since the mid–nineties in production vehicles. Some of the main functions are given in Table 2.1. The main function of the EBS system, when it comes to longitudinal braking,

Table 2.1: Some EBS functions.

Function	**Description**
Slip control	Controls the brake pressure so that each axle brakes in proportion to the load of the axle.
ABS	Prevents wheel locking to keep the lateral maneuverability of the vehicle
ESP	Uses the brake system to stabilize vehicle laterally (sideways braking)
ROP	Limits the lateral tyre forces to avoid rollover in a curve (the vehicle slides out instead of rolling over)
CFC	Distributes the brake force between truck and trailer in proportion to the mass of truck and trailer
ACC	Keeps a set distance to the vehicle in front (**adaptive cruise controller**)

is the **slip control**. The basic idea with the slip control is to distribute the brake force so that each axle brakes in proportion to the load of the axle using only information about the wheel speed. By controlling the brake force (i.e. the brake cylinder pressure) of each axle so that there is no wheel slip difference between the axles, the correct force distribution is obtained and each axle brakes in proportion to the load on the axle. This is so, since the longitudinal tyre force is the

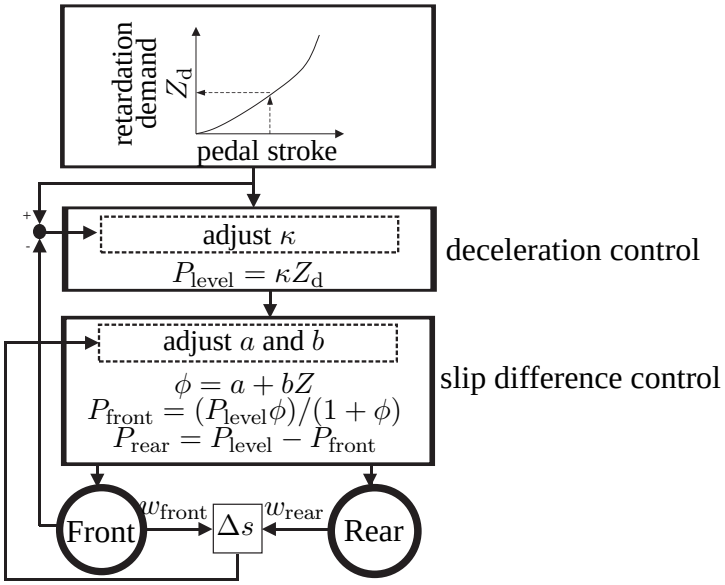


Figure 2.11: Simple foundation brake control loop.

normal force (load) times the utilized friction (slip). In earlier systems the strategy was basically the same but an electronic axle load sensor (ELS) had to be used to distribute the force accurately, and in even earlier systems a load-dependent pneumatic proportion valve was used (not EBS).

A typical FB activation from brake pedal stroke to generated brake force is as follows: When the driver presses the brake pedal a retardation request is sent to the EBS electronic control unit (EBS-ECU). A reference pressure value proportional to the retardation demand is calculated. This reference pressure is then divided (slip control) between each axle and transmitted to the **electronic pressure modulators**(**EPM**) on each axle.

In Figure 2.11, a simplified control schedule for the FBs on a two–axle vehicle is shown. By pushing the brake pedal, a pedal stroke is achieved and a retardation demand (Z_d) can be calculated. A pressure level value (P_{level}) is calculated from this retardation demand and is then distributed to the front and rear wheel brake cylinders (P_{front} and P_{rear}) in such a way that the tyre slip difference (Δs) is zero between front and rear wheels. By controlling the wheel slip difference (or wheel speed difference) between the axles to zero, each axle brakes in a well balanced way.

The total EBS pressure modification logic from driver or system demand to brake cylinders is illustrated in Figure 2.12. First, the total brake demand (de-

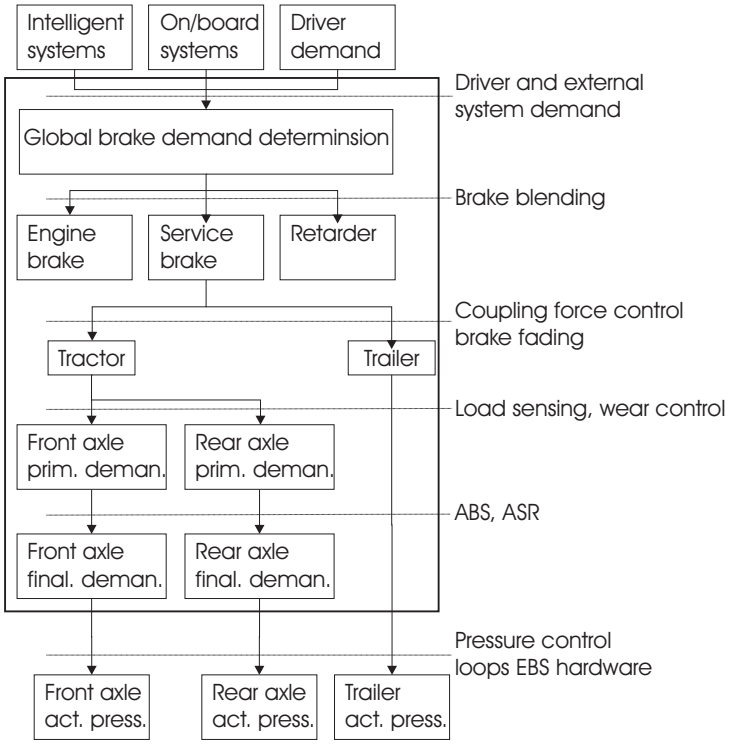


Figure 2.12: Pressure modification loop. Reproduced from Palcovics and Fries (2001).

celeration control) is divided between ABs and FBs. Next, the FB demand is divided between truck and trailer (CFC function) and then the slip control distributes the demands between all the axles, and if necessary the ABS system reduces the pressure to avoid wheel lock. The distribution between ABs and FBs is usually performed manually by the driver via multi-position levers (three to five positions) and the brake pedal. Sometimes also a button is used to engage the ABs (one or two positions). In the Volvo FM truck, introduced in 2002, the driver can also choose to drive the vehicle in a **brake blending** (**BB**) mode, which means that, when pressing the brake pedal, all the retardation force is requested from the ABs, and if the ABs cannot satisfy the request the FBs are used, as shown in Figure 2.13. The dashed line is the requested retardation force and the thin, medium, and thick solid lines are the AB, FB and their sum, respectively. Since the FBs mainly are used in the transient phase of the manoeuvre the, FBs can be said to handle the high frequency part of the demand and the ABs the stationary part (if they manage). FBs are used mainly in the beginning and a perfect blending is achieved when the sum of AB and FB equals the requested force. This strategy coincides

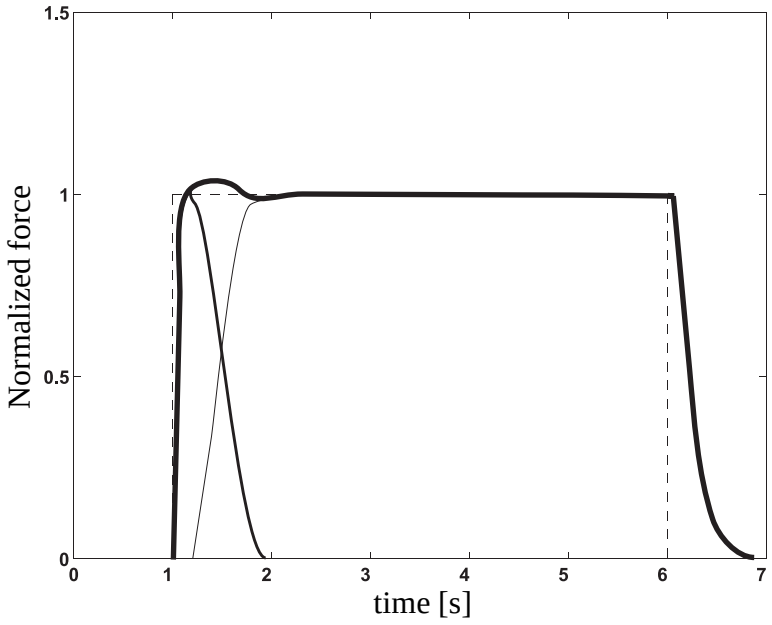


Figure 2.13: Brake blending. Dashed line: request; Thick solid line: foundation brakes and auxiliary brakes; Medium solid line: foundation brakes; Thin solid line: auxiliary brakes.

very well with the one preferred by drivers today in order to minimize wear of brake pads and discs.

Chapter 3

Vehicle features

In this chapter, vehicle features that are influenced by the retardation system are presented and discussed, from a customer point–of–view. Desired vehicle features are translated into technical requirements and used for the development of hardware configurations and control strategies at Volvo Truck Corporation (VTC). Some of these features conflict with each other and have to be balanced in order to achieve a good compromise. The aim with this chapter is to explain the need for investigating and designing control strategies for low–power retardation. Another aim is to give a background to the approach used in Papers 2, 3, and 4, focusing on controlling the retardation system for high transport efficiency (mean speed) and low component (pad, disc, and tyre) wear cost. In the truck industry, low–power retardation is sometimes referred to as **brake blending** or **co–ordinated brake control**, see Chapter 2. It should, however, be noted that either brake blending or co–ordinated brake control are not restricted to low–power brake manoeuvres. For example, brake blending according to Figure 2.13, can be applied in very demanding (high power) brake situations.

3.1 Retardation system features

In Table 3.1, vehicle features are listed and described. The features are divided into main features and subfeatures. Furthermore, feature conflicts are pinpointed in Table 3.2. These features stand as a base for the retardation strategy development and should be covered by it. When looking into the future for retardation systems and ABs in particular, there are several features that can conflict with each other, making the solution a compromise. In Table 3.2, feature conflicts are marked by numbers. Each number is explained in the following numbered list.

Table 3.1: Main and subfeatures.

Feature	**Subfeature**	**Comment**
Quality and uptime	Maintainability	Maintenance of pad, disc and tyre.
	Durability	Wear of components (pad, disc, and tyre).
	Reliability	Software robustness.
Safety	Retardation	Two situations: high and low–power retardation.
	Handling	Braking in curve.
	Driver assistance	Collision avoidance, collision warning, ROP, ESP, ACC, BC
Driver environment	Interior noise	Level and variation of noise. Example: switching actuator and brake squeal
	Dash and instrumentation	Number and type of controls
	Ride comfort (vibrations, jerk etc)	Jerk when switching actuator and gear. FB vibrations. System override situations (active speed control).
Environment	External noise	ABs produce noise and can be a problem in city environment.
	Material emissions	Tyre wear, pad wear (copper), oil wear, and road wear.
Price	Product price	
Fuel economy (FE)	Chassis configuration	Reduction of resistance forces like air drag, rolling resistance, and driveline drag losses.
	Powertrain configuration	Engine and transmission for low fuel consumption.
Brake economy (BE)	Chassis configuration	Number of axles, tyre type and configuration on axles (single or twin), type of brakes.
	Powertrain configuration	Type of ABs, gearbox, and final gear.
Transport effectiveness	Load capacity	Low chassis height increases load volume. Increased vehicle gross weight increases transport efficiency.
	Driving cycle	Suitable brake configurations and strategy depend on operating environment and vehicle utilization.
	Vehicle agility	The retardation strategy depends on traction conditions. ABs are not used in low friction conditions.
	Driveability	Good acceleration performance. Good retardation in high power brake manoeuvre. High mean speed on down–hill slopes.

Table 3.2: Feature conflicts. Each conflict is marked by a number and is explained in the numbered list.

Subfeature	Quality	Safety	Driver environment	Environment	Fuel economy	Brake economy	Transport effectiveness
Quality		1		2	3		4
Safety			5		6	7	8
Driver environment						9	10
Environment					11	12	13
Fuel economy						3	
Brake economy							14
Transport effectiveness							

1. Quality vs. Safety

   (a) As mentioned earlier, glazing is a problem and may deteriorate the brake performance. One way to improve the condition, is to clean the contact surface between pad and disc by transferring more brake load to the FBs. Another solution is to change the design of the brake pads to increase the wear during braking and thereby improve the contact surface cleaning. Either way, the FB wear increases.

   (b) Good continuous low–power brake performance may require high AB usage, i.e. high drive–axle tyre wear.

   (c) Active driver assistance safety systems, like the distance controlling ACC system, may require low or high FB usage (high wear or low condition). for example, ACC systems enable the driver, in a comfortable way, to obtain a very short following distance to the vehicle in front. In order to control this small distance, FBs have to be utilized more frequently in situations where the speed must be reduced since the maximum allowed following distance overshoot must be lower, the closer the actual set distance is.

2. Quality vs. Environment

   (a) ABs produce noise (cf. restrictive usage of Jake[1] brake when driving in city environments in the USA). This requires more use of FB and therefore increased material emissions (copper and lead).

3. Quality vs. fuel economy

   (a) From a fuel economy point–of–view a very desirable feature is to reduce resistance forces like rolling, aerodynamical, and friction losses (also called natural brakes). This reduction of natural brakes transfers more load to the brake system, leading to increased component wear.

   (b) On some driving cycles and load conditions, smaller engine volume also reduces fuel consumption. Small engine volume reduces VEB power and therefore FBs are used more, yielding more wear and lower durability.

4. Quality vs. transport efficiency

   (a) Very desirable features in the road transport industry are increased payload, increased volume and increased vehicle mean speed, all transferring more load to the brake system. There is a trade–off between increased payload and mean speed (income) and wear cost (expense).

5. Safety vs. driver environment

   (a) In an automated system, noise and vehicle motion may be interpreted differently than in a passive system where the driver is more active. Improving the low–power retardation performance requires automatic control of the complete retardation system, i.e. active vehicle speed control. It may be difficult to achieve acceptance for this by the driver.

6. Safety vs. fuel economy

   (a) See first point 3(a). A compact retarder (CR) may be necessary in safety-increasing driver assistant systems like vehicle following (ACC). The CR reduces the driveline efficiency (a power loss of approximately 1–2 kW) which increases fuel consumption.

7. Safety vs. brake economy

[1] The Jake brake is an engine brake manufactured by the Jacobs Company.

   (a) See first point 1(c). Active, safety-improving, driver assistant systems have to consider not only brake performance and safety but also component wear and cost.

8. Safety vs. transport efficiency

   (a) In order to increase vehicle load volume the chassis must be lower, resulting in less space for FBs. Together with increased vehicle mass this could reduce the vehicle brake performance.

9. Driver environment vs. brake economy

   (a) Distributing the retardation force between ABs and FBs to obtain good brake economy (i.e. to minimize component wear) may result in a variable speed solution (see Paper 2 and 4). Driver acceptance may be hard to to achieve.

10. Environment vs. environment

    (a) Using ABs less to reduce noise requires more FB usage, yielding more material emissions.

11. Environment vs. fuel economy

    (a) On some driving cycles and load conditions, smaller engine volume reduces fuel consumption. Small engine volume reduces also the VEB power and therefore FBs are used more, yielding more material emissions.

12. Environment vs. brake economy

    (a) High utilization of FBs increases material emissions but may constitute good brake economy.

13. Environment vs. transport efficiency

    (a) Increased vehicle load requires more brake power and therefore leads to more wear and emissions.

14. Brake economy vs. transport efficiency

    (a) Increased vehicle mass requires more brake power and therefore more expensive components and increased wear.

Clearly, increased vehicle weight, increased load volume, increased mean speed, reduction of pad and disc wear, safety, and copper and lead emissions from the FBs promote the development of more powerful ABs. On the other hand, expensive, heavy and noise-emitting ABs, tyre wear, and also perhaps increased drive line drag torque (CR), promote less AB usage for the future.

In the near future, vehicle loads and vehicle masses are unlikely to decrease. In fact, it is not unrealistic to think that vehicle gross weights will rise in order to increase transport effectiveness, see Lindkvist (1983) and Aurell (2000) for discussions of this issue. Also the vehicle weight and allowed axle loads have historically always increased. For example, during the last ten years allowed vehicle gross weight has increased from 52 tonnes to 60 tonnes in Sweden. Increasing the loadable volume and mean speed in order to improve transport effectiveness and customer satisfaction are other important issues. If this is the scenario, clearly stronger ABs and systems that increase the retardation capacity are needed.

To summarize:

1. Increased vehicle mass $\Rightarrow$ increased demand on the retardation system.
2. Increased vehicle mass $\Rightarrow$ more FBs $\Rightarrow$ FB have to be used also for low–power retardation.
3. Increased load volume $\Rightarrow$ less space for FBs $\Rightarrow$ reduced potential for using FBs also for low–power retardation.
4. Reduced natural brakes $\Rightarrow$ increased demand on the retardation system.
5. Energy regeneration $\Rightarrow$ more ABs $\Rightarrow$ more tyre wear if ABs are mounted on driveline.
6. To increase the AB performance, ABs can be mounted on the trailer $\Rightarrow$ integration of several AB systems.
7. A balance of wear cost between tyre, pad, and disc must be made.
8. System too complex to be handled by the driver $\Rightarrow$ driver assistance is required.

To manage all this, intelligent control systems that control the total retardation system including disc brakes, VEB, CR, engine cooler, and gear box have to be designed. The focus should be on utilizing the whole retardation system to increase retardation performance (continuous braking), safety, and retardation economy (component cost and wear), which all are important vehicle features for the customer.

Chapter 4

Low–power retardation: Downhill driving

It is quite fascinating to watch a skilled truck driver in action. First of all, the driver has to handle the steering wheel to keep a 25 m long vehicle on a usually rather narrow road. Then the driver has to switch to the right gear using the shift stick and clutch pedal. If a reduction of speed is required the driver has to press the brake pedal to engage FBs and push a button or stalk to engage ABs. This means that when entering a downhill slope the driver has to perform all these actions simultaneously, i.e. release the accelerator pedal, press the brake pedal, engage ABs, use clutch and shift stick to shift, and keep the vehicle on the road with the steering wheel. An even more difficult task for the driver is to decide on a proper speed depending on current carriage weight, road slope, road friction, and traffic in order to manage the upcoming descent.

4.1 Manual downhill driving

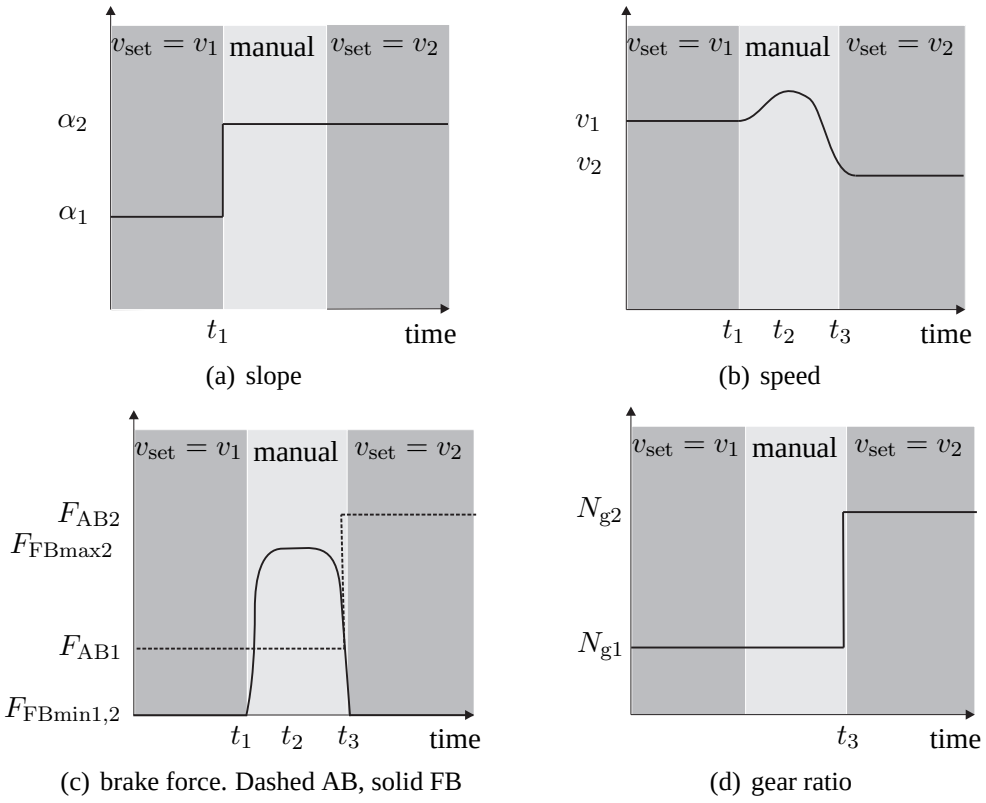
A problem with the retardation systems in use today is that the driver is given almost no feedback regarding vehicle mass, brake disc temperature, and other important vehicle and environment states. Even if the driver knows the loading condition of the vehicle, making use of that information is not an easy task in all situations. Accidents have occurred where a driver has chosen an excessive speed combined with a bad distribution of retardation force between ABs and FBs, resulting in overheating of both the disc brakes (fading) and the ABs (cooling system saturation). Examples of a fading accidents are given in Kress and Kress (2001) and at www.oregon.gov. Other more cautious drivers choose a low set speed yielding a low utilization of the capacity of the brake system and perhaps also low mean speed depending on the driving cycle. A third type of driver (most

common!) is the very experienced and skilled driver with good knowledge of both vehicle and environment. By observing engine speed, coolant temperature etc., such a driver uses the AB system to keep the vehicle speed constant but never uses the FB for long periods of time in order to avoid fading and minimize FB wear. A typical "normal (skilled) driving" sequence is shown in Figure 4.1. The driver has engaged the **cruise controller** (**CC**) to track a set speed $v_{\text{set}} = v_1$. At time t_1 the downhill slope increases from α_1 to α_2. Since the available AB torque is insufficient for the CC to track v_1, the vehicle will **overspeed**. Therefore, it is necessary for the driver to interact with the FB shortly after time t_1. The CC is then turned off and the vehicle is manually controlled by the driver until time t_3. At time t_3 the vehicle speed is v_2 and the driver can shift down to increase N_g (by selecting a lower gear position) and engage the CC with the new lower set speed v_2. This new set speed, v_2, is not trivial to find and in practice the driver has to carry out an iterative procedure of lowering speed and shifting down to find a suitable v_2. In Figure 4.2, the maximum stationary speed (= "suitable set speed") a driver can achieve on different slopes is shown. Two different combinations are shown: VEB only (thin line) and VEB+CR (thick line). The vehicle weight was 60 tonnes. The strategy (as used by a skilled driver) is to utilize the highest possible engine speed to achieve high retardation power and the best possible cooling (high fan speed and coolant pump speed). Using a discrete gearbox causes the steps seen in Figure 4.2. Maximal cooling capacity was around 280 kW. Figure 4.2 is the same as Figure 4 in Paper 2 except for the resolution and the lower cooling capacity, which explains the steps of the curve and the lower speed values compared to Paper 2. The improvement obtained by using a CR together with a VEB decreases with increasing slope. This is due to the fact that the CR is a secondary retarder that has low performance when the propulsion shaft speed is low, i.e. for low vehicle speed.

Even if most drivers are very skilled it is still a fact that some are not. For example, in America there is large shortage of truck drivers (an estimated 20,000 drivers in 2005). It is not unusual for carriers to have 100% annual driver turnovers, resulting in situations with less qualified drivers, see e.g. Min and Lambert (2002). During 2005 logging of driver braking behaviour was started at Volvo. The data collection is far from finished and the number of vehicles that are logged is still only between 5–10 vehicles, distributed around Europe. Preliminary investigations show, however, that some drivers utilize also FBs for long durations, i.e. during downhill driving. A brake phase is defined as

$$\begin{aligned} &\text{if} \quad P_{\text{front}} > \ 0.3 \text{ bar} \ \text{ and } \ \text{v} > 5\text{km/h} \quad \text{Start brake phase,} \\ &\text{if} \quad P_{\text{front}} \leq 0.3 \text{ bar} \ \text{ or } \ \text{v} \leq 5\text{km/h} \quad \text{End brake phase,} \end{aligned} \tag{4.1}$$

where P_{front} is the front brake cylinder pressure and v is the vehicle speed. The



(a) slope (b) speed

(c) brake force. Dashed AB, solid FB (d) gear ratio

Figure 4.1: Typical manual (skilled driver) downhill driving sequence. Foundation brakes are engaged for a short time to reduce vehicle speed to enable a gear down shift. The new vehicle speed is the stationary speed that can be held using only auxiliary brakes.

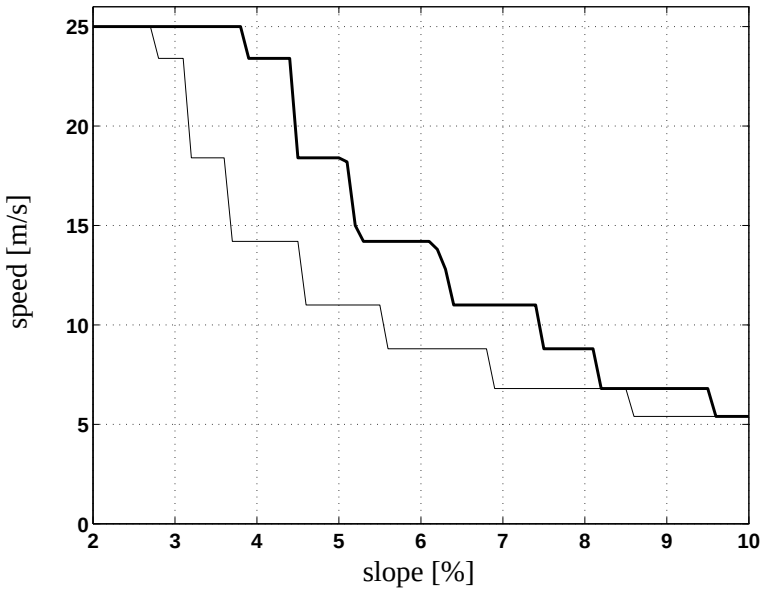


Figure 4.2: Maximum stationary speed as a function of road slope using VEB (thin line) or VEB and CR (thick line).

brake phase starts when the driver engages the brake pedal and the brake cylinder pressure exceeds the **foundation brake threshold pressure**. Additionally, the vehicle speed must be above a predefined limit, in order to avoid registering stand-still brake applications. When the brake cylinder pressure falls below the threshold pressure or when the vehicle speed falls below the speed limit, the brake phase ends. In Figure 4.3, histograms of registered brake phases are shown for a vehicle registered during 9 days of driving. The number of brake phases is shown as functions of initial speed (21 bins, bin width=10 km/h, starting at 5 km/h), maximal pressure during the brake phase (44 bins, bin width=0.2 bar, starting at 0.3 bar), and length of brake phase (50 bins, bin width=1 s, starting at 0 s, the last bin has unlimited bin width). The two lower graphs are zoomed in from the two upper graphs, in order to show clearly the area of interest for low–power braking. It can be seen that brake applications longer than 15 s occurred 72 times (or 8/day). Three times the brake duration exceeded 50s. During these three brake manoeuvres, the maximum pressure was relatively low and the initial speeds were relatively high, indicating speed adjustments on downhill slopes. Note that the driver did not know the temperature of the FBs and therefore did not know if fading was imminent.

Other extreme conditions are described at www.oregon.gov where large downhill slopes in Colorado and Oregon cause accidents due to brake overheating. For example, on Interstate 70 in Colorado (see Figure 4.5) the 5 to 7% downhill

Table 4.1: Recommended speed based on vehicle mass in DTSWS, www.oregon.gov/ODOT/MCT/DOWNHILL.shtml. With permission from: Oregon Department of Transportation Traffic Engineering & Operations, ITS Unit.

Weight [tonnes]	Recommended speed [km/h]
18-22	56
22-25	40
25-36	25
$>$36	16

slope of approximately 16 km registered in average, between 1995 to 1999, nearly two trucks per month that had to use runaway ramps due to lack of brake power caused by overheated brakes, indicating that the drivers had chosen a an excessive downhill cruising speed. To improve the safety, the Oregon department of transportation has installed Downhill Truck Speed Warning Systems (**DTSWS**). The DTSWS produces a custom made message with recommended downhill driving speed that is presented to the driver well ahead of the downhill slope, as shown in Figure 4.4. In the DTSWS the recommended speed is mainly based on the mass of the vehicle, which is measured using a dynamic axle weighing system. In addition, the number of axles is sent (wirelessly) to the off–board DTSWS system. In www.oregon.gov/ODOT/MCT/DOWNHILL.shtml it was concluded that vehicles with 5 axles or more should have a recommended speed calculated according to Table 4.1. It was also concluded that the average speed was decreased with the DTSWS but no conclusion was made on whether or not a reduction in accidents had occurred after the DTSWS was implemented. The DTSWS is a rather extensive system integrated into both the infrastructure and vehicle, making it unrealistic to install on downhill slopes that are not causing many accidents. These downhill slopes may, however, still cause drivers to either under– or overspeed, resulting in low mean speeds or FB damage. Another drawback with the DTSWS system presented at www.oregon.gov is that vehicle internal states such as foundation brake type, auxiliary brake type, gear ratios, wheel radius, brake temperatures etc are not utilized to further improve the calculation of recommended speed. It is also a fact that the road grade varies along the downhill road section. Therefore, ideally, the speed should vary in order to increase mean speed and save components. Note that according to Table 4.1 a vehicle weighing $>$36 tonnes is recommended to drive at 16 km/h, i.e. the 16 km downhill slope will take 1 hour to overcome!

In addition to safety reasons, drivers prefer to use AB over FBs also to minimize the service cost of the FB. It is, however, not obvious that this strategy is

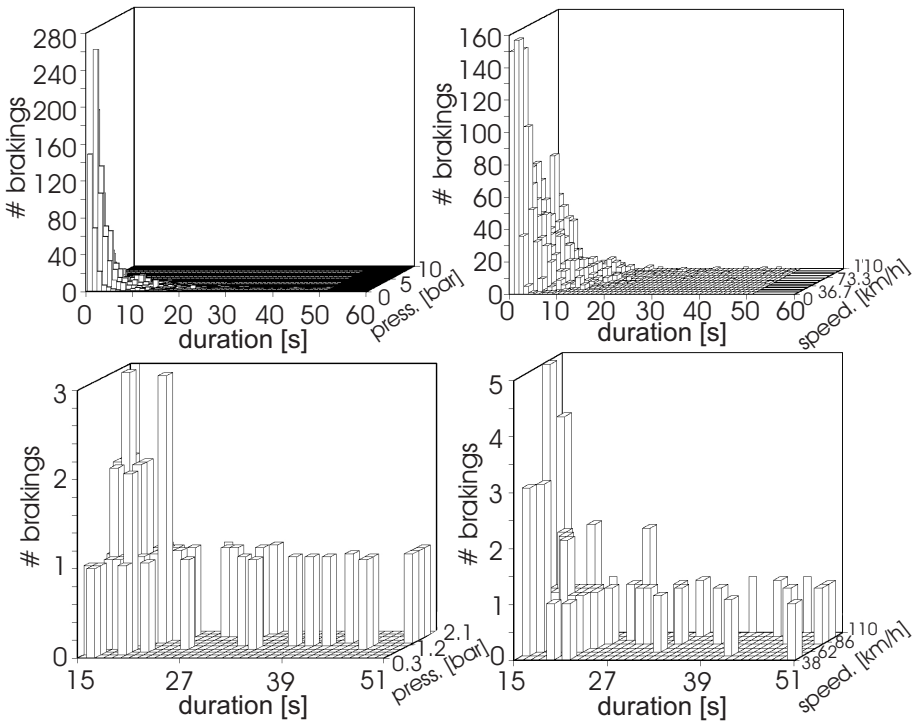


Figure 4.3: FB brake duration for one truck registered during 9 days in March 2005. The total driving distance during 9 days was 3100 km. Clearly these drivers manually used the FBs very much in downhill cruising. Note that the drivers did not know the temperature of the FBs.

Figure 4.4: Recommended downhill cruising speed that is custom made to the driver. Emigrant hill, Interstate 84, Oregon, USA. With permission from: Oregon Department of Transportation Traffic Engineering & Operations, ITS Unit.

Figure 4.5: Interstate 70 in Colorado USA. 5–7% 10 miles long downhill slope with downhill truck speed warning system. With permission from: Oregon Department of Transportation Traffic Engineering & Operations, ITS Unit.

optimal, since, when performing a retardation the force is transferred to the road via the tyres, causing wear. When using ABs, the total retardation force is transferred to the road via the drive axle tyres only, whereas when using the FB it can be distributed to all axles. The tyre, pad, and disc wear cost using ABs and FBs is discussed in Section 4.4. In Table 4.2 pad, disc, and tyre wear cost is calculated for three different vehicle configurations and strategies to overcome a 3% and a 5% downhill slope of 3 km. The three different vehicle configurations are: VEB only, VEB+CR, and the complete brake system VEB+CR+FBs. The two first combinations only utilize ABs and can therefore be compared to the strategies used by normal drivers whereas the last combination utilizes the complete brake system in an optimal way according to Paper 4. On the 3% slope, all combinations can maintain 23 m/s and using only ABs is the best strategy in order to minimize the wear cost. If the slope is increased to 5% the VEB configuration can only maintain 11 m/s, resulting in a wear cost of 1.267 Euro. By using also the CR, the speed can be increased to 18 m/s, and due to increased resistance forces, the cost is reduced when compared to using only VEB. Using the complete brake system in an optimal way significantly reduces the wear cost and enables a higher driving speed, as shown in Table 4.2. This is a clear indication that, on steep downhill slopes, it is advantageous to distribute the required brake force to all axles of the vehicle, rather than concentrating it to the drive–axle.

4.2 Including FBs for continuous retardation

As mentioned earlier, drivers try to maximize the usage of ABs and never use FBs for continuous retardation, in order to avoid fading and save FBs. The main reason for this conservative approach is that the driver has no information of disc temper-

slope (%)	speed (m/s)	VEB	VEB+CR	Optimal Blending
3	23	0.078	0.078	0.078
5	11	1.267	—	0.390
	18	—	1.086	0.384
	23	—	—	0.363

Table 4.2: Results for wear cost in a 3 km constant slope road. All cost values are presented in Euro. In order to minimize the wear cost on a 3% slope it is optimal to use only ABs. If the slope is increased to 5%, optimal blending using the complete brake system is cost efficient. Optimal blending is calculated accoring to Paper 4.

ature, i.e. there is no warning of fading. Some drivers have noticed unusually high drive tyre wear when utilizing ABs, but in general drivers are more concerned with FB wear. In Figure 4.6 the maximum stationary speeds for different slopes are shown for four different hardware configurations VEB only, VEB+CR, VEB+FB, and VEB+CR+FB. The maximum FB temperature was set to 500°C in order to maintain a realistic FB safety margin for emergency situations. Combining not only VEB, CR, and gear but including also FBs improves the retardation performance of the vehicle and it can therefore cruise downhill with higher speed than when only using ABs. Comparing, for example, the combination {VEB, CR} (thick solid line) with {VEB, CR, FB} (thick dashed line) in Figure 4.6, a performance difference is evident, especially for slopes over 6% where the vehicle speed is low and the CR displays mediocre performance. For road slopes between 7% and 10%, the smallest speed gain is around 15% and the largest is around 60%. By also using the FBs, the downshift to increase VEB torque on the driveline can be performed at higher slopes compared to a situation wherein only VEB and CR are used. FBs can be used to increase performance on driving cycles that require much retardation power. On some driving cycles, the VEB alone does not allow high mean speed, and neither can the installation of a CR be motivated from a cost or weight point–of–view. For such situations, controlled FB usage is a good alternative.

As discussed earlier, FBs have to be used with restrictions on temperature, in order to avoid head fading. This is, in fact, the main limiting factor for usage of non–ventilated FBs (solid disc). However, if the vehicle is equipped with ventilated discs then, in addition to fading, also the temperature distribution of the disc must be considered in order to avoid thermal fatigue of the disc. For example, as can be seen in Figure 2.3, the surface temperature is not evenly distributed, causing thermal stresses in the disc material. Therefore, when using ventilated

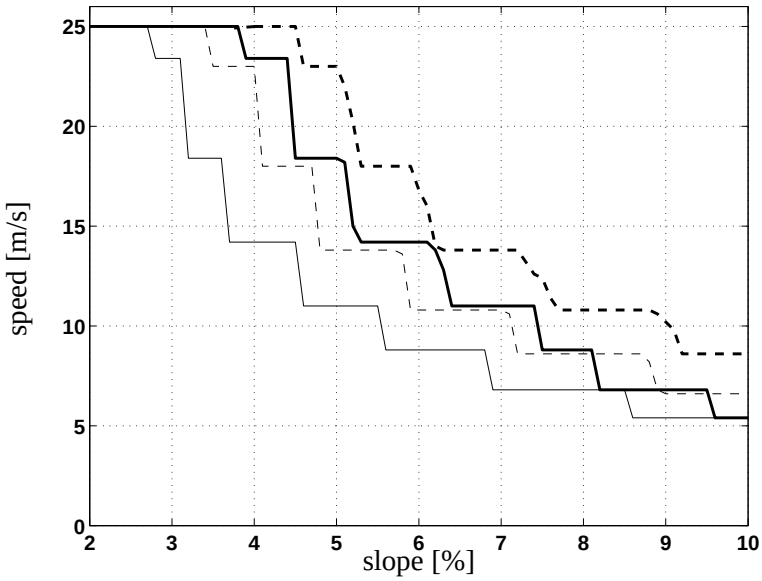


Figure 4.6: Maximum stationary speed as function of road slope using VEB (thin solid line), VEB and CR (thick solid line), VEB and FB (thin dashed line) and finally VEB,CR and FB (thick dashed line).

FBs, the brake duration should be constrained, not only to avoid fading, but also to avoid fatigue damage. Furthermore, it is, from a fatigue point–of–view, advantageous to avoid brake applications with long durations and low pressures. If the brake application pressure is low, the actual pad and disc contact area is usually small, creating an uneven disc surface temperature. More specifically, brake applications, when using non–ventilated FBs, should have an upper limit in brake duration and a lower limit for application pressure. The fatigue issue is not considered in the appended Papers. However, it is possible to include this constraint into both the open–loop and closed–loop strategies derived in Papers 2, 3, and 4. For the open–loop case, an additional constraint in brake duration and a lower boundary on application pressure would have to be included at every time step. This would result in an increased dimension of the Jacobian but it would not introduce any difficulties in terms of discontinuities or other non–linearities. For the closed–loop case in Paper 3, one possibility to take the fatigue phenomena into consideration in the temperature mode controller (TM) and velocity mode controller (VM) is to apply a relay function with hysteresis on the integrated control signal. This relay function would achieve a lower boundary on the FB force, hence creating an intermittent usage of FBs, i.e. the duration of the brake application would automatically by restricted. In Figure 4.7 the relay with hysteresis (fatigue protection) has been applied to the TM controller. The continuous con-

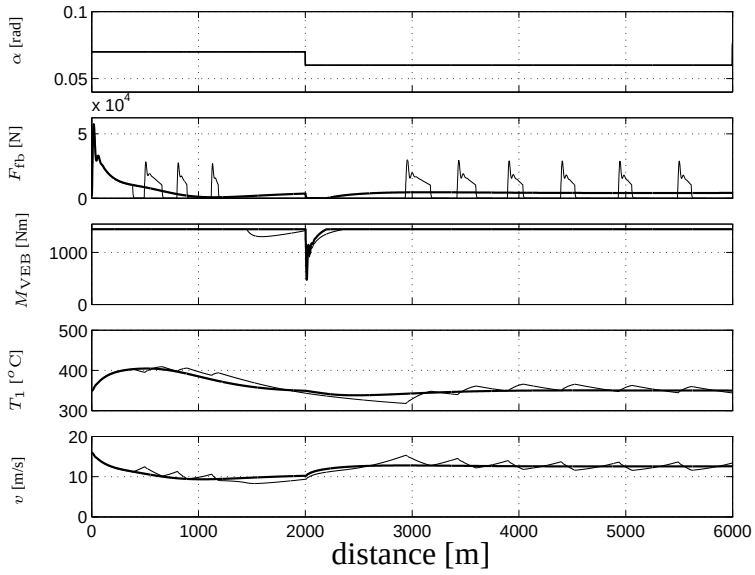


Figure 4.7: Step responses for temperature mode with (thin line) and without (thick line) fatigue protection. At 0 m the slope is changed to 6%. When using the fatigue protection a limit cycle appears with a brake duration of 10 s and a resting brake of 20 s.

trol is shown with thick lines and the fatigue protection is shown with thin lines. As can be seen, the fatigue protection results in a periodic behaviour of the FB application in which the brake is applied for 10 s, released for 20 s, applied for 10 s etc.. Depending on how long brake applications that can be allowed for the ventilated disc, the hysteresis width and position can be tuned in such a way that an adequate protection is achieved.

4.2.1 Automated downhill driving

In a modern truck there are almost no systems that aid the downhill cruising situation for the driver and ABs, FBs, and gear are usually manually handled. There are, however, some systems that integrate ABs and gear and even some systems that integrate ABs, FBs, and gear. Downhill cruise or brake cruise is a system on Volvo trucks that enables the driver to choose a set speed that, on downhill slopes, engages the ABs to obtain the set speed. If the truck is equipped with an automatic gearbox of some type, then also the gear is controlled so that the AB are utilized as much as possible, i.e. a high engine speed is used. However, this system does not control the FBs at any time, even if the AB are insufficient to obtain the set speed. Instead driver intervention is required if the vehicle overspeeds. Downhill cruise can be compared to an ordinary cruise controller that utilizes ABs.

Another system that integrates the brake system is adaptive cruise control (**ACC**). ACC systems, that utilize a vehicle mounted radar, enable the driver to choose a set distance (actually a time gap) to the vehicle travelling in front, instead of a set speed. In the first generation of ACC systems only ABs were used to control the set distance in case of brake force was needed. If the AB brake force was insufficient, usually the driver received a collision warning after which the driver engaged FBs to reduce vehicle speed. The second and current generation of ACC systems utilizes also the FBs to improve the performance of the system. Usually the maximum allowed FB force is limited to a corresponding retardation of around $3\,\mathrm{m/s^2}$. In addition to this limitation, FB usage must also be restricted to ensure that FB fading and FB fatigue does not occur due to ACC over–usage. In a Volvo truck this is achieved by using one fading protection algorithm based on temperature information and one fatigue protection algorithm based on the duration of the brake application. When either of these two protection algorithms reaches predefined limits, the ACC system shuts down. In fact, if the driver set distance is small, the system, before shutting down, increases the FB force to increase the distance to the vehicle in front before handing over to the driver. This procedure is executed in order to minimize the risk of colliding with the vehicle in front when the ACC is turned off. The FB fading protection is based on an estimated (i.e. not measured) FB temperature from the EBS system. If the FB temperature reaches a predefined value (T_{lim1}, well below fading) the system increases the distance to the vehicle in front and ACC is turned off. ACC can be activated again only after the FB temperature reaches T_{lim2} ($T_{\mathrm{lim2}} < T_{\mathrm{lim1}}$, i.e. hysteresis is used). The fatigue protection algorithm is designed as follows

$$\begin{aligned} \text{if } \mathrm{P_x} > \mathrm{P_L} \text{ and } \mathrm{v} > \mathrm{v_{limit}} \text{ then } \mathrm{L_{fp}(t)} &= \int \mathrm{d}t + \mathrm{L_{fp}}(t_0), \\ \text{if } \mathrm{P_x} \leq \mathrm{P_L} \text{ or } \mathrm{v} \leq \mathrm{v_{limit}} \text{ then } \mathrm{L_{fp}(t)} &= a \int \mathrm{d}t + \mathrm{L_{fp}}(t_0), \\ \text{if } \mathrm{L_{fp}}(t) &> b, \text{ brake and turn off ACC.} \end{aligned} \quad (4.2)$$

Thus, as shown in Equation 4.2, if the vehicle speed exceeds a predefined limit (v_{limit}) and FBs are applied (i.e. brake cylinder pressure $P_{\mathrm{x}} >$ pressure loss P_{L}) the fatigue protection value, $L_{\mathrm{fp}}(t)$, is increased by a term obtained by integrating the time variable. If the brakes are released or if the vehicle speed is low, then $L_{\mathrm{fp}}(t)$ is decreased by integrating time with a scale factor $a \in [-1, 0]$ ($L_{\mathrm{fp}}(t) \geq 0$ always). b is the upper limit for $L_{\mathrm{fp}}(t)$ before the ACC is shut down. This ensures that the FBs are never utilized for long periods of time. After the ACC has been forced to shut down, it can be turned on when $FP(t)$ reaches 0 again. An example of a vehicle following situation, where the FB temperature reach T_{lim1} (at time 94) and initiates an evacuation phase, is shown in Figure 4.8. By applying the FBs, the distance to the forward vehicle is increased before handing over to the driver.

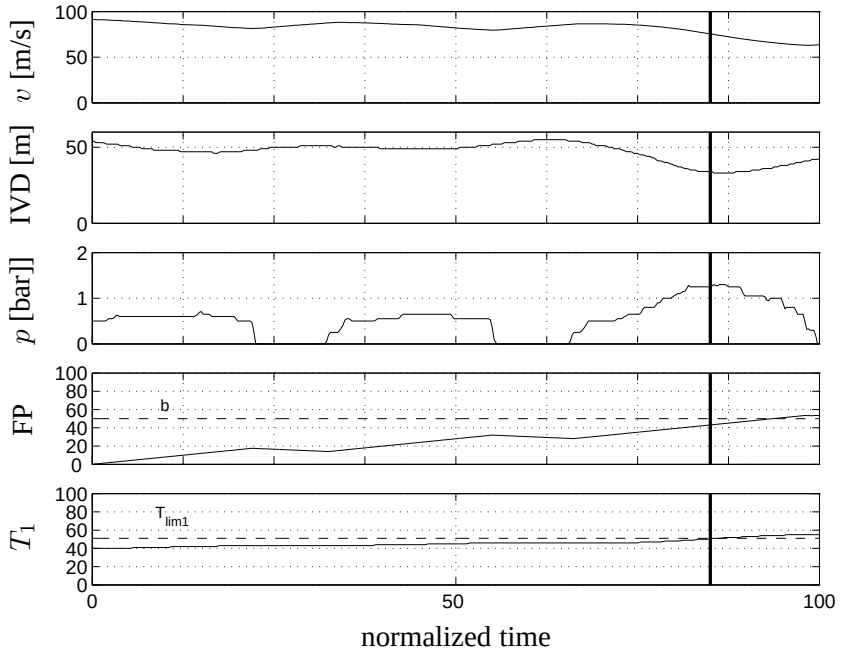


Figure 4.8: Evacuation when using the ACC system. FBs are applied in order to control the driver–selected inter–vehicle distance (IVD). The fatigue protection (FP) and FB temperature (T_1) increases due to the FB applications. At time 94, T_1 reaches the limit value of T_{lim1}, and an evacuation is initiated resulting in increased IVD before handing over to the driver. Time, FP, and T_1 are given in normalized units.

Also the fatigue protection is shown in Figure 4.8.

4.3 Adding road profile preview

Comparing a real driver and the controllers in Paper 2 and 3 a significant difference is that the controllers do not get any information about the road profile ahead, whereas a real driver can see a couple of hundred meters in front of the vehicle, depending on the terrain. Therefore, the controllers are basically equivalent to a driver that knows the internal states of the vehicle, but is blind. In Lingman (2002) road profile preview was added to the neural network (NN) controller. It was concluded that the road profile preview reduced the conservative behavior (low mean speed) of the NN on moderate slopes. In this section the usage of road profile preview in the temperature mode (TM) controller in Paper 3 is presented.

In reality, road profile preview requires either a combination of a map and a GPS receiver, or that the vehicle should be supplied in advance with road profile

as a function of position. It is believed that road profile information using GPS together with digital maps will be feasible for use in commercial vehicles in the future, and that road profile is necessary information for Advanced driver assistant systems (ADAS) to function, see e.g. the NextMAP project (run by the European commission) for discussions about combining digital maps and GPS receivers for use in different driver assistant systems working in both the lateral and longitudinal direction.

One way to utilize preview information in the TM controller is to include samples of the road profile into the feedback. By extending the original state vector used for the controller design ($x = [\dot{T}_1, \dot{T}_2, v\dot{v}, \dot{\alpha}, T_1]$) with

$$\begin{aligned} \dot{\alpha}_1(t+1) &= \dot{\alpha}_2(t) \\ \dot{\alpha}_2(t+1) &= \dot{\alpha}_3(t) \\ \dot{\alpha}_3(t+1) &= \dot{\alpha}_4(t) \\ \vdots &\quad \vdots \\ \dot{\alpha}_{\mathrm{n}-1}(t+1) &= (1 - \mathrm{dt}/\mathrm{T}_\alpha)\dot{\alpha}_{\mathrm{n}}(\mathrm{t}) + \nu(\mathrm{t}), \end{aligned} \tag{4.3}$$

where $\dot{\alpha}_{\mathrm{j}}(t)$ is the road slope rate $j-1$ samples ahead of the vehicle. Note that the road slope rate is used according to the velocity–based formulation utilized in Paper 3. The structure of Equation 22 in Paper 3 is then expanded into

$$\begin{bmatrix} \dot{T}_1(t+1) \\ \dot{T}_2(t+1) \\ v\dot{v}(t+1) \\ T_1(t+1) \\ \dot{\alpha}_1(t+1) \\ \vdots \\ \dot{\alpha}_{\mathrm{n}-1}(t+1) \end{bmatrix} = \begin{bmatrix} F & E \\ 0_{(\mathrm{n}-1)\times 4} & D \end{bmatrix} \begin{bmatrix} \dot{T}_1(t) \\ \dot{T}_2(t) \\ v\dot{v}(t) \\ T_1(t) \\ \dot{\alpha}_1(t) \\ \vdots \\ \dot{\alpha}_{\mathrm{n}-1}(t) \end{bmatrix} +$$

$$\begin{bmatrix} 1/c_1 \\ 0 \\ -1/m \\ 0_{\mathrm{n}\times 1} \end{bmatrix} \dot{u} + \begin{bmatrix} 0_{(2+\mathrm{n})\times 1} \\ 1 \end{bmatrix} \dot{\alpha}_{\mathrm{n}}(t),$$

where F and E build up the discrete system matrix from Equation 22 in Paper 3, n is the number of preview points and $0_{\mathrm{p}\times\mathrm{q}}$ is a $\mathrm{p} \times \mathrm{q}$ matrix of zeros. The matrix D is

$$D = \begin{bmatrix} 0 & 1 & 0 & \cdots & & 0 \\ 0 & 0 & 1 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \\ & & & & & 0 \\ & & & & & 1 \\ 0 & & & & & (1 - \mathrm{dt}/\mathrm{T}_\alpha) \end{bmatrix},$$

where dt is the sampling time. Note that for convenience states four and five in Equation 22 are shifted within the state vector in order to group all road slope states.

The size of the new state vector depends on the required preview time. If a preview time of t_{pv} is required, the state vector will consist of the original state vector augmented with $t_{\mathrm{pv}}/\mathrm{dt}$ new states. In Figure 7 in Paper 4 the open–loop optimal solution required roughly 800 m in preview for a step in road slope from 6% to 8 %. This preview corresponds to approximately 40–50 s of preview (for that particular speed and slope). In Figure 4.9, three different system trajectories are shown, one for the open–loop optimal (dash–dotted lines), another for the TM controller without preview (solid lines), and finally the TM controller with a 40 s preview (dotted lines). Note that in the open–loop case the preview covers the whole problem horizon.

From Figure 4.9 it can be seen that the temperature transient decreases significantly with preview information. Comparing no preview at all to 40 s preview shows a 70% reduction of maximum temperature overshoot. Naturally, increasing the preview will bring the closed–loop controller closer to the open–loop performance. Additionally the control signal activity is significantly reduced when using road profile preview. Another benefit of using preview information in the TM controller is in situations where the upcoming road slope is rapidly decreasing and a safe acceleration, propelled by the gravity force acting on the vehicle, would be desirable in order to avoid unnecessary braking and fuel consumption. In Figure 4.10 one example of this is shown. In the beginning, the vehicle is in velocity mode (VM) with a driver selected set speed of 17 m/s. When the downhill slope begins, FBs are applied to control the vehicle speed. This increases the FB temperature which leads to the TM eventually being engaged. In TM the vehicle speed is reduced significantly in order to keep the FB temperature within resonable limits. At 3500 m the slope decreases and due to the preview, the FB force (and VEB force) is reduced in advance, and the gravity force accelerates the vehicle. Thus, eventually VM is engaged and the vehicle speed is back at the driver set speed.

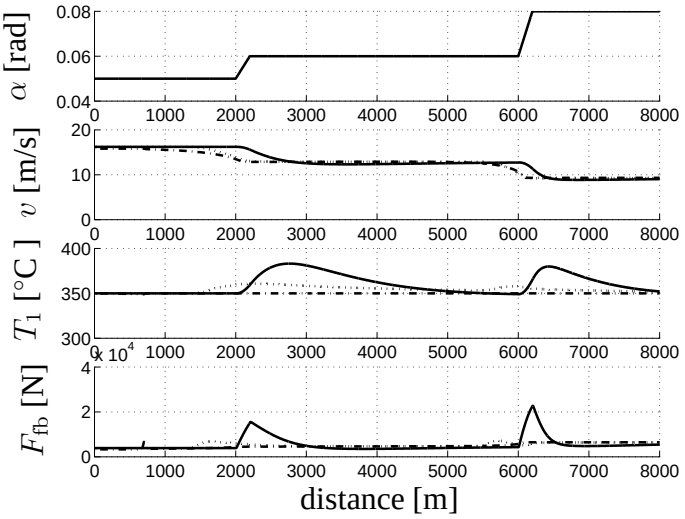


Figure 4.9: Comparing step responses for closed–loop control using no preview (solid line), using 40 s preview (dotted line), and open–loop control (dash–dotted line). A 40 s preview significantly reduces the overshoot in FB temperature when compared to the no preview at all. Note that the 40 s preview case approaches the open–loop (infinite preview) case.

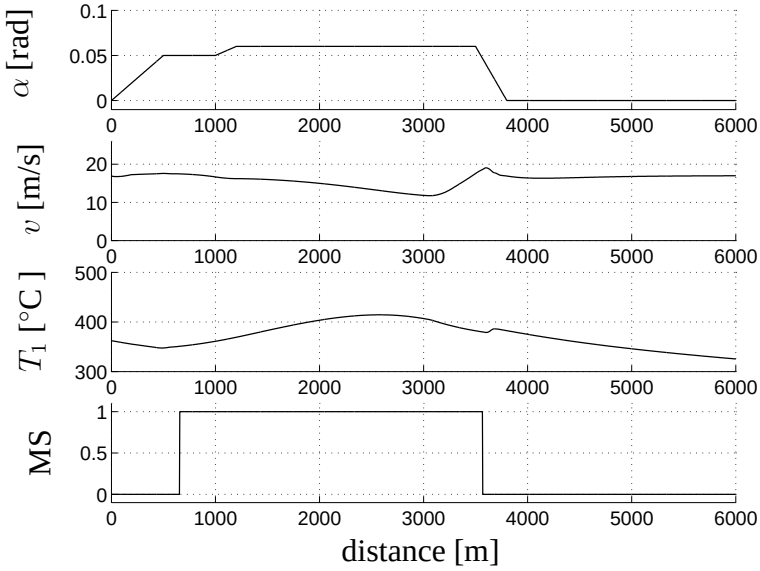


Figure 4.10: Preview in the TM controller results in reduced usage of brakes and fuel. In situations where the road slope decreases rapidly, the gravity force can be used to accelerate the vehicle to the driver–selected set speed.

4.4 Retardation economy

In this section, some important factors that affect transport economy are presented. Then, vehicle retardation is discussed from a transport economy perspective, and the concept of retardation economy is defined.

As discussed by Lindkvist and Gustavsson (1980), transport economy can be described differently depending on where the system boundary is set. If the system boundary is set around the vehicle and the owner of the vehicle, transport economy can be defined using:

- Income from transport mission
- Fuel cost
- Component wear (pads, tyres etc.)
- Other maintenance cost (engine service etc.)
- Utilization of vehicle (hours per day)
- Driver cost (salary etc.)
- Cost of vehicle purchase or rent
- Other costs (tax, road fee etc.)

Looking at transport economy from a retardation control point–of–view, component wear for low maintenance cost and system performance for high mean speed to improve transport effectiveness are important. Mathematically this leads to a multi–objective optimization problem, where the objective function (or fitness function, i.e. the inverse of the objective function) is built up by three objective functions. The objective function is given in Equation 4.4 where x represents the system states.

$$\begin{aligned} F(x) &= f_1(x) + f_2(x) + f_3(x) \\ &= f_{\text{wearABs}}(x) + f_{\text{wearFBs}}(x) + f_{\text{speed}}(x). \end{aligned} \tag{4.4}$$

In most applications, the objective function consists of objectives that are in conflict with each other. It is very unrealistic to expect that a feasible solution, x^*, can be found that minimizes all the objectives. In Equation 4.4, $f_1(x)$ and $f_2(x)$ are given in euro whereas $f_3(x)$ is given in s/m, see Equations 4.5 and 4.6. Comparison is therefore very difficult to make even if it, at least theoretically, would be possible to assign a cost for time. If one would assume a cost for time, different

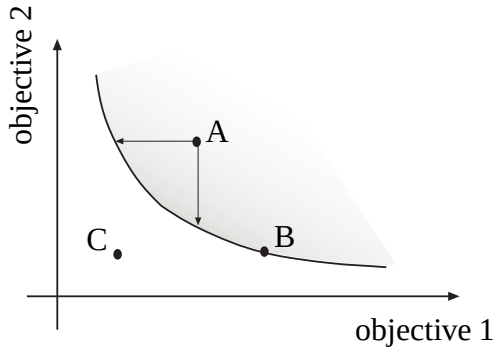


Figure 4.11: Typical Pareto curve for 2 objectives. A–Inefficient B–Efficient C–Infeasible.

assumptions (depending on, for example, different companies and what goods that are transported) of the cost of time would result in different solutions. A different definition of optimality would therefore be useful. One such criterion is **Pareto optimality**[1]. The solution x^* is said to be optimal (or Pareto optimal, or efficient, or non-dominated) if there is no other solution x such that $f_j(x) < f_j(x^*)$ for some j and $f_{\mathrm{i}}(x) \leq f_{\mathrm{i}}(x^*)$ for all $i \in [1, N]$, where N is the number of objectives. The set of points for which the optimality conditions hold is called the Pareto surface (or Pareto front or Pareto curve in the bi-objective case) and it constitutes a graphical representation of the trade-off between the objectives.

$$J_{\mathrm{speed}} = \overline{v}. \tag{4.5}$$

$$J_{\mathrm{wear}} = \frac{1}{C_{\mathrm{tot}}}. \tag{4.6}$$

The concept of Pareto optimality was used in Paper 4 to study the trade–off between mean speed (f_3) and component wear cost $f_1 + f_2$ (See Equation 4.7. In Paper 2, the component wear was investigated under constant speed conditions and therefore the criterion function consisted of only $f_1 + f_2$.

Before proceeding with the definition of wear cost, the concept of transport efficiency requires some clarification. It should be mentioned that efficient use of the brake system will increase the mean speed on downhill slopes, but it will not necessarily affect the mean speed on the whole driving cycle. If large parts of the driving cycle consists of relatively flat road sections, the mean speed (transport time) is not much affected by an increase in the speed on single down hill

[1]Vilfredo Pareto was an Italian economist (1848–1923).

slopes. However, high mean speed on down hill slopes is often desirable also to increase the driver comfort and satisfaction (i.e. driveability). Here, a comparison with the development of engines can be made. Reducing emissions and fuel consumption are prime targets for engine developers but other customer demands like engine response, power reserve, top gear gradeability etc.also have to be considered. Stronger and more powerful engines are developed, even if the mean speed on the whole driving cycle does not necessarily increase with more powerful engines.

The wear cost can be formulated in several different ways, and one of the main questions is whether or not to include the time of maintenance stops, i.e. the time the truck is not productive. As mentioned above, the cost of time is not considered here. Therefore $f_1 + f_1 = f_{\text{wear}}$ can be given as

$$\begin{aligned} f_{\text{wear}} &= W_{\text{disc}} \frac{C_{\text{discm}} + C_{\text{discw}}}{\delta_{\text{disc}}} + W_{\text{pad}} \frac{C_{\text{padm}} + C_{\text{padw}}}{\delta_{\text{pad}}} + \\ & W_{\text{tyre}} \frac{C_{\text{tyrem}} + C_{\text{tyrew}}}{\delta_{\text{tyre}}}. \end{aligned} \tag{4.7}$$

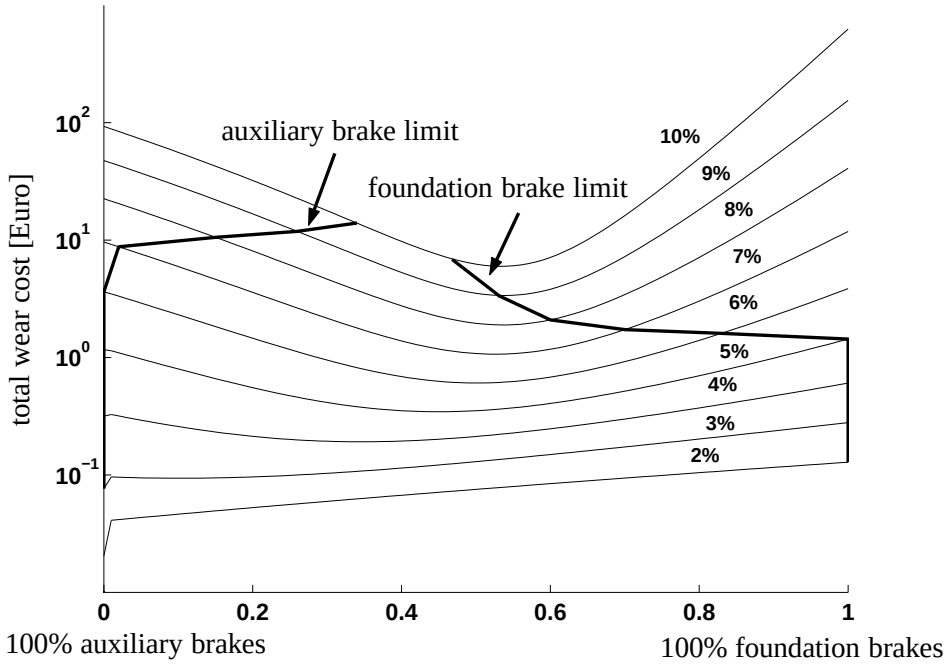
Numerical values for work cost and material cost used here are what the truck owner has to pay at an Volvo workshop in Sweden. Table 4.4 is a description of the components used in Equation 4.7.

Table 4.3: Components in Equation 4.7.

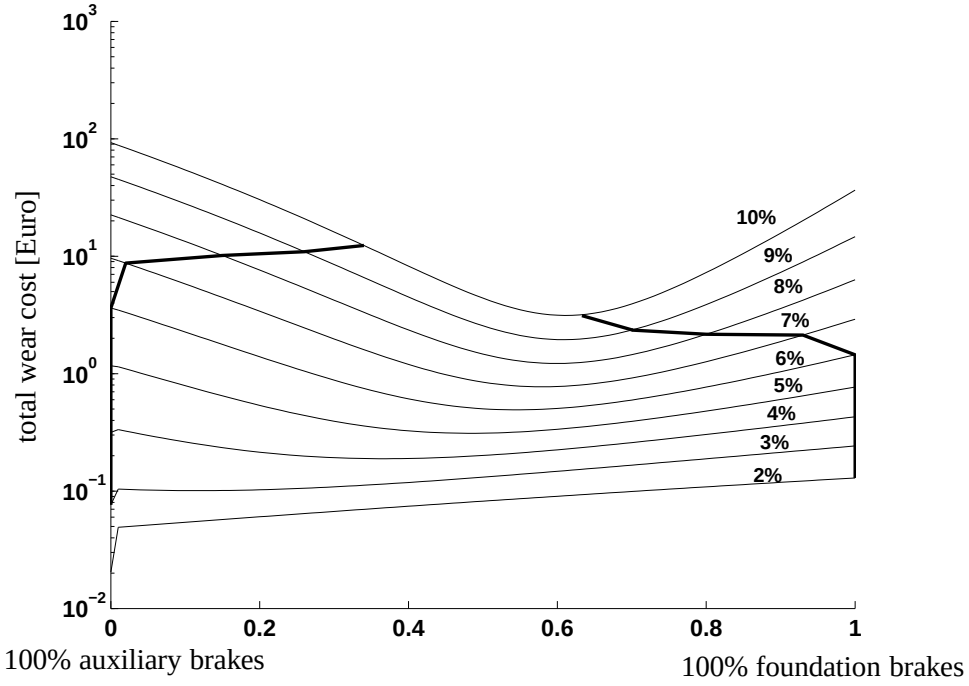
W_{disc}	Disc wear [mm]
C_{discm}	Material cost of disc [euro]
C_{discw}	Work cost to change disc [euro]
δ_{disc}	Available thickness of disc
W_{pad}	Pad wear [mm]
C_{padm}	Material cost of pad [euro]
C_{padw}	Work cost to change pad [euro]
δ_{pad}	Available thickness of pad [mm]
W_{tyre}	Tyre wear [mm]
C_{tyrem}	Material cost of tyre [euro]
δ_{tyre}	Available tyre thread [mm]

In order to evaluate Equation 4.4, a model of the vehicle is needed. The model used is explained in Paper 2 and 4. The disc wear is not modelled, but in practice the rule of thumb used at workshops is that every second time pads are changed, discs are also changed. Thus, the model for pad wear described in Paper 2 is used for evaluating both pad and disc wear.

As mentioned earlier, drivers prefer to utilize the ABs to the fullest in order to minimize FB wear cost. In Figures 4.12 and 4.13 the wear costs (J_{wear}^{-1}) for four different 60 tonnes carriage configurations are shown: 6x2 with 6 axles, 6x2 with 8 axles, 6x4 with 7 axles, and finally 6x4 with 9 axles. Note that configuration NxM means a truck with a total of N wheels out of which M are connected to the drive shaft. The length of the descent was 3000 m and the slope was varied between 2% and 10%. For each carriage combination and slope, an optimal constant distribution can be found that maximizes f_{wear}. This indicates that the strategy used by drivers today is non-optimal when it comes to brake economy and that the name ”non-wear brakes” can be somewhat misleading due to the tyre wear.

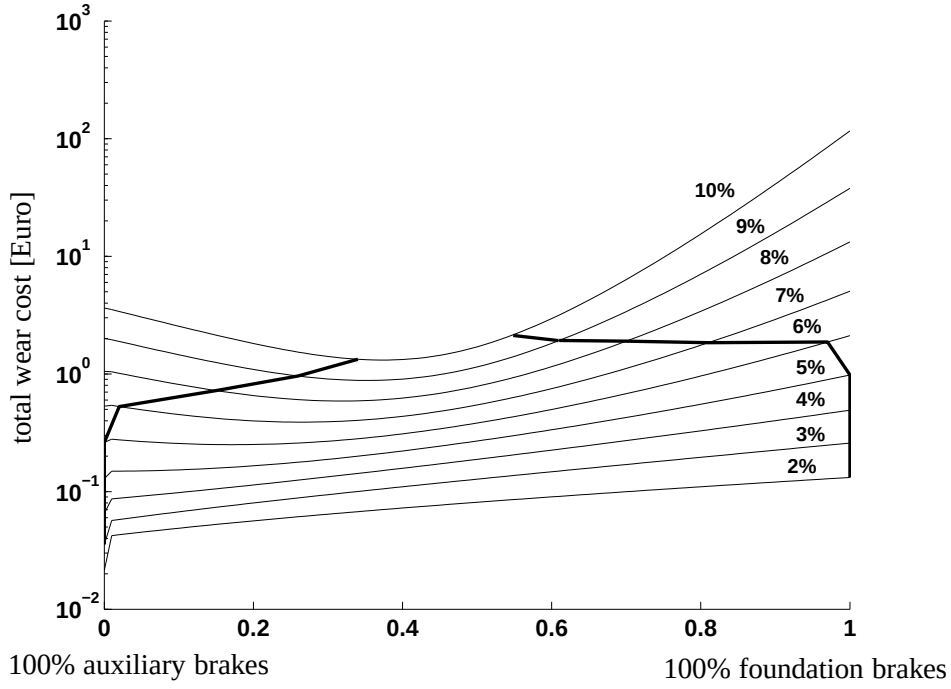


(a) 6x2 with 6 axles

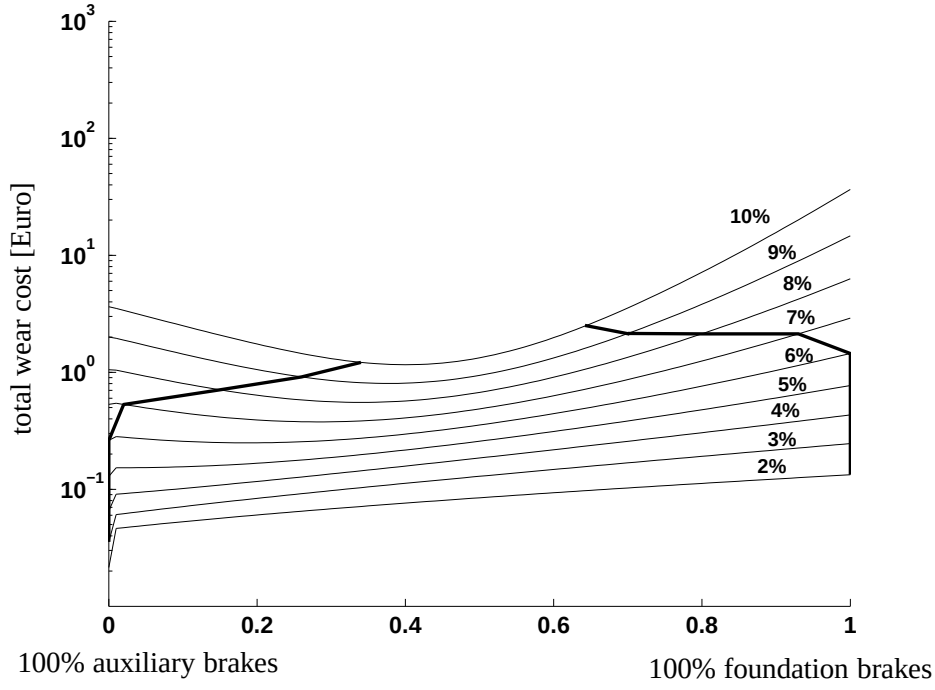


(b) 6x2 with 8 axles

Figure 4.12: Wear cost vs. distribution between ABs and FBs. 6x2 case where the drive–axle has twin mounted tyres and the pusher or tag axle has single mounted tyres. Vehicle speed is 15 m/s and the length of the slope is 3000 m.



(a) 6x4 with 7 axles



(b) 6x4 with 8 axles

Figure 4.13: Wear cost vs. distribution between ABs and FBs. 6x4 case where both drive–axles have twin mounted tyres. Vehicle speed is 15 m/s and the length of the slope is 3000 m.

Chapter 5

Models and methods

In this chapter, the models and methods used in the development of retardation strategies are explained (see also Papers 2, 3, and 4). The method used to acquire information of vehicle states is explained as well (see also Paper 1).

5.1 Retardation controller design

The first, and often most difficult step, when designing a controller for a system is to define the objective for the control, i.e. to define a criterion function. The criterion function should be minimized or maximized using an optimization algorithm. For longitudinal retardation control, the criterion could be mean speed maximization or component wear cost minimization. Additionally, constraints such as disc temperature have to be within limits for safety reasons.

Maximization of mean speed can be transformed into a minimum time problem which constitutes a special class of so called **optimal control problems**, see for example Ljung and Glad (2000). The solution to minimum time problems is often expressed in the form of a time-varying control signal. No feedback controller is obtained, making direct implementation very difficult. However, optimal control will reveal the fundamental properties of the dynamic problem; the resulting trajectories can be used in the verification and development of implementable strategies, and one could thus say that the solution to the optimal control problem probes the limit for what is theoretically achievable in an implementation. In Paper 4, the optimal control problem is defined and solved using state–of–the–art optimization software.

As for implementable (closed–loop) controllers, model predictive control (MPC) is an appealing method that can be used on problems where boundary conditions have to be considered (like in downhill driving). The basic idea with MPC is to use a model of the process that calculates future system outputs for a set of dif-

ferent future control signals. The best control signal is then chosen according to a certain criterion function (like maximum vehicle speed). A common problem in the implementation of MPC controllers is that the procedure requires much computational power since an optimization with constraints have to be performed at each sampling. MPC is often used in process control applications, e.g. in the petrochemical industries. In this thesis, two other approaches have been used in the design of a controller. In the first approach (Paper 2) the controller is represented by a **feedforward** neural network (**FFNN** or simply **NN**) and in the second approach (Paper 3) linear quadratic control (**LQ**) with gain scheduling was used. The different methods will now be explained.

5.1.1 Open–loop optimization

In many situations it is important to know what can be achieved theoretically, i.e. to determine the upper bound for the system performance. An example is the design of a controller for downhill driving were, clearly, knowledge of the theoretical upper speed (subject to velocity constraints, actuator constraints etc.) for a truck would be very useful. The theoretical limit can then be used both to inspire and to verify the development of implementable, closed–loop, strategies. In general, the optimal control problem is reduced to finding a control signal $u(t)$ that minimizes the criterion function

$$F = \phi\left[y(t_{\mathrm{F}}), t_{\mathrm{F}}\right], \tag{5.1}$$

subject to the state equations

$$\dot{y} = f\left[y(t), u(t)\right], \tag{5.2}$$

the boundary conditions

$$0 = \psi\left[y(t_{\mathrm{F}}), u(t_{\mathrm{F}}), t_{\mathrm{F}}\right], \tag{5.3}$$

and the path constraints

$$c_{\mathrm{L}} \leq c(x) \leq c_{\mathrm{U}}, \tag{5.4}$$

where t_{F} is the final time. Note that the concept of **phases** is left out in the above equations, see Paper 4. In general, it is very hard to find the analytical solution to this infinite-dimensional (continuous–time), problem. In the theory for solving optimal control problems (see e.g. Betts (2001)), the strategy is instead to transform the problem into a finite-dimensional one and then apply numerical optimization methods to find the optimal solution $u^*(t)$. In Betts (2001) the method for transforming the problem into a finite set is referred to as a **transcription**

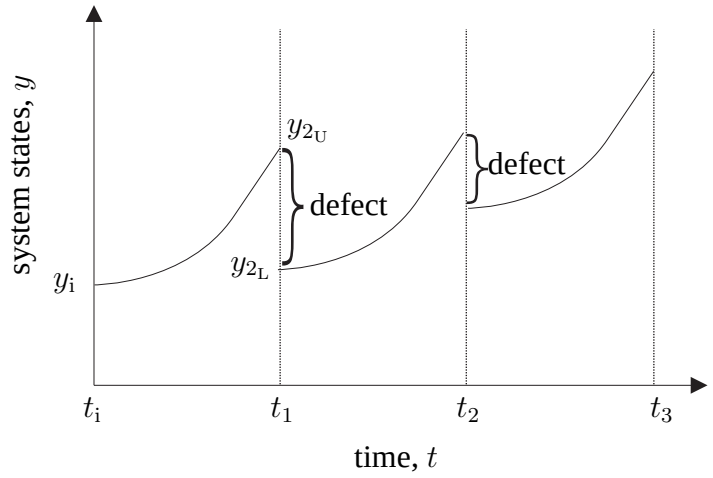


Figure 5.1: The multiple shooting method. State equations result in a set of defect constraints. In this example the number of grid points $n_{\text{grid}} = 4$.

method (**multiple shooting**). Scalar variables are used to represent the values of the continuous functions (See Equations 5.1, 5.2, and 5.4) in a finite number of time instants n_{grid}. A numerical integration method is then used to propagate the differential equations between consecutive time instants, as shown in Figure 5.1. The points corresponding to the time instants ($t_1, t_2, \ldots$) are called *nodes* or *grid points*. When the defect, that arise from the discretization of the ODEs, approach zero, the system state equations (i.e. the defect constraints) are fulfilled. Nonlinear programming (**NLP**) is usually applied to solve the finite optimization problem that arises from the transcription method.

Frequently used integration methods are Euler, trapezoidal and Hermite-Simp son. These methods vary in ease of implementation, accuracy and computational cost associated with the constraint evaluations. The Euler integrator is of order one (low accuracy) and the Hermite–Simpson integrator is of order four (high accuracy). The choice of appropriate integrator is strongly dependent on the problem at hand. In Paper 4, the trapezoidal integrator is used since it combines good accuracy with relatively low computational effort. Applying the trapezoidal integrator to Equation 5.2 results in constraints according to Equation 5.5, where Δt is the time step in the grid.

$$\mathbf{y}_{k+1} - \mathbf{y}_k - \frac{\Delta t}{2}\left(\dot{\mathbf{y}}_{k+1} + \dot{\mathbf{y}}_k\right) = 0. \tag{5.5}$$

The resulting NLP is thus reduced to the minimization of the criterion function

$F(x)$, subject to the ODE (i.e. the defect constraints) and path constraints

$$\begin{aligned} C_{\mathrm{e}} &= 0, \\ C_{\mathrm{i}} &\geq 0. \end{aligned} \tag{5.6}$$

Note that x now represents both the system states, y, and the control signals, u, i.e. ($x = (y, u)$).

There exists a wide range of different methods to solve special cases of NLPs, like quadratic programming, **QP** (quadratic criterion and linear constraints) and linear programing, **LP** (linear criterion and linear constraints). A very successful method to solve the general NLP is **sequential quadratic programming** (**SQP**). Below, a simplified version of a SQP algorithm is described, Betts (2001).

1. Terminate optimization if convergence is reached, i.e. terminate when there are no constraint violations and when the solution vector does not improve anymore.

2. Use first-order Taylor expansion to calculate an approximation to the constraints, i.e. calculate the Jacobian matrix (J) where m is the total number of constraints.

$$J = \nabla_x C_{\mathrm{e,i}} = \begin{bmatrix} \frac{\partial C_1}{\partial x_1} & \frac{\partial C_1}{\partial x_2} & \cdots & \frac{\partial C_1}{\partial x_n} \\ \frac{\partial C_2}{\partial x_1} & \frac{\partial C_2}{\partial x_2} & \cdots & \frac{\partial C_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial C_m}{\partial x_1} & \frac{\partial C_m}{\partial x_2} & \cdots & \frac{\partial C_m}{\partial x_n} \end{bmatrix} \tag{5.7}$$

   In optimal control problems, the Jacobian matrix will be sparse. In general the number of constraint evaluations equals $n_{\mathrm{grid}} n_{\mathrm{var}} + 1$, were n_{var} represent the number of states and control signals. However, since only adjacent points of the defect constraints will interact, the sparsity of the Jacobian can be further exploited by, for example, grouping the terms by grid point, also referred to as discretization separability, Betts (2001). Exploitation of the sparsity reduces the computational effort involved in calculating the Jacobian matrix.

3. Evaluate the gradient (g) and Hessian (H) in order to calculate the second order approximation of the criterion function:

$$g = \nabla_x F = \left[\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}, \ldots, \frac{\partial F}{\partial x_n} \right]^\top \tag{5.8}$$

$$H = \nabla_{xx} F = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 x_2} & \cdots & \frac{\partial^2 F}{\partial x_1 x_n} \\ \frac{\partial^2 F}{\partial x_2 x_1} & \frac{\partial^2 F}{\partial x_2^2} & \cdots & \frac{\partial^2 F}{\partial x_2 x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n x_1} & \frac{\partial^2 F}{\partial x_n x_2} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix} \tag{5.9}$$

The Hessian is also sparse.

4. Solve the QP subproblem, which is derived using the results in steps 2 and 3, to obtain the search direction $p = x_{\mathrm{t}} - x_{\mathrm{t}-1}$.

5. Since the search direction is based on approximations that are valid only in the close vicinity of $x_{\mathrm{t}-1}$, a line search must be performed in order to find a suitable step length, a, along the search direction p. The step length, a, is thus the result of an optimization procedure that gives a suitable reduction in the merit function (this is called a globalization strategy), Betts (2001), i.e. $x_{\mathrm{k}} = x_{\mathrm{k}-1} + ap$.

6. Return to step 1.

In Paper 4, the nonlinear program has been solved using the powerful optimization package TOMLAB provided by Tomlab optimization, Holmström and Göran (2003).

5.1.2 Controller design using genetic algorithms

A feedforward neural network (**FFNN**) is a nonlinear function mapping from a set of input signals (Y) to a set of output signals (U). The computational nodes, called neurons, are connected to each other forming a network structure as shown in Figure 5.2. Each neuron is represented by a summation operator and a nonlinear **activation function** ($\sigma(s)$) that limits the output from each neuron. In this thesis the **sigmoid function** was chosen as activation function: $\sigma(s) = 1/(1 + e^{-s})$, where s is the sum of all input signals to the neuron and the output (z) from each neuron is therefore given by: $z = \sigma(\Sigma w_{\mathrm{ij}} x_{\mathrm{j}} + w_{\mathrm{bias}}\Theta)$ as shown in figure 5.3. This output signal is multiplied by a weight (w) and then transmitted as an input signal to a neuron in the next layer, as illustrated in Figures 5.2 and 5.3. $w_{\mathrm{bias}}\Theta$ ($\Theta \equiv +1$) is a bias term which sets the output level of the neuron in the absence of input. Typical input and output sets used in this thesis are Y:{speed, disc temperature, road slope, coolant temperature, engine speed} and U:{retardation force request, force ratio between FBs and ABs, force ratio between VEB and CR, gear change}, as illustrated in Figure 5.4. Since the activation function limits the output to [0,1], post– and preprocessing of signals measured in the vehicle and

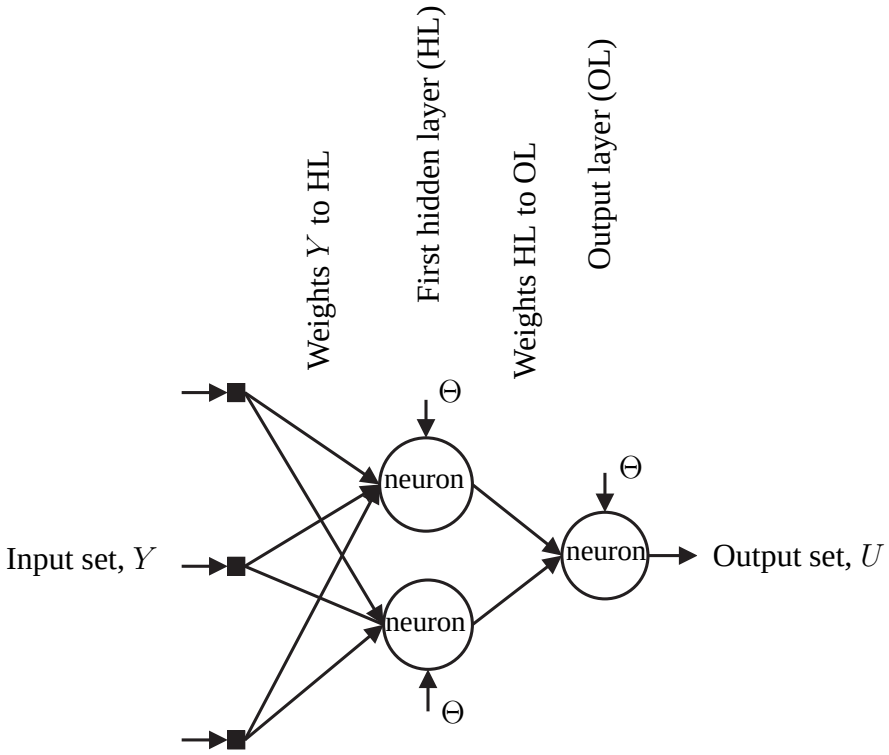


Figure 5.2: Neurons forming a net structure, neural network.

signals controlling the actuators have to be performed as shown in Figure 5.4 (see also Haykin (1994)).

In the design of the NN controller both the structure (i.e. the number of layers and connections) and the weights have to be chosen in such a way that the criterion function is minimized. In this thesis the NN is optimized using a genetic algorithm (**GA**).

Genetic algorithms are stochastic search methods, ideal for complex optimization problems. GAs are strongly inspired by biology and therefore the terminology regarding GA has its origin in biology, see e.g. Wahde (2005). The starting point for all GAs is to form a set of different candidate solutions to the problem. This set is usually referred to as the **population** and the candidates are called **individuals**. The properties for each individual are numerically coded into **genes** forming strings known as **chromosomes**. In Figure 5.5, a chromosome for an individual is shown. In this example, real–number encoding is used so that each gene is a real

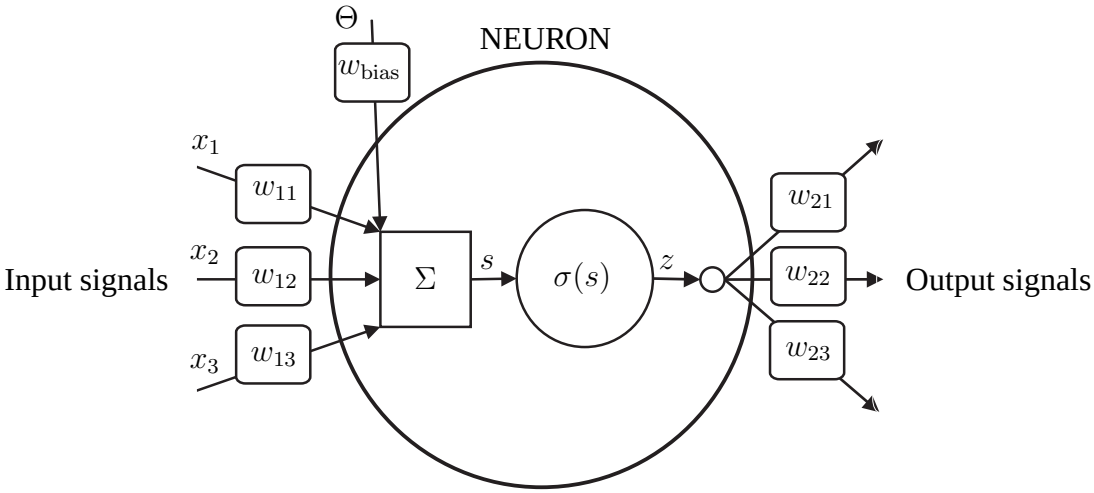


Figure 5.3: A neuron.

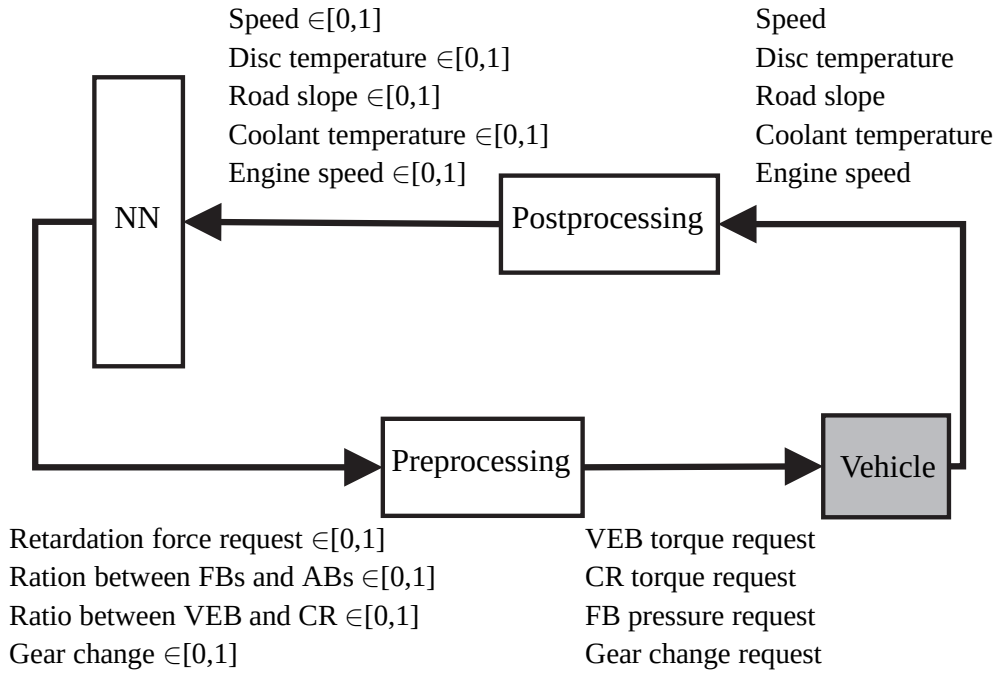


Figure 5.4: Closed–loop feedback.

number in the range [0,1] representing one variable. Assuming for example that the variables are in the range [-10,10], the decoding will give the corresponding variable v as

$$v = -10 + 2 \times 10 \times g, \tag{5.10}$$

were g is the value (allele) of the gene. The procedure is illustrated in Figure 5.5. In this thesis the structure of the neural network was chosen manually and therefore each individual in the population represents the weights of a neural network with a given structure.

By applying the different steps of a GA, the aim is to move the whole population to the point of optimality defined by the criterion function (or **fitness function**). The different steps of the GA are based on three main operators: **selection**, **mutation**, and **crossover**.

Selection is a fundamental part of the GA. Individuals are selected in pairs from the population using fitness–proportional selection. These individuals are then transformed into two new individuals, by performing crossover and mutations.

Crossover is an exchange of information between two individuals. A point of crossover is randomly chosen along the two strings, dividing each string into two parts. Information is then exchanged by switching one part of one string with the corresponding part in the other string, as shown in Figure 5.6. Mutation operates by drawing a random number for each gene along the chromosome. If this number is smaller than a predefined number (the **mutation probability**), the corresponding gene is assigned a new random value in the allowed range. The construction of new individuals proceeds until a new population has been created to replace the old one. The procedure of evaluating and creating new populations continues until a satisfactory solution to the problem has been found, see Figure 5.6.

It is important to understand that GAs are far from random search methods: Mutation is random, as is the choice of crossover points, but selection is not, since it is carried out (although stochastically) in proportion to fitness.

The main reasons for using a NN representation of driving strategies is that a feedback controller is obtained for a process with hard and soft nonlinearities. Using GAs to optimize the NN is a natural approach in this case. Traditional gradient descent methods such as e.g. back–propagation, see Haykin (1994), cannot easily be applied here, since there exists no training set consisting of input–output pairs. An additional advantage is that both the NN and the GA are easy to implement in software.

From the user's point–of–view an additional advantage with these methods is that one can affect the final result, by tuning comprehensive parameters rather than directly tuning the weights of the NN. For example, if the NN performs badly on

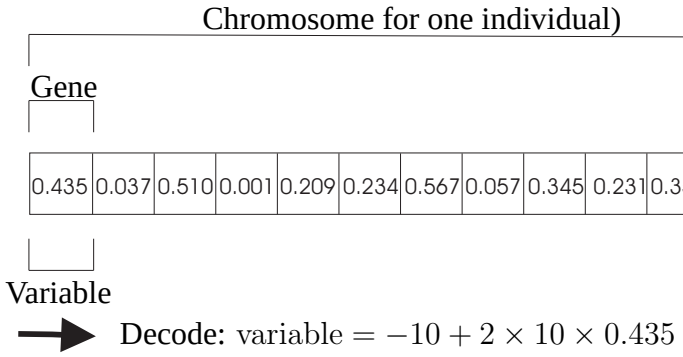


Figure 5.5: Chromosome representing one individual. In this example real–number encoding is used.

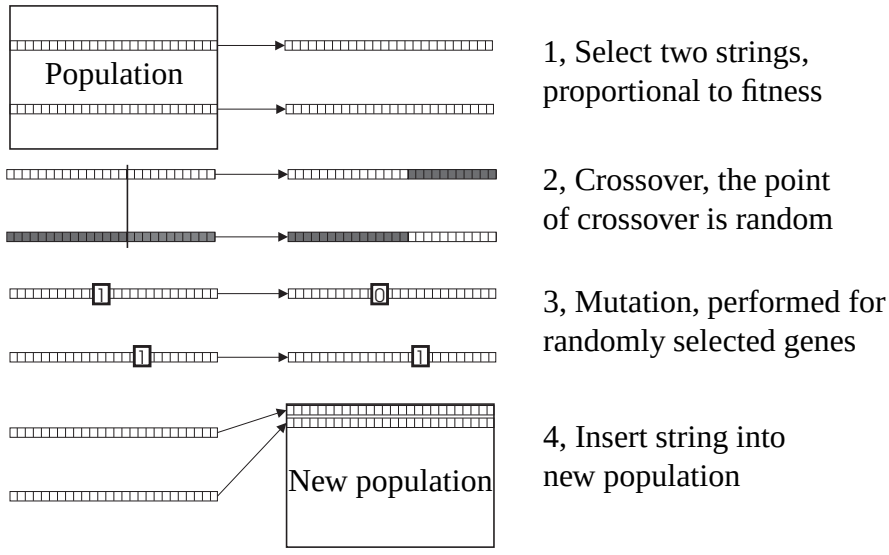


Figure 5.6: The formation of two new individuals in the GA.

one type of road profile, this road profile can easily be added to the training road, indirectly influencing the final NN to manage also this obstacle.

An objection to the use of this kind of approach is that it can be difficult to verify closed–loop robustness. There exists no standard methods to verify robustness for systems were the feedback gain is an NN. In this work the robustness analysis of the NN controller is made by simulations, i.e. the obtained NN is tested on a set of different road profiles (measured road profiles) designed to cover all possible situations that are likely to occur in reality. The robustness is then quantified according to Equation 12 in Paper 2. Another problem is that it may be difficult, due to the relatively high computational effort involved when using GA, to perform trade–off studies. If, for example, multi–objective criterion (fitness) functions are used, obtaining Pareto curves may require very much computational effort due the large number of optimizations involved. Instead, using open–loop optimization showed to be a powerful method to study the trade–off between component wear cost and transport efficiency.

5.1.3 Closed–loop control using gain scheduling

As pointed out in Paper 2, real vehicle implementation of the FFNN controller is not straightforwardd since, for example, tuning in vehicle may be rather time consuming. Therefore another method for controller synthesis was required. In the literature several nonlinear controller synthesis methods can be found: model predictive control, generalized predictive control, feedback linearization etc, see for example Ljung and Glad (2000) and Schmidtbauer (1999). Another method, which has proved to work in a variety of practical applications, is gain scheduling. The basic idea with gain scheduling is to find a set of linear process models that together describe the nonlinear dynamics of the process. Once the set of linear models is obtained, controller synthesis can be made with a suitable linear design method. Gain scheduling can be seen as a divide-and-conquer approach for the design of nonlinear control systems. The general algorithm is:

1. Linearize (using a first-order Taylor expansion) the nonlinear process about a number of equilibrium points.

2. Design a linear controller for each of the process linearizations.

3. Combine the linear controllers to obtain a nonlinear controller.

The procedure can be described mathematically as

$$\dot{x} = F(x, u). \tag{5.11}$$
$$y = G(x, u). \tag{5.12}$$

$$\delta\dot{x} = \nabla_x F(x_0, u_0)\delta x + \nabla_u F(x_0, u_0)\delta u. \quad (5.13)$$

$$\delta y = \nabla_x G(x_0, u_0)\delta x + \nabla_u G(x_0, u_0)\delta u. \quad (5.14)$$

$$\delta u = u - u_0,\ \delta y = y - y_0,\ \delta x = x - x_0,\ \delta\dot{x} = \dot{x}. \quad (5.15)$$

Here, linearization of the nonlinear system F, G in Equations 5.11 and 5.12 about an equilibrium point, yields the linear system in Equations 5.13 and 5.14, with system state vector δx and output vector δy. This linear approximation is valid only in the close vicinity of the equilibrium point. Therefore, linearization has to be performed at several equilibrium points in order to cover the complete operating range of the process. The choice of suitable linearization points is not trivial. In order to describe the nonlinear process well, a large number of linearization points is required. On the other hand, the use of a large number of linearization points results in a large number of different controllers, and therefore a trade–off between model accuracy and real time target performance must be made. One of the more challenging aspects of gain scheduling is the choice of scheduling variables (θ), i.e. the quantity that is used for the selection of the correct controller in each situation. Generally, from a closed–loop stability point–of–view, the scheduling variable should vary slowly in order not to introduce unstable modes into the nonlinear controller. Usually the appropriate scheduling variables are found after some testing. It should be mentioned that gain scheduling is generally considered as an ad-hoc approach to obtain a nonlinear controller.

In Paper 3, two different scheduling strategies are used, divided into case 1 and case 2. In case 1, the road slope (α) is used as scheduling variable. The road slope is a natural scheduling variable candidate. Firstly, the road slope is rather slowly varying (even though the dynamics of the FB temperature is sometimes in the same range). Secondly, the whole downhill driving process is forced by the road slope which causes the gravity force to accelerate the vehicle. One drawback with this approach was, however, that the transient performance of the controller was not good. Tuning of the parameters of the control design method (altering the grid of linearization points etc.) did not improve the situation. The problems arise from the fact that the system was not, during transient manoeuvres, in the vicinity of any equilibrium point.

In Figure 5.7 a set of equilibrium points is shown as function of FB force (control signal) and vehicle velocity (state). For example, for one velocity several equilibrium (or stationary) points can be found, uniquely separated by different combinations of FB force, FB temperature and road slope. Since the aim of the TM controller in Paper 3 is to control the FB temperature to a predefined level, $T_{1\mathrm{ref}}$, the scheduling points in case 1 were defined along the trajectory of constant FB temperature. This trajectory is marked with the dashed line in Figure 5.7 ($T_{1\mathrm{ref}}$ =500 °C in this example). Each point along this trajectory is defined by the scheduling variable $\theta = \alpha$. One problem is, however, that the process is far away

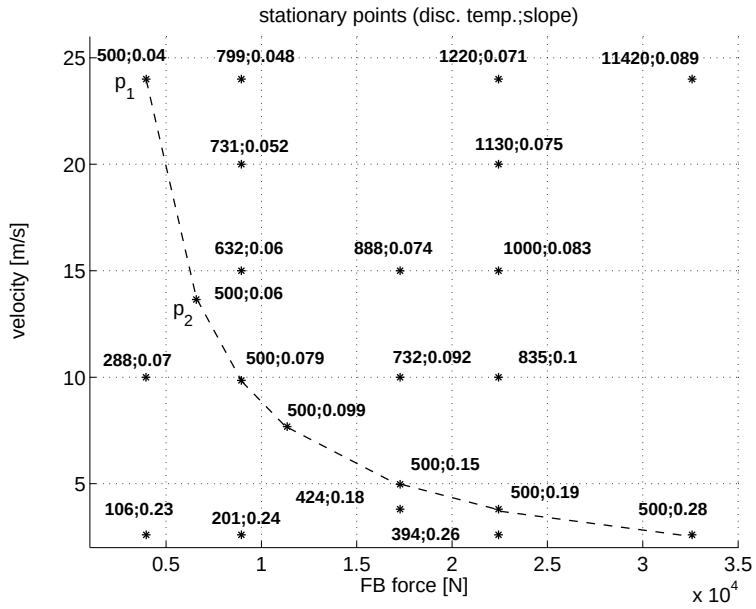


Figure 5.7: Stationary operation points.

from the equilibrium points, especially during transients. For example, if the road slope increases from 0.04 rad (p_1) to 0.06 rad (p_2), the linear representation of the system, which the controllers are based on, changes from equilibrium point p_1 to p_2. However, before reaching equilibrium at the new level of slope (p_2), the system states will travel far outside p_2, with high FB forces (close to 35,000 N) and a temperature transient reaching 600 °C (depends of course on the tuning of the controller). It should also be noted that the actual velocity of the truck has to be reduced from p_1 to p_2. However, as soon as the slope is changed, the system is assumed to be in the close vicinity of p_2, which it obviously is not.

To improve the transient performance, another scheduling strategy, **velocity-based linearization** (Leith and Leithead (1998)), was used in case 2 in Paper 3. A general, first–order approximation of the nonlinear system in Equations 5.11 and 5.12 is given by (note that $\nabla_x(a) = \nabla_x(x)_{x=a}$)

$$\begin{aligned} \dot{x} &= \{F(x_n,u_n) - \nabla_x F(x_n,u_n)x_n - \nabla_u F(x_n,u_n)u_n\} + \\ &\quad \nabla_x F(x_n,u_n)x + \nabla_u F(x_n,u_n)u. \end{aligned} \tag{5.16}$$

$$\dot{x} = w. \tag{5.17}$$

$$\dot{w} = \nabla_x F(x_n,u_n)w + \nabla_u F(x_n,u_n)\dot{u}. \tag{5.18}$$

The state and the input are the same at every operating point n. This system represents an approximation of the original nonlinear system at each operating

point (not only at equilibrium). However, Equation 5.16 is not a linear system representation due to the inhomogeneous terms caused by the off-equilibrium approximation. To overcome this problem, velocity-based linearization is used. By taking the time derivative of Equation 5.16, Equation 5.18 arises and the number of states is now twice the number of states in the original system. It should also be noted that the control signal now is $\dot{u}$.

The system in Equation 5.18 is, in fact, a linear representation of the original nonlinear system that is valid at every operating point of the system, i.e. not only in the close vicinity of an equilibrium operating point. The velocity-based scheduling was used in case 2 and the scheduling variables were chosen to be the road slope and velocity. It should also be noted that gain scheduling based on velocity-based linearization does not inherit the strong condition on slowly varying scheduling variables. In both case 1 and case 2 in Paper 3, the selection of the controller was based on the scheduling variables and linear interpolation.

For both the VM and TM, the controller design was performed using linear quadratic (**LQ**) control. LQ control is a model–based, linear, design method that results in a state feedback gain L. This state feedback gain is obtained by solving, at each scheduling point, the algebraic Riccati equation, i.e. Equation 5.20 with $\dot{S} = 0$, which minimizes the criterion function J in Equation 5.21. In fact, to allow calculation of state feedback gains in advance, which is preferable in real-time targets, the stationary Riccati equation was solved, i.e. the resulting state feedback gain, L, was an approximation to the optimal gain.

$$\dot{S} = SA + A^T S - SBR_2^{-1}B^T S + R_1. \quad (5.19)$$

$$L = R_2^{-1}B^T S. \quad (5.20)$$

$$J = E\{x^T R_1 x + u^T R_2 u\}. \quad (5.21)$$

5.2 Retardation system model

A model used for control purposes should be simple yet able to describe the main characteristics of the real process in the frequency range of interest. The force dynamics for the FBs and ABs are relatively fast, with time constants of around 0.5 s whereas the FB temperature and coolant temperature dynamics are significantly slower with time constants of around 100–500 s and 20–50 s, respectively.

The models used for temperature dynamics of FBs and coolant are thoroughly explained in Paper 2. They are built using a combination of physical modelling and experiments. The main nonlinearities in the FB temperature dynamics are introduced by the radiation and convection, and yield two characteristic parts, one

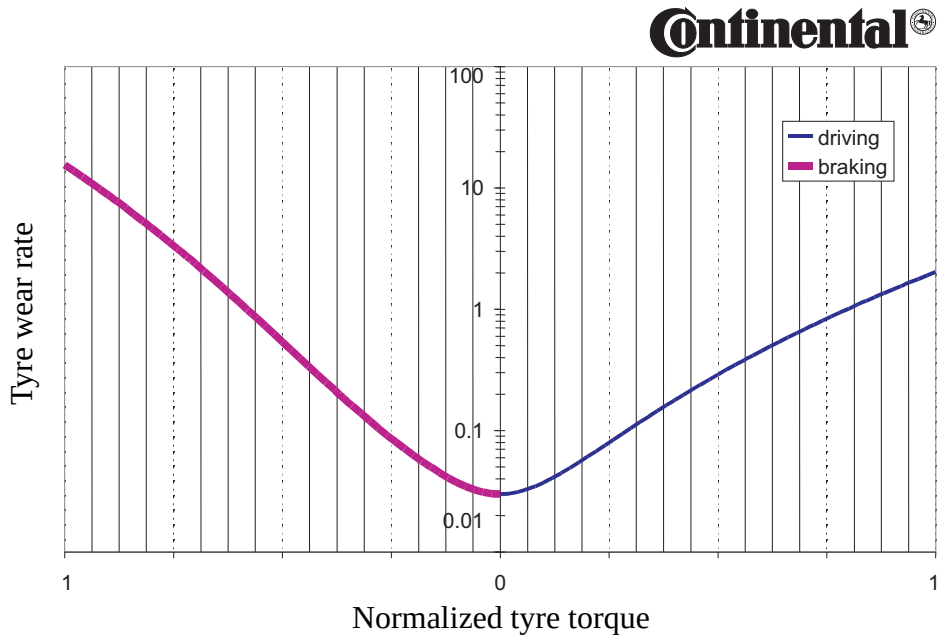


Figure 5.8: Tyre wear as function of tyre torque. Braking (Thick line) and driving (thin line).

fast and one slow, when subjected to a step in the clamp force, and for the coolant temperature it is introduced in the heat transfer from water to air where the ϵ-NTU method is used (the air flow through the radiator is also nonlinear in the vehicle speed and fan speed), see Andersson (1999) and Paper 2.

Several approaches for modelling the tyre wear can be found e.g in Tsai (1973) and Saleh (1979). A usual approach is to describe the tyre wear as a function of wheel slip and tyre temperature. In this thesis a tyre wear model describing the wear caused by longitudinal forces is used. The model used here was developed by Continental in Germany. It is based on experiments, and the tyre wear is formulated as a function of the torque applied on the tyre, during both driving and braking. In Figure 5.8 the tyre wear rate is shown as a function of tyre torque. It can be seen that the tyre wear rate is nonlinear in the applied torque and that braking causes more wear than driving, as discussed in Paper 2.
To summarize, the model used here consists of:

- Foundation brake force dynamics
- Auxiliary brake force dynamics
- Foundation brake temperature dynamics
- Foundation brake wear

- Tyre wear
- Cooler system dynamics
- Vehicle mass

5.3 State estimation

Today, in a real production vehicle, mass and road slope cannot be measured directly and therefore an estimation of these states must be performed. For this problem a model–based state estimation technique, called Kalman filtering, was applied, see Schmidtbauer (1999), Anderson and More (1974), and Ljung and Glad (2000).

A Kalman filter is an estimation method for linear systems in state–space form. If the system is nonlinear, linearization around the estimated state vector must be performed, often referred to as **extended Kalman filtering** (**EKF**). A Kalman filter is a state observer that considers the statistical behaviour of process and measurement disturbance

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + v(t). \quad (5.22)$$
$$y(t) = C(t)x(t) + w(t). \quad (5.23)$$

In Equations 5.22 and 5.23, process and measurement equations are given including process disturbance ($v(t)$) and measurement disturbance ($w(t)$). The process states, control signal, and measurement are represented by: $x(t)$, $u(t)$, and $y(t)$, respectively. Assuming time invariant dynamics and forming the observer as a simulation of the above system, the equation for the estimated state vector $\hat{x}(t)$ becomes

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t). \quad (5.24)$$

The estimation quality can be quantified via the innovation in the measurement, $y(t) - C\hat{x}(t)$, which is therefore used as a feedback to the simulation in equation 5.24:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu(t) + K(y(t) - C\hat{x}(t)). \quad (5.25)$$

The choice of the gain matrix K, is a trade–off between how fast the estimation error $x(t) - \hat{x}(t)$ should approach zero and how much the measurement $y(t)$ can be trusted, i.e. large values of K give eigenvalues of $A - KC$ that represent fast dynamics, but in that case errors in the measurement strongly affect the estimation. For small K, on the other hand, measurement errors have less influence whereas

the estimator becomes slow in tracking forced dynamic process variations. If the process and measurement disturbances can be modelled, a stochastic framework can be formulated to handle this trade–off. The gain matrix K is calculated to minimize the estimation error variance, $E\{(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T\}$, yielding a Riccati equation for the solution of K, see Schmidtbauer (1999).

A crucial part with this approach is the formulation of stochastic models of different force mechanisms entering the vehicle. Disturbances originating from road elevation profile, wind speed, road and tyre condition, discrepancy between engine and brake maps and resulting true traction force, measurement errors etc. all have to be modelled.

Chapter 6

Experimental tests

In order to validate further the simulation results for the mode switching cruise controller (MSCC) presented in Paper 3, the controller was implemented into a real truck. The vehicle used for the experiments was a Volvo FH 12 with two axles and a total weight of 18 tonnes. The MSCC was implemented using a **rapid control prototyping** (**RCP**) equipment, i.e. an additional control unit was mounted on the truck. Both the brake pressure and VEB torque were commanded via the 1939 CAN bus. Even though the truck was equipped with an automated manual transmission, which enables automatic control of gear position, this possibility was not used in these first tests. Instead, the driver handled the gear shifting manually, by receiving a gear shift indication based on the algorithm presented in Paper 3. Furthermore, the truck was equipped with two auxiliary sensors:

1. Frame-mounted accelerometer measuring the specific force in the longitudinal direction of the truck.

2. Two temperature sensors (thermocouples of type-K) mounted on the disc according to Figure 6.1.

In Figure 6.1 the positions of the two temperature sensors, T_1 and T_2 are shown. T_1 measures a temperature 3 mm from the disc surface at radius of 170 mm. T_2 represents a temperature 15 mm from the disc surface at a radius of 32 mm, i.e. 3 mm from the spline that mounts the disc to the axle.

Since not all required signals can be directly measured (like road slope and disc temperature) an observer was designed using the method of Kalman filtering. The framework of Kalman filtering provides a recursive estimation method for linear systems on state-space form. Since the filtering procedure is recursive, the need for scheduling of filter gains is reduced compared to the control case. In cases where nonlinearities appear the gain calculation is based on a model that is linearized around the estimated system trajectory. As concluded

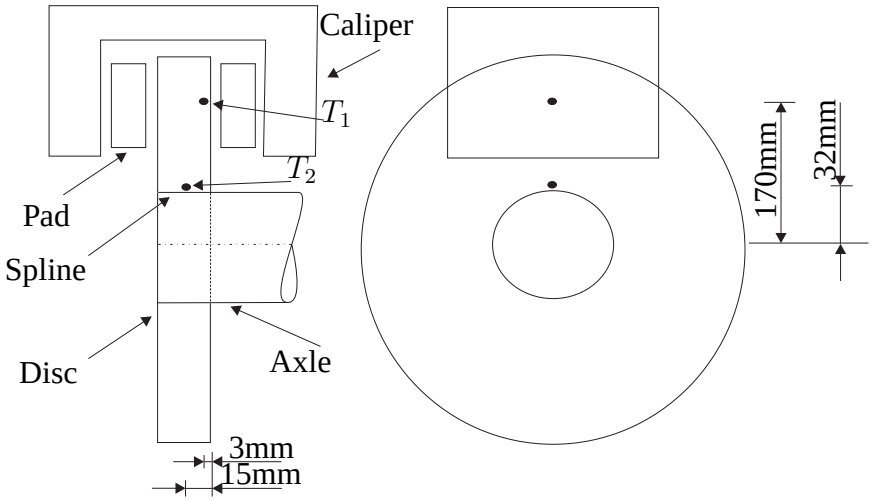


Figure 6.1: Measurement points p_1 and P_2 that correspond to foundation brake temperatures T_1 and T_2.

in earlier chapters, linearization in non–stationary conditions (transients) results in a linearization constant $F(x_0, u_0) \neq 0$. In the controller design the influence of $F(x_0, u_0) \neq 0$ was canceled by using **velocity–based linearization**. The influence of the linearization constant is automatically handled when using an EKF (extended Kalman filter). This can be seen by the fact that any linearization constant is canceled in the expression for the estimation error covariance ($E\{(x(t) - \hat{x}(t))(x(t) - \hat{x}(t))^T\}$), and that the time update $\hat{x}(t|t-1) = f(\hat{x}(t-1|t-1), u(t-1))$ is based on the nonlinear system rather than on the linearized system description which includes $F(x_0, u_0) \neq 0$; see for example Chapter 5 of this thesis, as well as Schmidtbauer (1999), Ljung and Glad (2000), and Anderson and More (1974) for further description.

From an observer point–of–view, the dynamics of the system can be divided into two parts, one for the temperature dynamics (T_1, T_2, $\dot{T}_1$, $\dot{T}_2$) and one for the motion (v, $\dot{v}$, α, $\dot{\alpha}$).

Motion estimator, ***ME***
The motion estimator (**ME**) is based on measurements of specific force, a, and vehicle speed, v. The model consists of a stochastic road slope process and specific force disturbance equation according to

$$\begin{aligned} \dot{v} &= g\sin\alpha - a - x_3, \\ \dot{\alpha} &= -\omega_c\alpha + x_4, \\ \dot{x}_3 &= -x_3\omega_d + \nu_3, \end{aligned}$$

$$\dot{x}_4 \;=\; \nu_4. \tag{6.1}$$

Further details are given in Paper 1. The rates of change are calculated using a simple Euler approximation.

Temperature estimator, ***TE***
For the MSCC, the temperature close to the disc surface (T_1), which is used as reference, is important. Estimation algorithms that function without temperature measurements exist and are sufficient for driver warning systems. A problem with algorithms that do not utilize a measurement of temperature is that they usually are very conservative and hence produce an overestimation of the temperature. This is so, since it is generally not possible to trace the influence of the disc temperature on the generated brake force, i.e. the system is non–observable when using measurements of only pressure and vehicle speed.

However, in the future, it is very likely that FB temperature will be measured. This temperature can be measured in several different positions (disc, pad, and hub) using different types of sensors (thermocouples and IR sensors). Disc surface temperature can be measured using both contact (thermocouples) and non–contact (IR) sensors. The temperature inside pads, disc, and hub can also be measured using thermocouples. Measuring the temperature inside the pad, which is a non–rotating part, is rather easily done. However, since the pad transfers heat very poorly, other techniques to predict fading are more suitable. Measuring the temperature inside the disc, using wireless transfer of data from the rotating disc is a rather usual setup for test vehicles. Production vehicles have not utilized this technique due to high cost and robustness problems with the data transfer. Today, however, there is some development in this area making it a realistic approach for the future.

The main benefit of measuring the temperature inside the disc, compared to e.g. surface measurements, is that the in-disc sensor is better protected from the environment, resulting in a reduced maintenance need. It is, however, not suitable to measure the disc temperature close to the disc surface even if that would be desirable in order to estimate fading. Firstly, the disc surface wears down when braking and therefore the sensor may be damaged. Another problem it that is not desirable to change the sensor when the disc is worn and needs to be replaced. It is therefore likely that the temperature measurement will be performed close to the interface where the disc mounts to the axle. Therefore, T_2, according to Figure 6.1 is considered as a realistic instrumentation.

The equations for the temperature dynamics are given by

$$\begin{aligned} \dot{T}_1 &= q_{\text{in}} - c_1(T_1 - T_2) - c_2(v)(T_1 - T_{\text{amb}}) - c_3(T_1^4 - T_{\text{amb}}^4), \\ \dot{T}_2 &= c_4(T_1 - T_2) - c_5(v)(T_2 - T_{\text{amb}}) - c_6(T_2^4 - T_{\text{amb}}^4), \end{aligned} \tag{6.2}$$

were q_{in} is the FB brake power, c_1 and c_4 are heat conduction constants, $c_2(v)$ and $c_5(v)$ are vehicle speed dependent convection variables, and c_3 and c_6 are radiation constants.

By using these equations, with the FB power $q_{\rm in} = F_{\rm FBd}v$ as a control signal, T_2 as a measurement signal and v as an external signal (from the motion estimator), an EKF was designed. In the Volvo truck there is no measurement of brake force or brake torque. Instead one has to rely on measurements made by a pressure sensor in the brake valve combined with the assumption that the brake force is proportional to the brake pressure, i.e.

$$F_{\rm FB} \;\; = \;\; C_{\rm BG}(P_{\rm FB} - P_{\rm L}), \tag{6.3}$$

where $P_{\rm FB}$ is the measured pressure in the brake valve, $P_{\rm L}$ is the brake application pressure threshold (friction losses etc.), and $C_{\rm BG}$ is the brake gain or brake factor. The pressure loss ($P_{\rm L}$) is constant (short time-scale dynamics) but may increase during the lifetime of the truck (long time-scale). The brake gain ($C_{\rm BG}$), however, varies significantly with brake temperature, application pressure, wheel speed (short time-scale) and also with the brake usage and environmental effects like rust and dirt on the disc (long time-scale). Brake gain variations in the range of $\pm$20% are not unusual and have to be robustly handled, see Granqvist and Rundqvist (2005) for further information. Therefore, in order not to underestimate the brake temperature, $C_{\rm BG}$ is chosen relatively high and $P_{\rm L}$ is chosen relatively low. This should not be confused with the level of conservativeness that is necessary if no measurement of temperature is available. When compared to the open–loop estimation, the measurement of T_2 will reduce the estimation error even if $C_{\rm BG}$ and $P_{\rm L}$ are chosen in a conservative way. The level of improvement will simply depend on the tuning of the Kalman filter, i.e. the trade–off between the quality of the measurement and the model accuracy.

The motion filter and temperature filter were first tuned and tested in simulations and then implemented into a real–time target. In Figure 6.2 and 6.3 experimental results for the temperature estimator and the motion estimator are shown. Estimation of the disc temperature is rather good with a maximum error (difference between measurement and estimation of T_1) of approximately 8% in the cooling phase (after 250s). The error that occurs in the cooling phase (after 260 s) can be explained by process disturbances, such as wind drag and water on the road which are slowly transferred to the measurement point (p_2, see Figure 6.1) in the disc. The numerical values used for $C_{\rm BG}$ and $P_{\rm L}$ were 3800 Nm/bar/axle and 0.3 bar, respectivelly. Estimation of the road slope is also concluded to work good, see also Lingman and Schmidtbauer (2001). After 5 s the initial transient in the slope estimation is stabilized and the truck enters a 5 % uphill slope for approximately 1200 m.

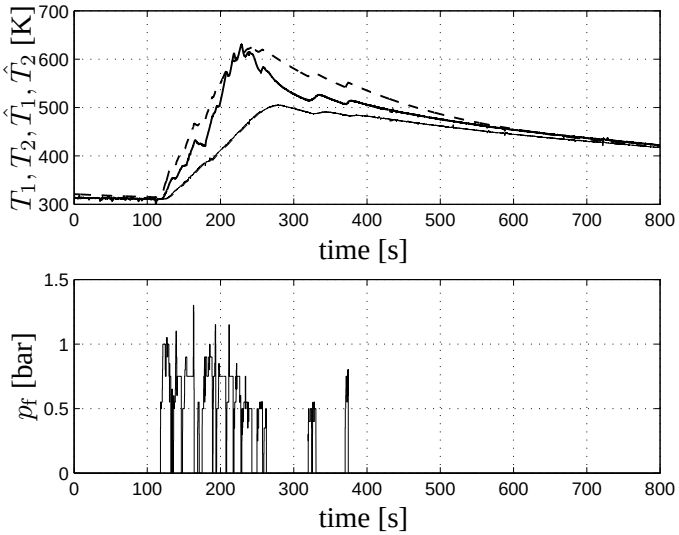


Figure 6.2: Test run for the temperature estimator. Temperatures and brake cylinder pressure are shown. Dashed lines are estimated values. Thick lines represent T_1 and thin lines are for T_1.

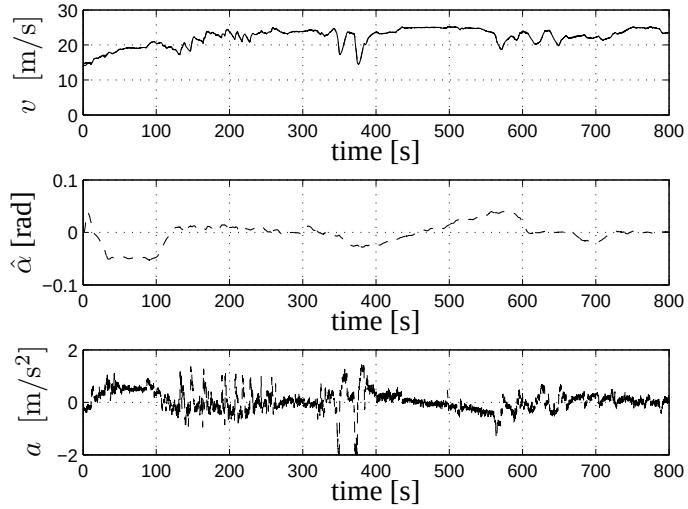


Figure 6.3: Test run for the motion estimator. Estimated road slope and specific force measurement are shown and dashed lines represent estimated values.

Figure 6.4 presents an overview of the TM (temperature mode) closed-loop controller and Figure 6.5 presents a detailed view of the controller and vehicle interfaces. Actuator requests (retardation, engine torque, and VEB torque) are sent via the CAN bus and the brake system ECU (EBS) and engine ECU handles the actuator control. The auxiliary sensors are connected to the RCP equipment via analog wires. The complete controller was first tuned in simulations followed by real tests. All tests were performed at a downhill slope (Kallebäck) of approximately 5% and 1200 m located in Göteborg, Sweden. Test runs were performed with different driver-selected set speeds and with different initial FB temperature. In Figure 6.6 the driver set speed was 21.5 m/s. When the vehicle enters the downhill slope (at 40 s) the velocity mode (VM) is active and the estimated FB temperature ($\hat{T}_1$) is 300 °C . Due to the slope, the vehicle accelerates and the VEB is engaged.

Since the VEB is insufficient to obtain the set speed, the FB are also engaged. The pulsating behaviour of the FBs is due to the fatigue protection, see Chapter 4, where the switch-on value was chosen to 0.82 bar and the switch-off value was chosen to 0.62 bar, in order to avoid small pressure requests with long durations. During the downhill slope, the vehicle speed varies between 22 m/s and 21 m/s, and the FB are applied for 2–3 s every 5-8 s. Since the whole slope is passed without exceeding the FB temperature limit, TM is never engaged in this example, i.e. it was possible to obtain the driver-selected set speed during the complete downhill slope. In Figure 6.8 the same downhill as in Figure 6.6 is shown, except that now also a manual run is shown. The driver was asked to drive downhill in a regular way. Approximately 25 s before entering the downhill slope the driver releases the accelerator pedal and uses the VEB to reduce the vehicle speed to a suitable level. Since the AB force is not enough, FBs are applied in a pulsating way, much resembling what was archived by the MSCC in Figures 6.6 and 6.7.

In Figure 6.7 the driver has selected a slightly higher set speed of 22.2 m/s. In addition, the initial FB temperature is higher (340 °C), causing the TM to be activated at 45 s. Therefore, the vehicle speed is reduced to approximately 18.7 m/s (lowest value), creating an overshoot of approximately 50 °C in FB temperature.

When the slope starts to decrease, the VM is engaged and the vehicle is accelerated to the driver-selected set speed of 22.2 m/s. Note also that in both these examples the maximal VEB torque was restricted to 18% of the maximal torque. This was done in order to compensate for the relatively low vehicle load (no trailer) and the relatively low road grade (5%).

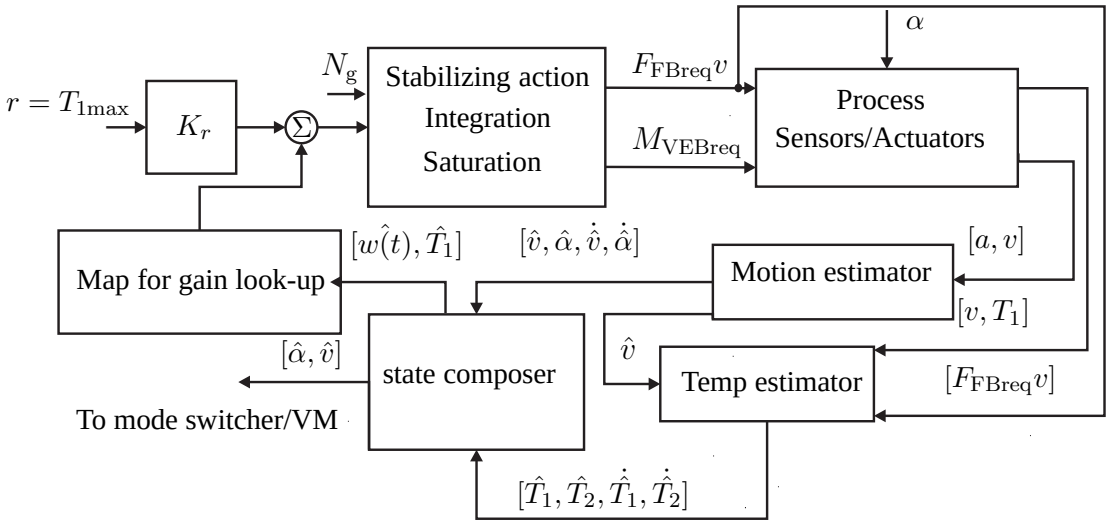


Figure 6.4: Overview of MSCC controller.

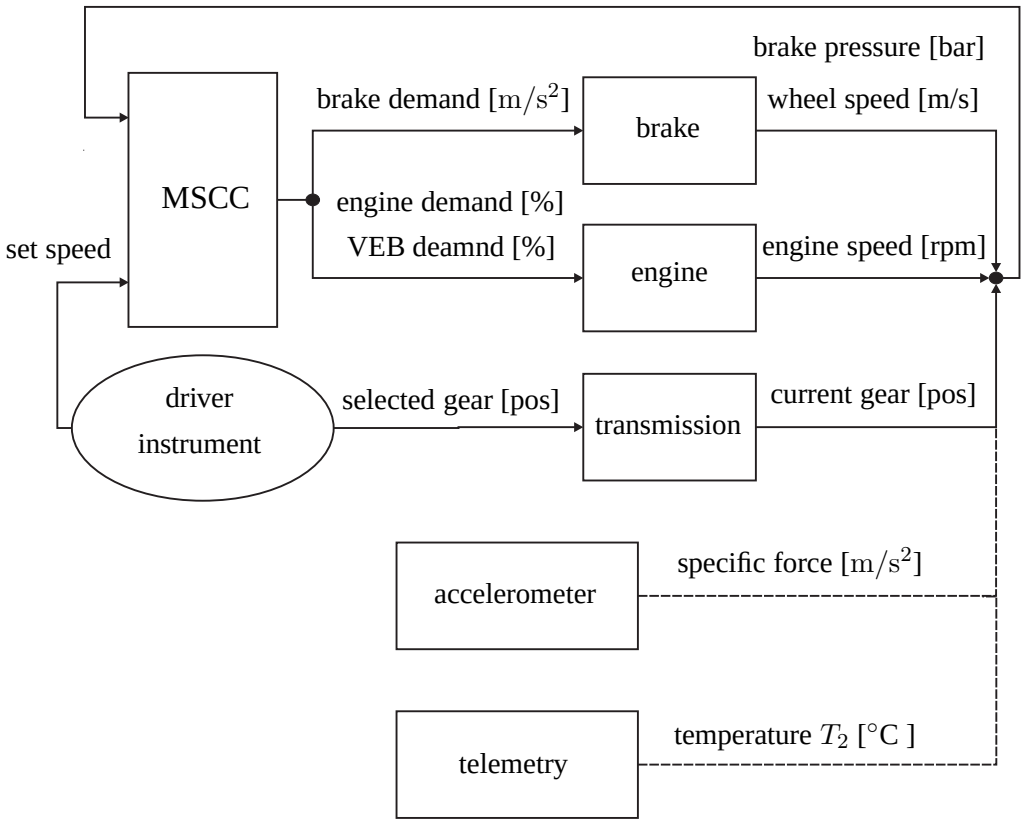


Figure 6.5: MSCC implementation in a vehicle. Solid lines represent CAN bus and dashed lines represent analog signals. VEB and engine torques are given in percentage of maximum torque.

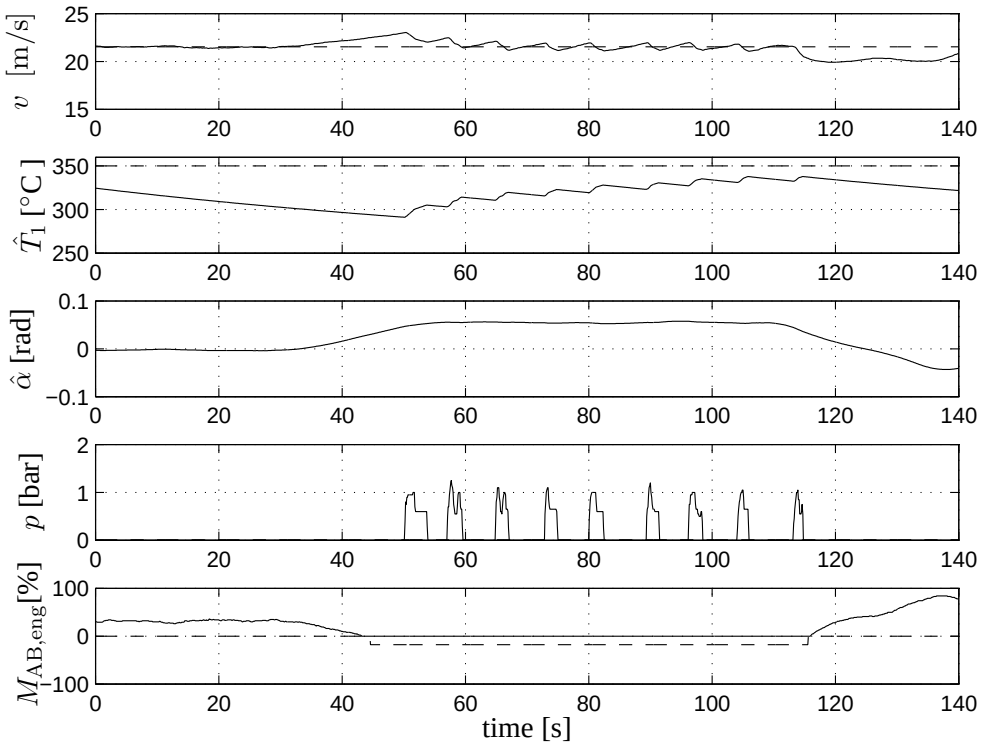


Figure 6.6: Downhill driving in Kallebäck. The driver has engaged the MSCC with a set speed of 21.5 m/s. Due to the relatively low FB temperature, the mode does not change to temperature mode. This enables the driver to keep, in a comfortable way, the set speed even during the whole downhill slope.

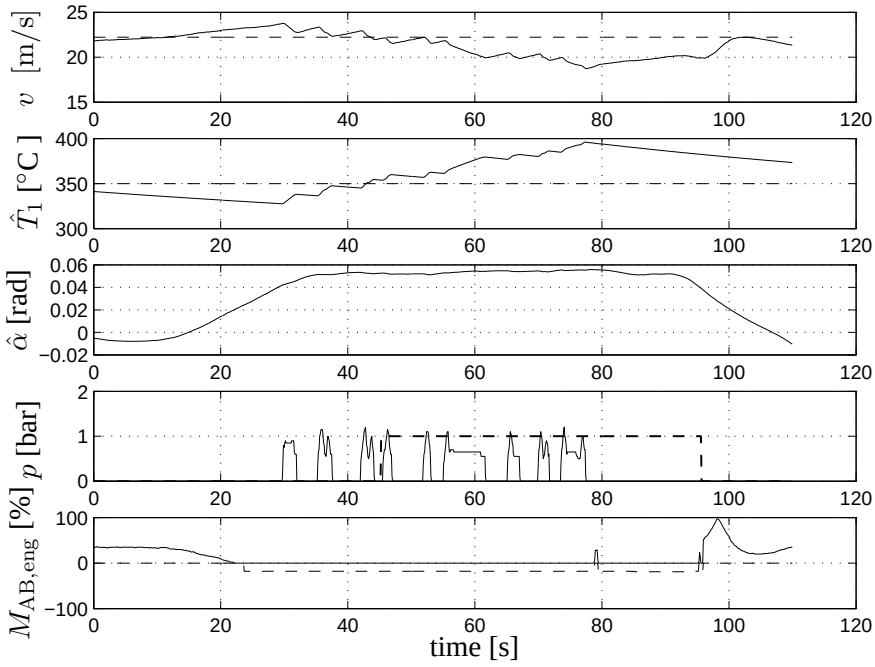


Figure 6.7: Downhill driving in Kallebäck. The driver has engaged the MSCC with a set speed of 22.2 m/s. Due to the relatively high FB temperature, the temperature mode is engaged at 45 s, resulting in a reduction of vehicle speed. The minimum speed is 18.7 m/s. This automatically reduces the vehicle speed to a proper speed based on the current load condition, road slope, and FB temperature.

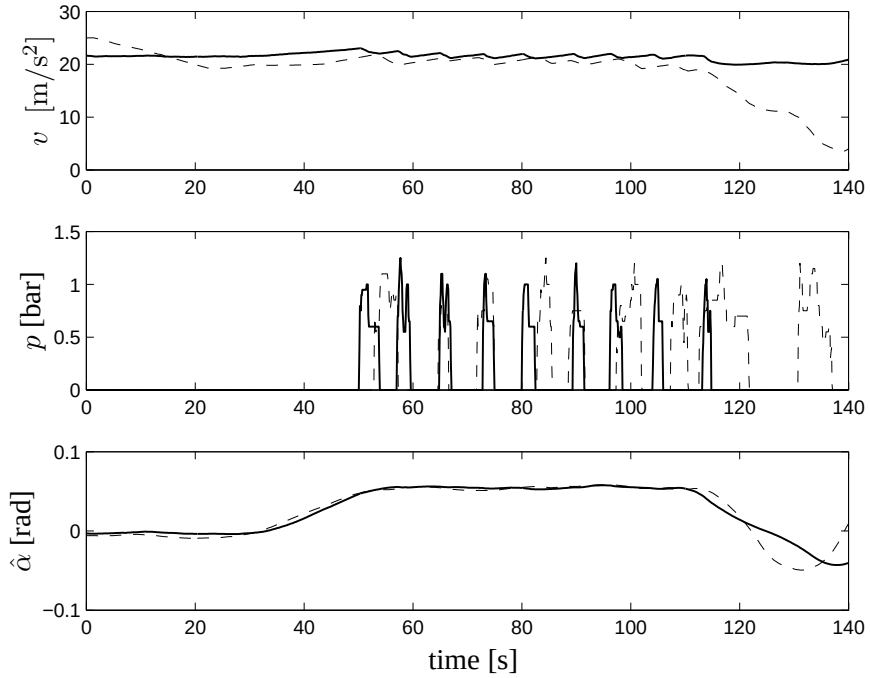


Figure 6.8: Manual and MSCC downhill driving in Kallebäck. Dashed lines represent manual driving and solid lines represent driving with MSCC. The driver reduces the speed before entering the slope. To maintain the speed the driver applies the foundation brakes in a pulsating way, much resembling what was achieved by the MSCC in Figures 6.6 and 6.7.

Chapter 7

Concluding remarks

The field of longitudinal low–power retardation control is a rather unexplored area. Previously, research within longitudinal retardation has mainly focused on the performance of individual actuators. For example, control, durability and performance studies of foundation brakes, hydraulic retarders, and engine retarders, are often found in research papers. Research on the integration of retardation actuators can be found in work published within intelligent highway systems, like the PATH project in California. In, for example, Moklegaard and Stefanopoulou (2000) the blending of brake actuators is performed as in contemporary vehicles that are sold to customers. Furthermore, in the same paper, it is also concluded that, since the natural brake capacity (e.g rolling and air resistance) of vehicles is decreasing, the brake system performance has to be increased using intelligent control strategies.

Intelligent driver–assistance systems like adaptive cruise controllers (**ACC**) have been on the truck market for approximately 2-3 years. In the first versions of ACC, only engine and AB control were used to correct the speed of the vehicle. During 2004 and 2005, ACC systems that also utilize FB, have been introduced on the market. The main application for ACC systems is, of course, not downhill driving but rather highway cruising. Even so, the ACC (and, in particular, the FB handling) has to perform well if the driver prefers to use ACC on downhill slopes. In fact, the issue of over-utilization of FBs is of particular interest for trucks that have, compared to passenger cars, a weak AB performance per mass unit, i.e. FB are applied more frequently for heavy trucks, when compared to passenger cars. Especially the work in Paper 3 can be applied directly to ACC control, as follows: (1) Use ACC as long as fading and fatigue are no threats, (2) When fading and fatigue become threatening, change reference to control the FB temperature instead of speed and following distance, and thereby maximize the brake system performance, (3) When fading and fatigue are no longer problems, apply the original, driver–selected, set speed (or distance, if there is a vehicle in

front).

Other approaches to control the vehicle speed can be found at www.oregon.gov and www.crfproject-eu.org. One common factor with these two approaches is that large parts of the intelligence is built into the infrastructure instead of into the vehicle. At www.crfproject-eu.org a European project called *Safe tunnel* is presented which aims at improving the safety of tunnels and more specifically, avoid fire due to overheated components, e.g. FBs (see Artus et al. (2003)). One benefit of implementing intelligence into the infrastructure is that almost all vehicles can be monitored, i.e. there is no need for the truck owner to purchase and install additional equipment (hardware or software). On the other hand, systems that are mainly integrated into the infrastructure, are expensive to install and maintain, and can therefore only be applied to downhill slopes that are very demanding. Furthermore the accuracy of, for example, equipment that weighs the vehicle in motion is rather low. The weighing system described at www.oregon.gov has an accuracy of $\pm 25\%$ for 80% of the trucks, and worse for the remainder. Therefore, in many situations implementations into on board systems, like presented in this thesis, are preferred.

The approach of using neural networks (NN) and genetic algorithms (GA) to derive retardation strategies is a reliable method, resulting in sensible driving strategies that, even if they may be suboptimal, are significantly better than what a normal driver manually can achieve. The key to obtain good solutions (i.e. NN) using GA lie in the definition of sensible criterion functions, training data, input signals, and output signals. Furthermore, knowledge about reasonable solutions is necessary to determine the quality of the GA solution, i.e. a large amount of engineering is needed.

The final NN obtained performs very robustly on different road profiles. However, it is not robust when exposed to road profiles that have very low resemblance with the training road used. It is therefore very important for the generalization of the NN to design the training road in such way that all realistic road profiles are considered. This is by no means an impossible task since real road profiles are well known and can therefore be taken into consideration, hence making the NN robust for realistic driving conditions. Another, very appealing feature, is that the method of combining NN with GA to generate driving strategies, does not enforce any fundamental requirements on the model, i.e. the states and control signals may be continuous or discontinuous. This enables a very quick and reliable generation of driving strategies. Discontinuities in particular, created severe problems in Paper 4 where a gradient–based optimization method was used whereas when using the GA-based method, no particular problems arose due to discontinuities.

As mentioned earlier, one problem with NN is that they do not in general allow interpretation of internal properties, i.e. the transparency of the control algorithm is rather low. However, there are several options to choose from in order

to improve this situation. One approach, which would be a natural extension of combining NN and GA, would be to use EA in combination with a finite state machine (**FSM**). An FSM parameterisation of the controller would clearly improve the transparency of the algorithm. However, yet another approach was taken in Paper 3. An FSM consisting of only two states and three input signals was designed to integrate the constant speed speed controller and an energy optimal controller. In the downhill cruising situations, energy optimality implies highest possible energy dissipation in order to utilize the brake system as much as possible. In each of the two states, the actuator commands were realized using gain scheduling and linear quadratic control. The approach taken in Paper 3 gives good possibilities (high algorithm transparency) to examine the robustness of the system. Due to the fact that measurement disturbances and process disturbances are relatively well known, simulations were the main tool used to examine robustness. In fact, the framework to verify robustness of gain scheduled controllers is not very well established in theory and gain scheduling is still considered as a relatively ad–hoc controller design method. Furthermore, as discussed in Paper 3, the introduction of actuator saturation and discrete states, make analytical assessment of the complete controller robustness very difficult to achieve.

Model–based state estimation including disturbance modeling has previously not been applied to acquire information of vehicle mass and road slope. In this thesis it is shown that Kalman filtering can be applied for robust and accurate estimation of mass and slope, see Paper 1. Compared to Kalman filtering a drawback with other methods found in the literature is that they make use of a gear shifting phase in which the propulsion force is zero. Such methods require measurements to be made under very difficult circumstances. In addition, the assumption of zero propulsion force during gear shifting is violated in case the vehicle is equipped with a gearbox using two parallel force transmissions. A very popular type of gearbox is the so called **automated manual transmission** (**AMT**). In the AMT, the gear shifting is performed by an electro–pneumatic–mechanical system. A very desirable feature is to reduce the time of zero propulsion force, i.e. to reduce the time duration of the gear shift. This is achieved through the use of sophisticated control, continuously improving the comfort and reducing the duration of the gear shift. This further lowers the possibility to obtain a good estimate of the vehicle mass during the even shorter time of zero propulsion.

The development of control strategies for integrating the different brake actuators of heavy duty vehicles will continue. In this thesis, the objective was to investigate the potentials of improving the performance using existing components only, and the results indicates that much can be done. Hybrid drivelines will come in the future also for heavy duty applications, automatically introducing more brake actuators, making system integration even more important. This development will

probably start with vehicles that are used for distribution purposes, i.e. not heavy long distance vehicles that are the main target of this thesis. In the field of low–power retardation for long distance vehicles there are in particular two areas that deserve special attention for the future. One is the usage of map and positioning systems in order to obtain preview information and another is improving the truck and trailer communication in order to safely utilize all the available brakes of the complete vehicle combination.

Chapter 8

Summary of the Papers

8.1 Paper 1

In Papers 2 and 3 (see below) vehicle mass and current road slope were used as input signals to the controller. Since neither mass nor road slope are directly measurable an estimation must be made. All of the current methods used for real time estimation of mass and slope have one common drawback: they all depend on the gear shifting phase where the propulsion force is zero. This is problematic since, first of all, the propulsion force is non-zero when shifting if the vehicle is equipped with a gearbox that uses two parallel force transmissions. Second, these methods require measurements to be made under very difficult circumstances. Oscillations induced by the gearshift arise both from the flexible driveline and from the play between truck and trailer.

Therefore a more continuous state–space estimation was performed in Paper 1. Kalman filtering was used as a powerful method to obtain accurate estimation of vehicle mass and road slope. First, the problem of estimating the slope when the vehicle mass is known was studied using two different sensor configurations: One where speed is measured and one where both speed and specific force are measured. A filter design principle was derived guaranteeing the estimation error under a worst case situation (assuming first order disturbance dynamics). The simultaneous estimation problem required an Extended Kalman Filter (EKF) design when measuring speed only, whereas the additional specific force case yielded a simple filter structure with a time-varying measurement equation. Additionally the filter needs the current propulsion force which, in the case considered in Paper 1, was calculated from the engine speed and the amount of fuel injected. When the vehicle uses the foundation brakes, the estimates are frozen since varying friction properties would make the braking force unknown. Both sensor configurations were concluded to be robust and accurate by simulation and experimental

field trials.

8.2 Paper 2

Using the powerful techniques of neural networks (NNs) and genetic algorithms (GAs), a brake system controller was designed. First, the problem of blending auxiliary brakes, foundation brakes, and gear for high mean speed in downhill cruising situations was investigated. An optimization problem with constraints, such as vehicle speed and disc temperature, was formulated and solved, resulting in a well performing controller even compared to experienced drivers. It was shown that the mean speed can be improved by controlling the whole brake system. Second, the issue of distributing a required force between auxiliary and foundation brakes in order to minimize the wear cost of pad, disc, and tyres was investigated. The neural network controllers obtained from the optimization procedure significantly outperform the traditional strategy of using non-wear auxiliary brakes in order to minimize pad and disc wear cost. One drawback with the GA-based method is that convergence to optimality is not ensured and that the optimization requires rather high computational power due to the large number of criterion function evaluations. However, the first problem can, to some extent, be compensated by the fact that comparison can, in our application, be made against skilled drivers, i.e. the relative improvement achieved by the optimization can be measured. The second problem cannot, however, be easily overcome and must be considered inherent to the optimization method. Particularly the high computational load causes problems if, for example, parameter studies are to be carried out. Therefore, the generation of optimal trajectories was targeted in Paper 4 (see below). From the implementation point–of–view, the rather low controller transparency (i.e. limited possibility to investigate, in detail, the complete input–output mapping of the controller) must be considered as a problem and the approach taken in Paper 3 offers an improved transparency. It should, however, be noted that there is a large variety of applications where NN have been used in practical control situations.

8.3 Paper 3

Paper 3 uses a divide-and-conquer technique in the design of the controller. First, the control problem was divided into two different modes of operation:

1. Velocity mode controller (**VM**). The purpose of the VM controller was to work as the interface for driver speed demand. It was basically a standard constant speed cruise controller, except for the fact that the VM controller

also commands foundation brakes if required. In standard cruise controllers the foundation brakes are never used due to the risk of heat fading, i.e. a total loss of friction between brake pads and disc.

2. Temperature mode controller (**TM**). When the VM controller has made considerable use of the FBs, due to a relatively high driver speed demand, heat fading might become a threat. In practice this means that the speed of the vehicle should be reduced to allow down gear shifting and by that, produce more auxiliary brake torque. When this situation arises, it is clear that the brake system of the truck is saturated and cannot maintain the driver–selected set speed. The best alternative in this situation is to maximize the brake system performance. This was achieved by switching the set speed reference to a foundation brake temperature reference. This means that the energy dissipation of the foundation brake was maximized at the temperature reference. Combining this with the fact that auxiliary brakes are also utilized to the fullest, an energy optimal strategy ensuring that the vehicle speed is as close as possible to the driver-selected set speed was obtained.

Simplifying, it can be said that switching from VM to TM is based on the foundation brake temperature, i.e. such that a switch to TM takes place when the foundation brake temperature is at the predefined temperature reference. And the decision to switching back to VM is based on the trucks stationary downhill cruising performance i.e. the speed that corresponds to a stationary point of operation that does not violate the boundary conditions of engine speed, velocity, and brake temperature.

Second, the controllers for each mode are designed based on families of linear process models, i.e. gain scheduling. Under the assumption that a **continuously variable transmission** (**CVT**) is used to control the engine speed to the highest possible value (in order to maximize auxiliary brake performance), a continuous, nonlinear (and unstable) process model was obtained for the TM. Two different scheduling strategies were performed on this nonlinear model: one conventional and one using velocity-based linearization. Robustness was verified via simulations by applying realistic measurement and process disturbances. It was also shown that it is possible to include a discrete gear box with good results, even if the controller design happens to be based on a CVT assumption.

8.4 Paper 4

The aim of Paper 4 was to investigate the theoretical limit for the downhill driving performance of a brake system. Optimal control was used to calculate optimal trajectories for high mean speed driving, low maintenance cost driving, and high

mean speed and low maintenance cost driving. The results have then been used in the development and verification of the proposed mode switching controller presented in Paper 3. For the transformation of the original infinite dimensional problem into a finite optimization problem, the transcription method in Betts (2001) was adopted. For the solution of the finite optimization problem the powerful, Matlab compatible, optimization package TOMLAB was used, see for example Holmström and Göran (2003).

Special attention was given to problem function scaling and simplification of the process model. Soundly based scaling and model simplification, allow for relatively low computational effort, which is a very desirable feature for parameter studies that usually require several optimizations to be performed. Implementation of the gearbox does, in its original form, introduce integer states due to the discrete gear positions, i.e. mixed–integer optimization algorithms are required. However, due to the very large optimization problem that arises from the transcription method, mixed–integer solvers were concluded not to work. Instead another approach was taken in which the discontinuous upper bound of the auxiliary brake torque, caused by the gearbox, was approximated with smooth functions. Combined with an implementation of phases, see Betts (2001), this proved to be the best solution to generate optimal trajectories. Also the multi–objective optimization of minimizing the component wear cost and maximizing the vehicle speed was studied using the definition of Pareto optimality.

Bibliography

A. Anderson. A model for engine thermal management simulation. In *Nordic Matlab Conference*, 1995.

J.J. Anderson and J.B. More. *Optimal Filtering*. Prentice-Hall, 1974.

A. Andersson. Transient multi disiplinary engine thermal management simulation. In *Proceedings of VTMS4 Conference, London*, 1999.

S. Artus, S. Hayat, M. Staroswiecki, and V. Cocquempot. A brake disc's temperature estimation module. In *Proceedings of VTMS4 Conference*, London, 2003.

J. Aurell. The world–wide trend on mass and dimensions. In *Proceedings of Technology 2000: Eight International Heavy Vehicle Seminar*, 2000.

J.T. Betts. *Practical Methods for Optimal Control Using Nonlinear Programming*. SIAM Philadelphia, 2001.

R. Boudjemaa, M.G. Cox, A.B. Forbes, and P.M. Harris. *Automatic differentiation techniques and their application in metrology*. SIAM Philadelphia, 2003.

S. Boyd and L. Vandenberghe. Convex optimization. Cambridge University Press, 2004.

I. Das and J. Dennis. Normal-boundary intersection: An alternate method for generating pareto optimal points in multicriteria optimization problems. Technical report, Dept. of Computational & Applied Mathematics, Rice University, Houston, 1996.

M. Druzhinina, L. Moklegaard, and A.G. Stefanopoulou. Compression braking control for heavy-duty vehicles. In *Proceedings of the 5$^{\text{th}}$ International Symposium of Advanced Vehicle Control, AVEC2000*, volume 5, Ann Arbor, 2000.

E.R.G. Eckert and R.M. Drake. *Heat and Mass Transfer*. McGraw–Hill, 1959.

M. Eriksson, M. Bergman, and S. Jacobson. On the nature of tribological contact in automotive brakes. *Wear*, 252:26–36, 2002.

C.A. Floudas. *Nonlinear and Mixed-Integer Optimization*. Oxford University Press, 1995.

P. Frank. Load dependent control of braking force on commercial vehicles. In *Proceedings of the Conference on Mechanical Engineering, Budapest*, volume 1, 1998.

P. Gill, W. Murray, and M. Wright. *Practical Optimization*. Academic Press, London, 1981.

P. Gill, W. Murray, and M.A. Saunders. *User's Guide For SNOPT 5.3: A Fortran Package For Large-Scale Nonlinear Programming*. Technical report, Standford University and University of California, San Diego, 2002.

F. Granqvist and C.J. Rundqvist. Estimation of disc brake performance using empirical modeling–master thesis. Technical report, Chalmers University of Technology and Volvo 3P, 2005.

S. Haykin. *Neural Networks – A Comprehensive Foundation*. IEEE Computer Society Press. Macmillan Collage Publishing Company, 1994.

K. Holmström and A. Göran. User's guide for tomlab 4.0.5. Technical report, Tomlab Optimization, Västerås, Sweden, 2003.

L. Hörmander. *Notions of convexity*. Birkhauser Boston, 1994.

T. Kress and R. Kress. A tool for downhill accident analysis and brake design evaluation. *International Journal of Vehicle Design*, 26:361–373, 2001.

D.J. Leith and W.E. Leithead. Gain–schedule and nonlinear systems: Dynamic analysis by velocity–based linearisation families. *International Journal of Control*, 70:289–317, 1998.

K. Lindemann, E. Petersen, M. Schult, and A. Kom. Ebs and tractor trailer brake compatibility. In *Proceedings of SAE Symposium: Heavy Duty Vehicle Braking and Steering, SP-1307*, 1997.

A. Lindkvist. Konsekvensen av ¨(a)ndrade viktbestämmelser för lastbilar. Technical report, Transportforskningskommissionen (TFK), 1983.

A. Lindkvist and B. Gustavsson. Lastbilskostnader. Technical report, Transportforskningskommissionen (TFK), 1980.

P. Lingman. Integrated retardation control–thesis for the degree of licentiate of engineering. Chalmers University of Technology, 2002.

P. Lingman and M Ekhed. Optimal control strategies for active and semi–active automotive suspensions-master thesis. Technical report, Chalmers University of Technology and Volvo Car Corporation, 1999.

P. Lingman and B Schmidtbauer. Utilizing foundation brakes in a heavy duty vehicle cruise controller application. Technical report, Chalmers University of Technology. Submitted to Vehicle System Dynamics 2004, 2005.

P. Lingman and B. Schmidtbauer. Road slope and vehicle mass esimation using kalman filtering. In *Supp. to Vehicle Systems Dynamics*, volume 37, pages 12–23, 2001.

P. Lingman and M. Wahde. Integrated retardation control using neural networks with genetic algorithms. In *Proceedings of the 6th International Symposium of Advance Vehicle Control, AVEC, Hiroshima*, volume 6, pages 751–756, 2002.

P. Lingman and M. Wahde. Transport and maintenance effective retardation control using neural networks with genetic algorithms. *Journal of Vehicle System Dynamics*, 42:89–107, 2004.

L. Ljung and T. Glad. *Control Theory - Multivariable and Nonlinear Methods*. Taylor and Francis, 2000.

X.-Y. Lu and J.K. Hedric. Modeling of heavy–duty trucks for longitudinal platooning. In *Proceedings of the 6th International Symposium of Advance Vehicle Control, AVEC, Hiroshima*, volume 6, 2002.

H. Min and T. Lambert. Truck driver shortage revisited. *Transportation Journal*, 42:5–16, 2002.

L. Moklegaard and AG Stefanopoulou. Advanced braking methods for longitudinal control of commercial heavy vehicles. Technical report, California PATH Program, Institute of Transportation Studies, University of California, Berkeley CA, 2000.

H. Ohnishi, J. Ishii, M. Kayano, and H. Katayama. A study on road slope estimation for automatic transmission control. *JSAE Review*, 21:322–327, 2000.

L. Palcovics and A. Fries. Intelligent electronic systems in commercial vehicle for enhanced traffic safety. *Vehicle Systems Dynamics*, 35:227–289, 2001.

L. Palcovics, A. Semsey, and E. Gerum. Roll-over prevention system for commercial vehicles - additional sensorless function of the electronic brake system. *Vehicle System Dynamics*, 35:227–289, 2001.

E. Ritzèn. Adaptive vehicle weight estimation. Technical report, Institutionen för systemteknik, Department of Electric Engineering, Linköping University, 1998.

A.H. Saleh. Development of mathematical models to describe the wear process in viscelastic materials. Technical report, Department of Mechanical and Aerospace Enginering, North Carolina State University of Raleigh, 1979.

Y. Sawaragi, H. Nakayama, and T. Tanino. Mathematics in science and engineering. *International Journal of Control*, 176, 1995.

B. Schmidtbauer. *Modellbaserade Reglersystem*. Studentlitteratur, Lund, Sweden, 1999.

P. Templin. Simulation and transient control of diesel engine compression brake– thesis for the degree of licentiate of engineering. Chalmers University of Technology, 2004.

C. Tsai. Models of tyre wear. Technical report, Department of Mechanical Engineering, Carnegi–Mello University Pittsburgh, Pennsylvania, 1973.

M. Wahde. An introduction to adaptive algorithem and intelligent machines, 4^{th} ed. Chalmers University of Technology, 2005.

X. Yao. Evolving artificial neural networks. *Proceedings of the IEEE*, 87(9): 1423–1447, 1999.

Paper 1

Road Slope and Vehicle Mass Estimation Using Kalman Filtering

Peter Lingman and Bengt Schmidtbauer

Proceedings of the 17th International Symposium Dynamics of Vehicles on Road and Tracks, IAVSD 2001, Copenhagen, Denmark

Road Slope and Vehicle Mass Estimation Using Kalman Filtering

Peter Lingman
peter.lingman@volvo.com

Bengt Schmidtbauer
bes@me.chalmers.se

SUMMARY

Kalman filtering is used as a powerful method to obtain accurate estimation of vehicle mass and road slope. First the problem of estimating the slope when the vehicle mass is known is studied using two different sensor configurations. One where speed is measured and one where both speed and specific-force is measured. A filter design principle is derived guaranteeing the estimation error under a worst case situation (when assuming first order dynamics). The simultaneous estimation problem required an Extended Kalman Filter (EKF) design when measuring speed only whereas the additional specific force case yielded a simple filter structure with a time-variant measurement equation. Additionally the filter needs present propulsion force which in our case is calculated form the engine speed and amount of fuel injected. When the vehicle uses the foundation brakes the estimates are frozen since varying friction properties makes the braking force unknown. Both sensor configurations are concluded to be robust and accurate by simulation and experimental field trials.

1. INTRODUCTION

The major parameters in the energy balance of moving vehicles, setting the requirements on the propulsion and retardation systems in the actual driving situation, are current values of vehicle speed, road slope and vehicle mass. In an effort to improve the control of longitudinal vehicle motion it is thus essential to have access to accurate on-line estimates of these parameters.

When it comes to simultaneous estimation of road slope and vehicle mass mainly one approach is seen in the literature. The basic idea is to calculate the vehicle acceleration at two successively points that are relatively close to each other in time (no more than 0.5s) so that gravity, rolling- and air-resistance forces can be considered constant. By stating Newton's second law at the two measurement points the only unknown is the mass of the vehicle. Since the acceleration is calculated from a rather noisy vehicle speed signal (usually the passive ABS wheel speed sensor is used) a clever choice of measurement point has to be made in order to predict the mass successfully. By making one measurement just before a gearshift and the following one during the shift (zero propulsion force) a large difference in acceleration is obtained and hence increasing the accuracy of the mass prediction.

There are some obvious problems with this approach. First of all this method requires measurements to be made under very difficult circumstances. Oscillations arise both from the flexible driveline and from the play between truck and trailer. Both these are induced by the discontinuous propulsion force when changing gear. Furthermore the method can not be used for vehicles equipped with automatic gearshift (powershift) simply because these vehicles don't separate the driveline during a gearshift. Vehicles equipped with automatically shifted manual transmissions are increasing their market shares. This type of vehicle separates the

driveline during a gearshift enabling for the above estimation method to work. It is though a fact that manufactures are trying to minimise the time of no propulsion and thereby decreasing the time left to calculate the vehicle acceleration consequently leading to a low accuracy mass estimate. Another typical situation where the algorithm fails to estimate vehicle mass is when a truck drives up on a relatively flat highway after for example a short refuel. Large uncertainties in the estimate arise simply because the vehicle manages to accelerate with only a few up-shifts.

This paper investigates the possibility of achieving slope and mass estimates using available information on propulsion and brake system characteristics, a vehicle speed measurement and the possible improvement through the addition of a longitudinal accelerometer. Estimation theory (Kalman Filtering) is used to obtain appropriate computation algorithms and make assessments of resulting estimator performance. With this approach a crucial part is the formulation of stochastic models of different force mechanisms entering the vehicle. Disturbances originating from road elevation profile, wind speed, road and tire condition, discrepancy between engine and brake maps and resulting true traction force, measurement errors etc. all have to be modelled.

Some earlier attempts to slope and mass computation, without explicit dynamic error modelling, was done by P.Frank [1], H.Ohnishi et.al [2] and E.Fritzen [3]. In A.G Stefanopoulou et.al. [4] a Model Reference Adaptive System is designed to control vehicle brakes, including and on-line estimator of vehicle mass and road slope utilising a Lyapunov based method.

2. KALMAN FILTERING

A Kalman filter is an estimation method for linear systems in state space form. It is an observer that considers the statistic behaviour of process and measurement disturbances. The system is mathematically formulated on state space form

$$\dot{x} = Ax + Bu + v\,; \quad y = Cx + Du + w \tag{1}$$

where x is the state vector, y is the measurement vector, u is the known system input and v and w being the process and measurement disturbance vectors (white noise). The Kalman filter generating the state vector estimate $\hat{x}$ minimising the (variance of) the estimation error x- $\hat{x}$ is obtained as:

$$\dot{\hat{x}}(t) = A\hat{x}(t) + Bu + K(y(t) - C\hat{x}(t)) \tag{2}$$

The optimal stationary filter coefficient vector is given by

$$K = PC^T R_2^{-1} \tag{3}$$

where P is the stationary estimation error variance (covariance matrix in the vector case) as given by the Riccati equation [1,2]

$$AP + PA^T - PC^T R_2^{-1} CP + R_1 = 0 \tag{4}$$

Parameters R_1 and R_2 are the process disturbance and measurement noise intensities.

3. SLOPE ESTIMATION WITH SPEED MEASUREMENT ONLY

In this section we formulate and investigate different filter formulations for road slope estimation when the vehicle mass is known. We begin by stating the motion equation of the vehicle in the longitudinal direction.

$$m\dot{v} = mg\sin(\alpha) + f_p - f_r \tag{5}$$

where α, f_p, f_r are road slope, propulsion force and retardation force respectively. The propulsion force is the positive engine torque filtered via the transmission of the vehicle and the retardation force consists of forces generated by wheel, auxiliary brakes and deterministic parts of rolling resistance and air drag ($\sim v^2$). Both these forces are considered as known filter input $u(t) = f_p(t) - f_r(t) = f(t)$. Choosing vehicle speed and road slope as states we chose the following simple state space equations

$$\begin{aligned} x_1 = v &\Rightarrow \dot{x}_1 = g x_2 + \frac{1}{m} f(t) + \upsilon_1 \\ x_2 = \alpha &\Rightarrow \dot{x}_2 = \dot{\alpha} = \upsilon_2 \\ y &= x_1 + w \end{aligned} \tag{6}$$

Slope variations are modelled as integrated white noise (random walk) whereas force (acceleration) errors and measurement disturbances are approximated as white noise processes. The intensities of v_1, v_2 and w are a, b and c respectively. The simplicity enables us to find an explicit solution to the stationary Riccati equation:

$$P = \begin{bmatrix} p_{11} & p_{12} \\ p_{12} & p_{22} \end{bmatrix} = \begin{bmatrix} \sqrt{ca + 2cg\sqrt{bc}} & \sqrt{bc} \\ \sqrt{bc} & 1 \Big/ g\sqrt{ab + 2gb\sqrt{bc}} \end{bmatrix} \tag{7}$$

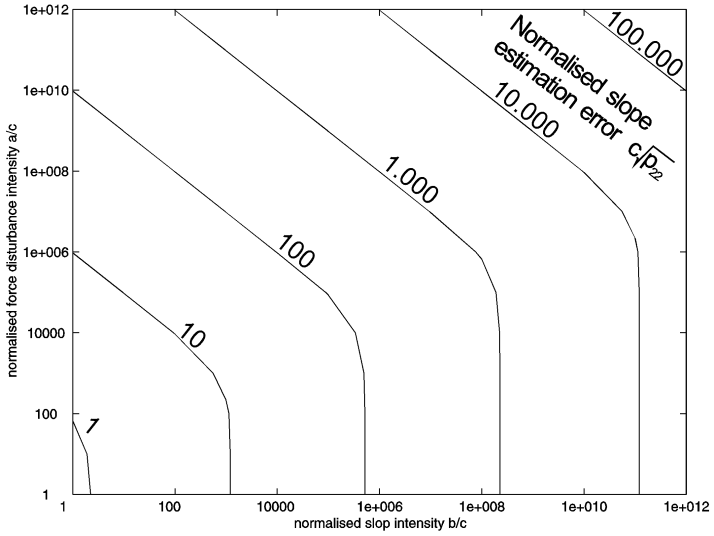


Figure 1. Normalised slope estimation filter performance (standard deviation; simple models)

A parametric study can be done to see how estimation error (filter performance) is affected by process and measurement disturbances, Figure 1. Everything is normalised with the intensity of the measurement noise c. The knees in Figure 1 clearly show areas where improvements in the force estimation does not affect the slope estimation.

3.1. Improved modelling of road slope

Figure 2 shows the power spectral density (PSD) of slope variations in a typical west-swedish road section is, the standard deviation being 0.028 rad. The integrated white noise model of the previous paragraph would correspond to a straight line with negative unit slope. From Figure 2 it is clear that a first order model with a cutoff frequency of f_c=0.002 cyc/m (spatial frequencies!) and a noise intensity of 0.8 $(\text{rad})^2/(\text{cyc/m})$ is a good representation of road slope, e.g. giving the same standard deviation as the experimental curve (dashed line segments in Figure 2).

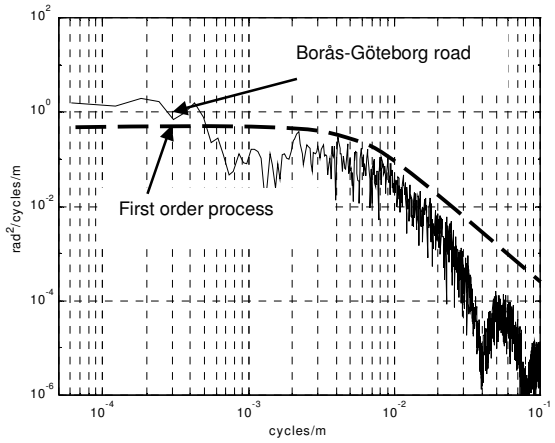


Figure 2. PSD for road section Borås-Göteborg (65 km). The dashed line is first order low pass filtered white noise which is the approximation used here fore the road slope process.

The complete state space model (still 2nd order) then turns to

$$\left.\begin{aligned} x_1 = v \Rightarrow \dot{x}_1 &= g x_2 + \frac{1}{m} f(t) + \upsilon_1 \\ x_2 = \alpha \Rightarrow \dot{x}_2 &= \dot{\alpha} = -\omega_c x_2 + \upsilon_2 \end{aligned}\right\} \Rightarrow A = \begin{bmatrix} 0 & g \\ 0 & -\omega_c \end{bmatrix} \quad \upsilon = \begin{bmatrix} \upsilon_1 \\ \upsilon_2 \end{bmatrix} \tag{8}$$

The Riccati equation for (8) is numerically solved and a parameter study can be done.

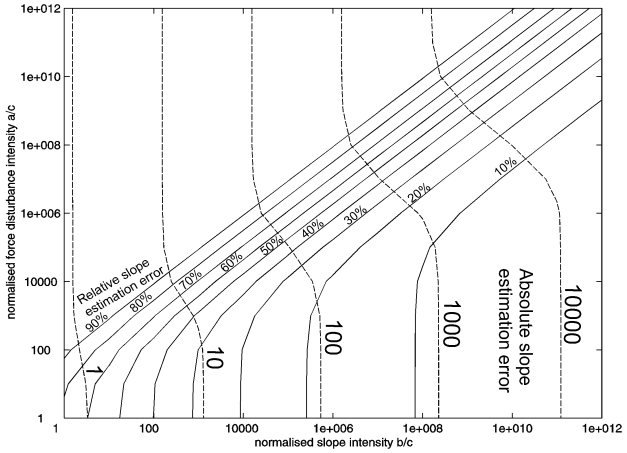


Figure 3. Parameter study of slope estimation error (relative is with respect to slope standard deviation)

In Figure 3 both relative and absolute slope estimation error is shown as function of normalised process disturbance. The relative estimation error is $\sqrt{p_{22}/2b}$ where b is the intensity of the slope process (temporal frequencies). In Figure 3 we see that the area where we like to make our estimation is the lower right part where the relative estimation error is low. The curves representing estimation error now have two knees each. For constant slope rate intensity we see that the estimation error now is affected only in an interval by the force disturbance.

3.2. Dynamic modelling of force disturbance

We now want to improve the unrealistic assumption of white noise force disturbance by replacing it with a first order model. The errors in the engine/brake torque estimation, rolling resistance and air resistance are relatively well known in magnitude but much less known in their frequency content. We have to enlarge our state space formulation with one additional state $x_3 = f_{dist}$.

$$A = \begin{bmatrix} 0 & g & 1/m \\ 0 & -\omega_c & 0 \\ 0 & 0 & -\omega_d \end{bmatrix} \quad Bu = \begin{bmatrix} f(t)/m \\ 0 \\ 0 \end{bmatrix} \quad v = \begin{bmatrix} 0 \\ v_2 \\ v_3 \end{bmatrix} \tag{9}$$

where ω_d is the disturbance force cutoff frequency and $v_3(t)$ is the forcing noise (with intensity d giving a disturbance force variance of $p_d = d/(2\omega_d)$).

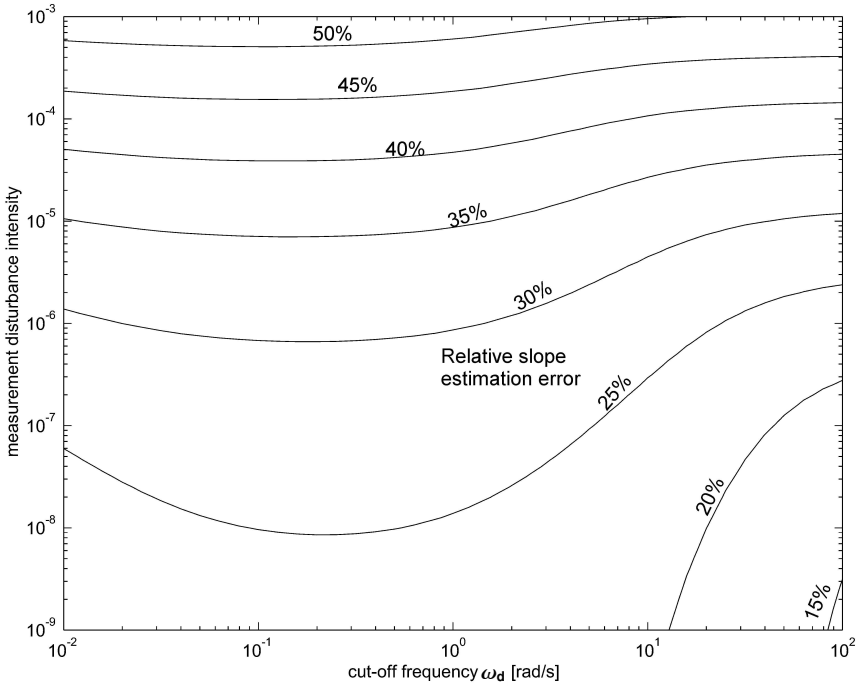


Figure 4. Influence of the force disturbance characteristics on relative estimation error. The cutoff frequency between 10^{-1} and 10^{0} affects the estimation error the most. The road is set to Borås-Göteborg and the vehicle is traveling with constant speed 20 m/s.

It is interesting to investigate the nature of the force disturbance that affects the estimation error the most. With a given disturbance variance p_d the cutoff frequency ω_d is varied with a noise intensity $d = 2\omega_d p_d$ (giving the same variance regardless of cutoff frequency). A parameter study of this is shown in Figure 4. It shows that the cut-off frequency that affects the estimation the most equals approximately the cut-off frequency for the road slope i.e. $\omega_d = \omega_c$. This leads to a non-observable but detectable system, meaning that the non-observable mode is stable and the Kalman theory can still be used (it is difficult for a filter to separate two signals with same assumed frequency content; the system is in reality of 2nd order).

Equal dynamic characteristics in force disturbance as in road slope variations thus represents a worst case for slope estimation. In the absence of knowledge on the real disturbance dynamics it makes sense to design the estimator for this worst case.

3.3. Worst case filter performance

Choose a cut-off frequency for the force disturbance that represents the worst case according to Figure 4 ($\omega_d = \omega_c$) and calculate the corresponding constant Kalman gain matrix K. How well does our observer perform if the real force disturbance has a characteristic (e. g. cutoff frequency) that differs from the one used to calculate K ? The process and observer are described by the following state equations:

$$\dot{x} = Ax + Bu + v \qquad \text{Process}$$

$$\dot{\hat{x}} = A_o\hat{x} + Bu + K_o(y - C\hat{x}) \qquad \text{Observer}$$

We like to calculate the covariance for $\hat{x} - x$ i.e.

$$E\left\{(\hat{x} - x)(\hat{x} - x)^T\right\} = E\left\{\hat{x}\hat{x}^T\right\} + E\left\{xx^T\right\} - 2E\left\{x\hat{x}^T\right\}$$

One way to obtain this is to expand our state vector to include both observer state and process state. Then we can solve the variance matrix equation (a Lyapunov equation) for covariance.

$$\dot{x}^* = \begin{bmatrix}\dot{x}\\ \dot{\hat{x}}\end{bmatrix} = \begin{bmatrix}A & 0\\ 0 & A_o\end{bmatrix}\begin{bmatrix}x\\ \hat{x}\end{bmatrix} + \begin{bmatrix}B\\ B\end{bmatrix}u + \begin{bmatrix}v\\ K_0(Cx + w - C\hat{x})\end{bmatrix} = \begin{bmatrix}A & 0\\ K_oC & A_o - K_oC\end{bmatrix}\begin{bmatrix}x\\ \hat{x}\end{bmatrix} + \begin{bmatrix}B\\ B\end{bmatrix}u + \begin{bmatrix}v\\ K_0w\end{bmatrix} \quad (10)$$

Where A_o, K_c are constant, for a constant cut-off frequency, and w and v are

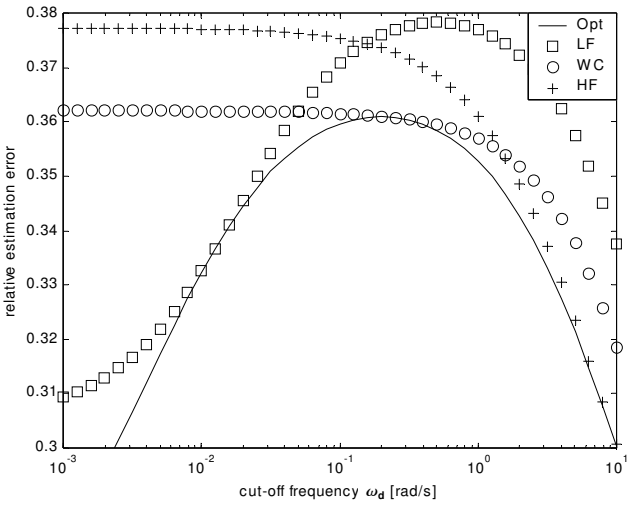


Figure 5. Relative estimation error as function of force disturbance cut-off frequency. The solid line represents optimal filtering i.e. new K is calculated at every ω_d.

uncorrelated process and measurement noises. To calculate the variance of $\hat{x}-x$ we simply have to solve the stationary Lyapunov equation. Figure 5 shows filter performance for 3 different filter design cases, Low Frequency (LF), Worst Case (WC) and High Frequency (HF) (constant disturbance variance in all three cases). In the worst case we see that the observer always keeps the relative estimation error under the maximum error 0.36. Though the performance does not improve for lower frequencies (bias) as for the optimal curve (solid) we always know that the error is equal to or under the maximum error. The HF case shows that designing the filter for higher frequencies deteriorates the observer performance in the low frequency region.

Two fundamental parameters of interest are the variance of the force disturbance (x_3) and the measurement disturbance *c*. How does the measurement disturbance affect the filter performance in a worst-case situation? In Figure 6 a parameter study is presented.

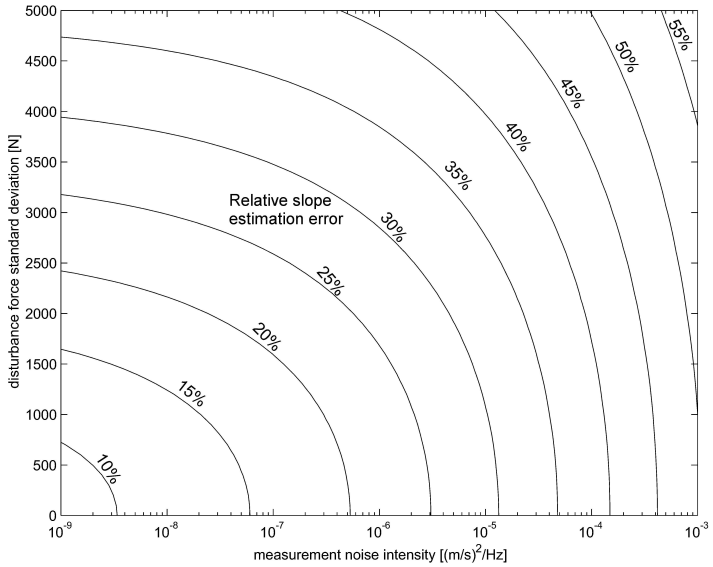


Figure 6. The influence of force disturbance and measurement noise on relative estimation error. Measurement noise as function of force disturbance. Worst-case design.

From the parameter study above we can set a limit for the maximum allowed measurement noise. The relative estimation error is clearly influenced by the measurement noise between 10^{-6} to 10^{-8}. For lower noise levels in the speed measurement the estimation is mainly influenced by the force disturbance. As noted earlier the WC design gave a detectable model and we can therefore reduce our observer by setting $x_{2tot} = x_2 + x_3$. Our system is redefined and the relation between

$$\left.\begin{matrix} x_2 \to g x_2 \\ x_3 \to x_3 / m \end{matrix}\right\} \Rightarrow A = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -\omega_c & 0 \\ 0 & 0 & -\omega_d \end{bmatrix}, \upsilon = \begin{bmatrix} 0 \\ g v_2 \\ v_3 / m \end{bmatrix}$$

$\hat{x}_{2tot}$ and $\hat{x}_2$ is proportional to the ratio of the second element of the full sized observer and the second element of the reduced observer.
By setting $x_{2tot} = x_2 + x_3$ we get

$$A_r = \begin{bmatrix} 0 & 1 \\ 0 & -\omega_c \end{bmatrix} \upsilon = \begin{bmatrix} 0 \\ g\upsilon_2 + \upsilon_3/m \end{bmatrix} \quad \text{and} \quad \hat{x}_2 = \frac{K_2}{K_{2tot}} \hat{x}_{2tot}$$

4. ESTIMATION OF SLOPE AND MASS WITH SPEED MEASUREMENT ONLY

To be able to simultaneously estimate vehicle mass and road slope we have to enlarge our state space model (section 3.2) with at least one additional state for the vehicle mass. This gives us the following equations

$$\dot{v} = g\alpha + \frac{f(t)}{m} + \frac{f_{dist}}{m} \qquad \begin{aligned} \dot{v} &= \dot{x}_1 = gx_2 + \frac{f(t)}{x_3} + \frac{x_4}{x_3} \\ \dot{\alpha} &= \dot{x}_2 = -\omega_c x_2 + \upsilon_2 \\ \dot{m} &= \dot{x}_3 = \upsilon_3 \\ \dot{f}_{dist} &= \dot{x}_4 = -\omega_d x_4 + \upsilon_4 \end{aligned} \tag{11}$$

This is a non-linear state-space equation system and we can no longer use standard linear Kalman filtering. Instead Extended Kalman Filtering (EKF) is used [5].

Linear/non-linear equations $\begin{aligned} \dot{x} &= f(x,t) + \upsilon \\ y &= g(x,t) + w \end{aligned}$

In our case $f(x,t)$ is a non-linear function and $g(x,t)$ is linear. The basic idea with EKF is to linearise the model around $\hat{x}$ and then use the time variant linear Kalman filtering equations for filter coefficient *K* and estimation error covariance *P*.

It is preferable to work with difference equations instead of differential equations in real-time applications. The state-space formulation is therefore discretisized with an Euler approximation ($\dot{x} = (x(t+1) - x(t))/h$, h being the time increment):

$$\begin{aligned} x_1(t+1) &= x_1 + hgx_2 + \frac{hf(t)}{x_3} + \frac{hx_4}{x_3} = f_1 \\ x_2(t+1) &= (1 - h\omega_c)x_2 + h\upsilon_2 = f_2 + h\upsilon_2 \\ x_3(t+1) &= x_3 + h\upsilon_3 = f_3 + h\upsilon_3 \\ x_4(t+1) &= (1 - h\omega_d)x_4 + h\upsilon_4 = f_4 + h\upsilon_4 \end{aligned} \tag{12}$$

As mentioned the next step in deriving a filter is to linearise the non-linear equations around the estimated state vector $\hat{x}$.

Linearization $\begin{aligned} \dot{\delta x} &= F\delta x + \upsilon \\ y &= g(x,t) + w \end{aligned}$ where $F = \frac{\partial f(x,t)}{\partial x}$ at $\hat{x}$ yielding:

$$\begin{bmatrix} \delta x_{1_{t+1}} \\ \delta x_{2_{t+1}} \\ \delta x_{3_{t+1}} \\ \delta x_{4_{t+1}} \end{bmatrix} = \begin{bmatrix} 1 & hg & -\frac{h(f(t) - \hat{x}_{4_t})}{\hat{x}_{3_t}^2} & \frac{h}{\hat{x}_{3_t}} \\ 0 & 1 - h\omega_c & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 - h\omega_d \end{bmatrix} \begin{bmatrix} \delta x_{1_t} \\ \delta x_{2_t} \\ \delta x_{3_t} \\ \delta x_{4_t} \end{bmatrix} + \begin{bmatrix} 0 \\ h\upsilon_2 \\ h\upsilon_3 \\ h\upsilon_4 \end{bmatrix}, [y] = [C] \begin{bmatrix} \delta x_{1_t} \\ \delta x_{2_t} \\ \delta x_{3_t} \\ \delta x_{4_t} \end{bmatrix} + [w] \tag{13}$$

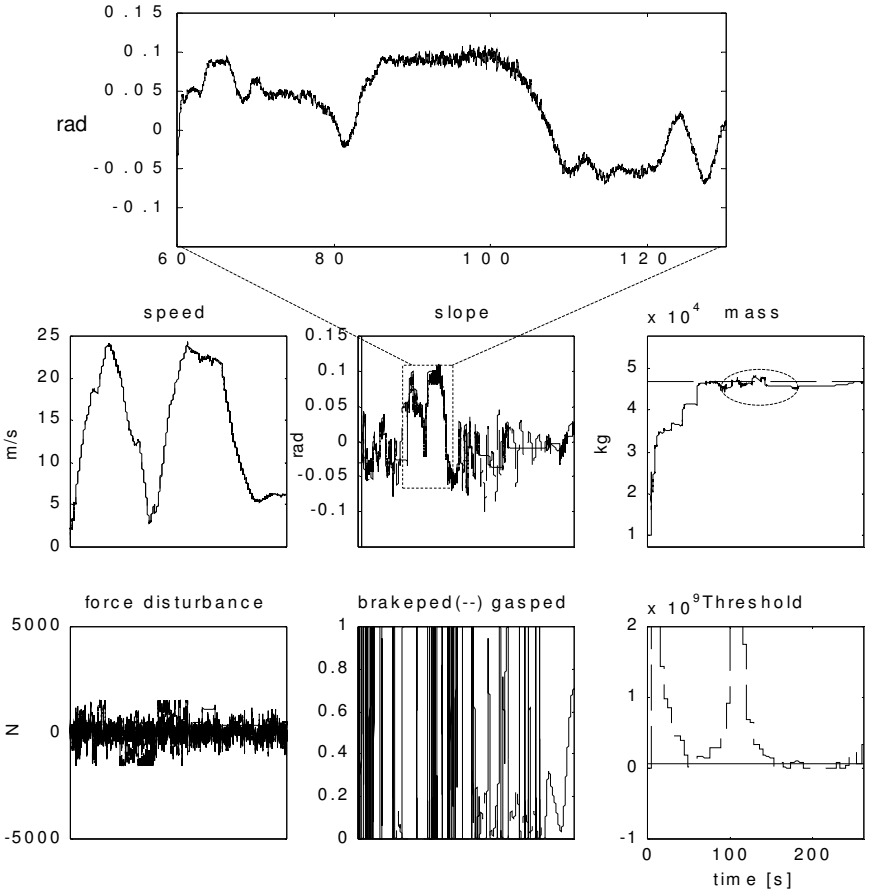


Figure 7. Simulation of estimator performance. Dashed lines are showing real parameter values and there estimates are marked in solid. In the marked area we have poor excitation of the system and the mass estimate would drift without thresholding. Thresholding simply turn the estimator on and off depending on the excitation conditions (to avoid covariance drift). Note that the slope can still be estimated in a parallel filter (based on equation 9) where the mass is set.

4.1. Simulation results for slope and mass estimation

Initial to practical field-tests the filter has been tested in a software environment. The model used here is a complete vehicle model built in Matlab/Simulink. In the simulation the vehicle travels on the measured Borås-Göteborg road profile.

Results from one simulation are shown in Figure 7. When the driver engages the foundation brakes the estimation process is temporarily frozen (uncertain friction between pad and disc; not the case for auxiliary brakes). The simulations indicate that the road slope and vehicle mass estimation algorithm using Kalman filtering gives reasonable performance (cf. also practical field results presented in section 7).

5. SLOPE ESTIMATION WITH SPECIFIC FORCE MEASUREMENT

In the near past accelerometers using the latest MEMS technology have gained interest in the automotive industry. Low cost and functionality improvement of various automotive systems makes accelerometers interesting sensors. One system that uses this kind of sensor is the ESP system (Electronic Stability Program) in heavy-duty vehicles. The focus in ESP systems is on the lateral and angular acceleration whereas here we are interested in the longitudinal acceleration or more correctly the longitudinal specific-force.

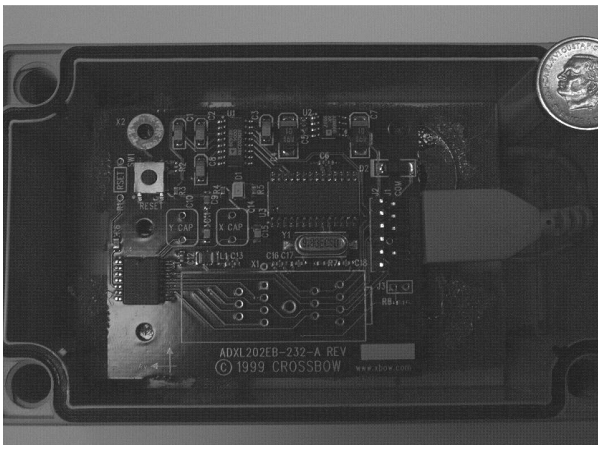


Figure 8. Accelerometer ADXL202EB-232-A mounted in a water tight case.

In the longitudinal direction the accelerometer measures the specific force with a dynamic error that will be regarded as a state variable $x_3(t)$, modelled by a first order process with cutoff frequency ω_d.

$$a(t) = g\sin(\alpha) - \dot{v} \approx g\alpha - \dot{v} + x_3 \tag{14}$$

The complete state model bears strong resemblance to the model of section 3.2:

$$\left.\begin{aligned} x_1 = v &\Rightarrow \dot{x}_1 = gx_2 - a(t) + x_3 \\ x_2 = \alpha &\Rightarrow \dot{x}_2 = -x_2\omega_c + \upsilon_2 \\ x_3 = a_d &\Rightarrow \dot{x}_3 = -x_3\omega_d + \upsilon_3 \end{aligned}\right\} \Rightarrow A = \begin{bmatrix} 0 & g & 1 \\ 0 & -\omega_c & 0 \\ 0 & 0 & -\omega_d \end{bmatrix} \quad \upsilon = \begin{bmatrix} 0 \\ \upsilon_2 \\ \upsilon_3 \end{bmatrix} \quad Bu = \begin{bmatrix} -a(t) \\ 0 \\ 0 \end{bmatrix} \quad C = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T \tag{15}$$

the only difference being that the control force measurement divided with vehicle mass is here replaced with an accelerometer measurement.

Since the control force error can be expected to be considerably larger than the accelerometer error it is reasonable to believe that the slope estimation accuracy increases when using an accelerometer (lower values on the *y*-axis in Figure3).

6. MASS AND SLOPE ESTIMATION WITH SPECIFIC FORCE MEASUREMENT

Using the accelerometer signal $a(t)$, the slope estimation is done without any coupling to vehicle mass, as shown in the previous section. Vehicle mass can then be estimated separately using calculated values of the control force (cf. section 3)

$f(t)$ from the simple relation $a(t) = -f(t)/m$. This means that when utilising an accelerometer measurement we can clearly separate the simultaneous estimation problem into two different filters, one "kinematic filter" (no equations of motion involved) for the road slope and one "dynamic filter" for the mass. The formulation of the dynamic filter used to estimate the vehicle mass is shown below:

$$\left.\begin{array}{l} x_1 = m \Rightarrow \dot{x}_1 = v_1 \\ y = f(t) = (a(t) - \hat{x}_3)x_1 + w \end{array}\right\} \Rightarrow \quad A = 0 \quad Bu = [0] \quad C = [(a(t) - \hat{x}_3)] \quad v = [v_1]$$

where $\hat{x}_3$ is the estimated accelerometer error calculated in the kinematic filter.

7. EXPERIMENTAL RESULTS

The estimation algorithms have been evaluated in field tests. A Volvo FH12 with drawbar trailer and a gear-tronic transmission was used. Required signals where sampled from the CAN SAE J1939 and J1576 bus and transmission ECU. The experiment was not done in real-time mainly to avoid the extra work required to implement software in an ECU and all filters are of course causal. In the below Figure10, a normal acceleration from standstill is shown. The actual

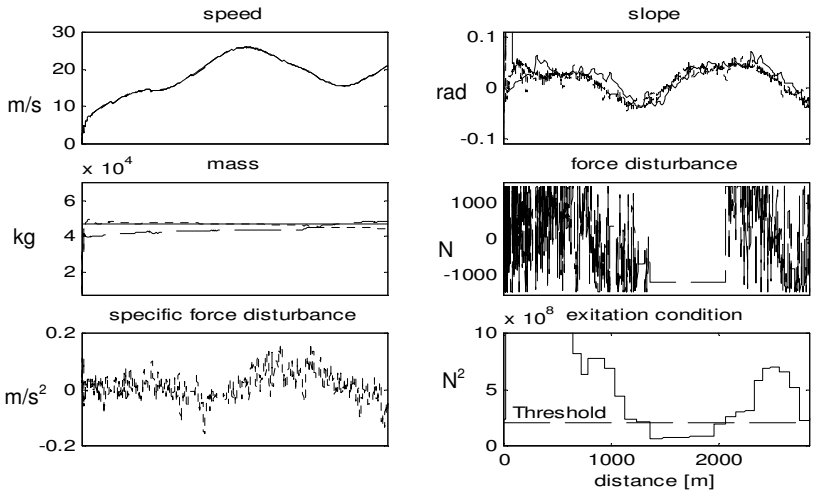


Figure 9. Experimental results for both sensor configurations. Notice that when the threshold work the slope is estimated with the filter derived in section 3 using the mass estimated with the filter in section 4. ——— reference ..... specific force and — — — Speed measurement.

vehicle weight was 47 000 kg and the vehicle passed a ridge after 800 meters which is seen in the slope estimation. Both road slope and vehicle mass approach there correct values respectively. The road slope reference was measured with an atmospheric pressure sensor. Notice that in the speed measurement only case two filters run in parallel to enable slope estimation even when the excitation was considered low.

Varying the process and measurement noise parameters is one way to study the robustness of the filter and this is shown below, Figure10.

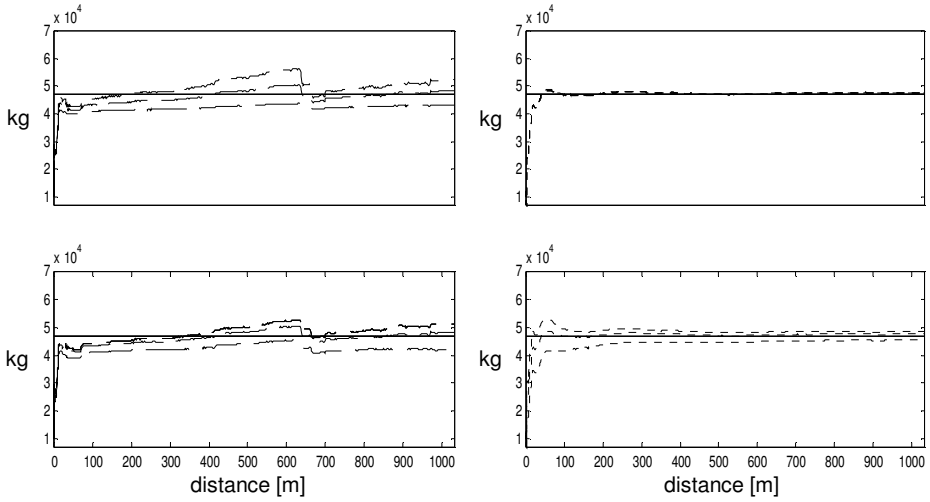


Figure 10. Left plots are for speed measurement only. The upper has constant measurement noise 1E-5 and the process noise is (1E3 4E3 8E3) and the lower is for constant process noise 4E3 and the measurement noise is (7E-6 1E-5 3E-5). The plots on the right are for the specific force configuration and the upper has measurement noise 1E8 and process noise (1E1 1E2 1E3) and the lower has process noise 1E2 and measurement noise (6E7 1E8 2E8).

Both filters perform well for these parameter variations. It is though clear that the additional specific-force measurement improves the somewhat sluggish speed measurement case.

8. FURTHER RESEARCH – ADAPTIVE FEATURES

Several parameters entering the estimation process are depending on varying environmental conditions, e. g. the road profile (the slope characteristics for roads in the alps are substantially different from those in the flatlands!), vehicle conditions (e. g. brakes; cf. section 4.1), etc. One subject for further research will be to investigate system performance under environmentally changing conditions using adaptive schemes for on-line tracking of relevant parameter values, e. g. the magnitude (variance) and time/space behavior (cutoff frequency) of road slope variations.

9. CONCLUSIONS

It has been shown through simulations and practical tests that Kalman filtering is a viable method for on-line estimation of road slope and vehicle mass in heavy duty vehicles. The method is robust and generally applicable and does not have any restrictions concerning different engine, transmission or brake system types, as long as the torque/force balance can be modelled with reasonable accuracy. In case an accelerometer is used as additional (to speed) sensor no torque/force-model is needed for slope estimation.

10. REFERENCES

[1] Frank. P, *Load dependent Control of Braking Force on Commercial Vehicles.* Conference on Mechanical Engineering 1 1998 Budapest, Hungary

[2] H.Ohninshi, J.Ishii, M.Katyano et.al, *A study on road slope estimation for automatic transmission control.* JSAE Review 21 pp 235-240, 2000

[3] Ritzén E. *Adaptive Vehicle Weight Estimation*. Institutionen för systemteknik, Department of Electric Engineering, Linköping University,1998.

[4] M.Druzhinina, L.Moklegaard and A.G. Stefanopoulou, *Compression Braking Control for Heavy-Duty Vehicles.* University of California, Santa Barbara, 2000

[5] Anderson B.D.O, More J.B, *Optimal Filtering, Information and System Science Series.* Prentice-Hall, University of Newcastle, New South Wales, Australia, 1979

[6] Schmidtbauer B. *Modellbaserade reglersystem*, Studentlitteratur 1999.

Paper 2

Transport and Maintenance Effective Retardation Control Using Neural Networks with Genetic Algorithms

Peter Lingman and Mattias Wahde

Vehicle System Dynamics, vol 42, pp 89–107, 2004

Comment: A shorter version of this paper was presented at the 6th International Symposium on Advanced Vehicle Control, AVEC 2002, Hiroshima, Japan. It was, out of 150 paper, nominated for best paper award together with 12 other papers.

Transport and Maintenance Effective Retardation Control using Neural Networks with Genetic Algorithms

PETER LINGMAN[*,†,1] AND MATTIAS WAHDE[*]

SUMMARY

A brake system controller is designed using the powerful techniques of neural networks and genetic algorithms. First, the problem of coordinating auxiliary brakes, foundation brakes, and gear for high transport effectiveness in down hill cruising situations is investigated. An optimization problem with constraints such as vehicle speed and disc temperature is formulated and solved, resulting in a well performing controller even compared to experienced drivers. Second, the issue of distributing a required force between auxiliary and foundation brakes in order to minimize the maintenance cost is investigated. The neural network controllers obtained from the optimization procedure significantly outperform the traditional strategy of using non-wear auxiliary brakes in order to minimize pad and disc wear cost. The performance of the brake system can be improved by controlling the whole brake system including gear.

Keywords: auxiliary brake, gear shift, cruising, wear, genetic algorithms, neural networks.

1. INTRODUCTION

Cruising downhill with a heavy duty vehicle requires several actions from the driver. First of all a proper set speed must be decided upon. To obtain and keep this set speed the driver has to change gear and engage both foundation (**FB**) and auxiliary brakes (**AB**), i.e. Volvo Engine Brake (**VEB**) and Compact Hydrodynamical Retarder (**CR**).

A problem is that the driver is given almost no feedback regarding vehicle mass, brake disc temperature and other important vehicle and environment states. Accidents have occurred where a driver has chosen an excessive speed combined with a bad distribution of retardation force between ABs and FBs, resulting in overheating of both the disc brakes (known as fading) and the ABs (cooling system saturation), see [2], for an example of a fading accident.

Other more conservative drivers choose a low set speed yielding a low utilization

[1]Address correspondance to: Peter Lingman, Department of chassis and vehicle dynamics, Volvo 3P, Gropegårdsgatan, 405 08 Göteborg, SWEDEN. Phone +46 31 327 4342.
e–mail: peter.lingman@volvo.com
[†]Department of chassis and vehicle dynamics, Volvo 3P
[*]Div. of Mechatronics, Department of Machine and Vehicle Systems, Chalmers University of Technology, 412 96 Göteborg, SWEDEN.

of the capacity of the brake system and perhaps also low mean speed depending on the driving cycle. The third type of driver (most common!) is the very experienced driver with good knowledge of both vehicle and environment. By observing engine speed, coolant temperature etc., such a driver uses the AB system to keep the vehicle speed constant but never uses the FB for long periods of time in order to avoid fading. In Fig. 11 the maximum stationary speed is shown as a function of slope. By choosing gear and level of the ABs, the driver finds the maximum stationary speed for downhill cruising. This is of course a difficult driver task, especially if also FBs are considered. Thus there is a need to design an integrated retardation control system for high mean speed (high transport effectiveness) while keeping the system within certain boundaries (safety and law).

Another important aspect of strategies for optimal retardation force distribution is how the tyre, pad, and disc wear are affected. FBs distribute the total retardation force demand on all vehicle axles whereas ABs use the drive axle only, see Fig. 1. The standard strategy is to utilize ABs (known as *non-wear brakes*) in order to save brake pads and discs. However, some drivers have noticed a very high drive tyre wear causing high maintenance cost when using this strategy. Therefore the trade-off between tyre, pad, and disc wear cost has to be investigated and considered in a criterion function for optimal retardation control. See also [4] for further discussions on the importance of integrating the whole retardation system of a heavy vehicle.

This paper investigates the possibility of achieving an integrated retardation controller for gear shifting and blending of different retardation actuators in a heavy duty vehicle. This is done in two different cases. **Case 1** focuses on maximum utilization (high mean speed, $\overline{v}$) of the brake system and **case 2** focuses on minimizing the pad, disc, and tyre wear cost in a constant speed keeping situation. In **case 1**, the scenario is that the vehicle enters a descent and the driver decides upon a maximum vehicle speed. The controller guides the vehicle downhill without exceeding the maximum speed and other constraints (fading, engine speed etc). **Case 2** focuses on a common type of constant speed controller requesting a retardation force to keep the vehicle speed constant on a downhill slope. The main question here is how the required retardation force should be split between ABs and FBs in order to minimize wear cost of tyres, pads, and discs. In both cases, the parametrization of the brake system controller is achieved by a neural network (NN) optimized with a genetic algorithm (GA).

2. THE RETARDATION SYSTEM

In Fig. 1, the main components of a Volvo heavy duty vehicle retardation system are presented. The FB power is distributed on all wheels (including trailer) whereas the ABs work only on the drive axle. Pressurized air, generated in the compressor and

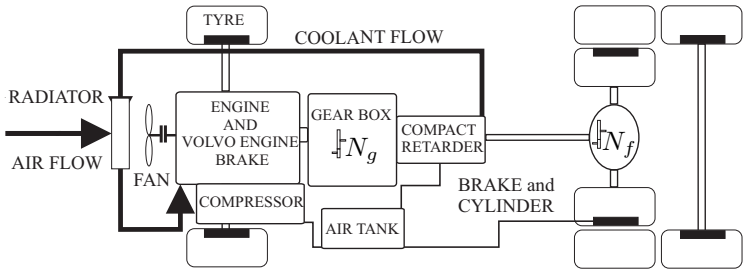


Fig. 1. Volvo retardation system

stored in air tanks, is used as FB energy source and as control medium for both the VEB and CR. The retarder cooling system is connected to the engine cooling system making it a closed loop system where the generated heat is handled in the main engine cooler (radiator).

2.1. Vehicle model

In order to design and evaluate an integrated retardation controller a mathematical vehicle model was built. Since the model is used in an optimization routine requiring many iterations, it is very important to have a well balanced model that describes the main dynamics of the system without being excessive in complexity. The main equations are presented below.

Longitudinal motion equation:

$$m\dot{v} = F_{\text{drive}} - F_{\text{air}} - F_{\text{roll}} - F_{\text{grade}} - F_{\text{aux}} - F_{\text{found}} \tag{1}$$

F_{drive} is the propulsion force, F_{air}, F_{roll}, and F_{grade} are resistance forces and F_{aux} and F_{found} are AB and FB forces, respectively.

Foundation brake dynamics
First order dynamics is assumed and $T_f = 0.4s$. T_1 and T_2 represent the temperature dynamics of the disc brake, discretized into two masses.

$$\dot{F}_{\text{found}} = \frac{1}{T_f}(F_{\text{req}} - F_{\text{found}}) \tag{2}$$

$$\dot{T}_1 = q_{\text{in}} - c_1(T_1 - T_2) - c_2(v)(T_1 - T_{\text{amb}}) - c_3(T_1^4 - T_{\text{amb}}^4) \tag{3}$$

$$\dot{T}_2 = c_4(T_1 - T_2) - c_5(v)(T_2 - T_{\text{amb}}) - c_6(T_2^4 - T_{\text{amb}}^4) \tag{4}$$

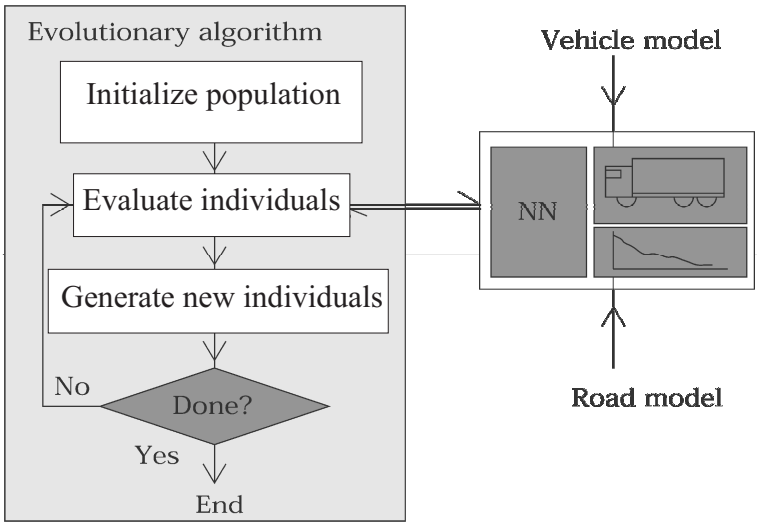


Fig. 2. The optimization loop

q_{in} is the brake force power, c_1 and c_4 are heat conduction constants, $c_2(v)$ and $c_5(v)$ are vehicle speed dependent convection variables, and c_3 and c_6 are radiation constants.

Auxiliary brake dynamics
First order dynamics is assumed with $T_{veb} = 0.3s$ and $T_{cr} = 0.5s$.

$$\dot{F}_{\mathrm{veb}} = \frac{1}{T_{\mathrm{veb}}}(F_{\mathrm{req}} - F_{\mathrm{veb}})\frac{N_g N_f}{R_w} \tag{5}$$

$$\dot{F}_{\mathrm{cr}} = \frac{1}{T_{\mathrm{cr}}}(F_{\mathrm{req}} - F_{\mathrm{cr}})\frac{N_f}{R_w} \tag{6}$$

F_{veb} and F_{cr} are the VEB and CR retardation forces, N_g is the gear ratio of the current gear position, N_f is the gear ratio of the rear axle final gear, and R_w is the wheel radius.

Cooling system
The radiator is modelled according to [1]. From measurements it can be concluded that approximately 40% of the VEB power and 100% of the CR power transfer to the coolant, respectively. In a radiator, as for all heat exchangers, the flow rate of the medium is used to control the cooling capacity. In a heavy duty truck the system is designed so that the coolant flow is directly proportional to the engine speed (v_E) and

the air flow is a function of vehicle speed, fan speed, and air temperature. The fan speed is a function of v_E and the slip in the viscous coupling, used to control the fan speed relative to v_E. Here the air density variation due to temperature variation has been omitted.

$$\text{Coolant flow} = c\, v_E \tag{7}$$

$$\text{Air flow} = f(v, \text{Fan speed}) \tag{8}$$

$$\text{Fan speed} = f(v_E, \text{slip}) \tag{9}$$

Wear dynamics
The models used for pad and tyre wear are shown below. As can be seen, the pad wear rate varies strongly with the temperature of disc whereas the tyre wear rate is only a function of the torque acting on the wheel. τ_{tyre} represents the torque on each wheel.

$$\dot{S}_{\text{pad}} = q_{\text{in}} S_0 e^{cT_1^{k_0}} \tag{10}$$

$$\dot{S}_{\text{tyre}} = v(a + b\tau_{\text{tyre}}^2 + c\tau_{\text{tyre}}^4 + d\tau_{tyre}^6) \tag{11}$$

S_0, c, and k_0 are disc pad wear constants and a, b, c, and d are tyre wear constants.

3. METHOD

Driving strategies are represented by two-layer feedforward neural networks (NNs). An NN is a non–linear function mapping a set of input signals (Y) to a set of output signals (U). The computational nodes, called neurons, are connected to each other forming a net structure as shown in Fig.3. Each neuron is represented by a summation operator and an activation function ($\sigma(s)$) that limits the output from each neuron. The sigmoid function was chosen as activation function: $\sigma(s) = 1/(1 + e^{-s})$, where s is the sum of all input signals to the neuron and the output (z) from each neuron is therefore given by: $z = \sigma(\Sigma w_{\text{ij}} x_{\text{j}} + w_{\text{bias}}\Theta)$ as shown in Fig.4. This output signal is multiplied by a weight (w) and then transmitted as an input signal to a neuron in the next layer, as illustrated in Fig. 3 and Fig. 4. $w_{\text{bias}}\Theta$ ($\Theta \equiv +1$) is a bias term which sets the output level of the neuron in the absence of input. Typical input and output sets used are Y:{speed, disc temperature, road slope, coolant temperature, engine speed} and U:{force ratio between FBs and ABs, gear change, force ratio between VEB and CR, retardation force request}, as illustrated in Fig. 9. Since the activation function limits the output to [0,1], post– and preprocessing of signals measured in the vehicle and signals controlling the actuators have to be performed as shown in Fig. 9, see [6] for discussion. We have chosen to use a genetic algorithm (GA) for the optimization

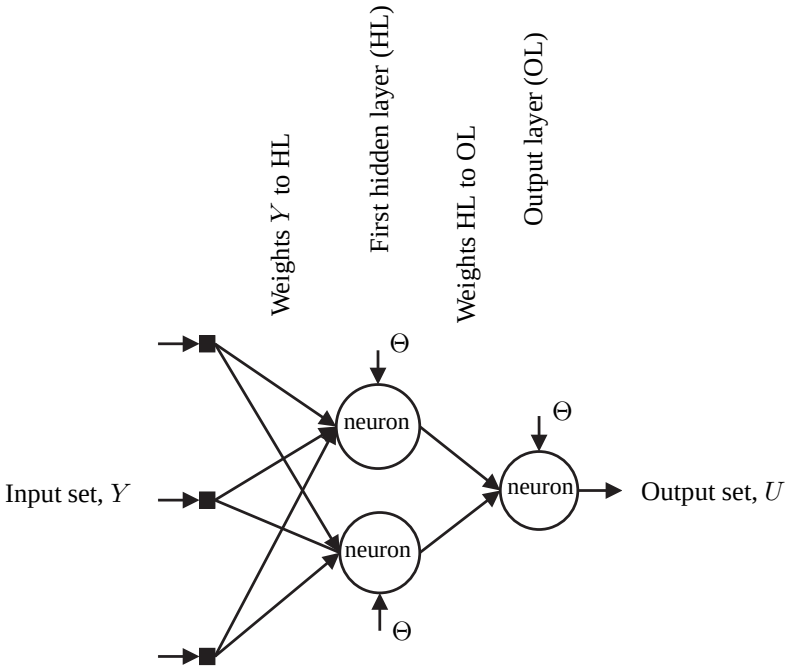


Fig. 3. A feedforward neural network consisting of three neurons, arranged in two layers. The input elements, shown as black filled squares, merely distribute the input signals

of the neural network representing driving strategies, see Fig. 2.

Optimization of neural networks is often performed using the backpropagation algorithm, see [6]. However, in order to apply backpropagation, a set of input-output pairs must be available, which is not the case here. In the problem considered here, the network generates a continuous stream of output signals, and the performance of the network is given as a single scalar value at the end of the evaluation, e.g. the distance traveled by the vehicle during a given period of time. In this situation, a GA is a natural choice of optimization method. GAs have been used by several authors in problems involving NN, see [5] for a review, and are particularly effective in large and complex search spaces.

A further motivation for using a GA is its ability to optimize both the structure (e.g.

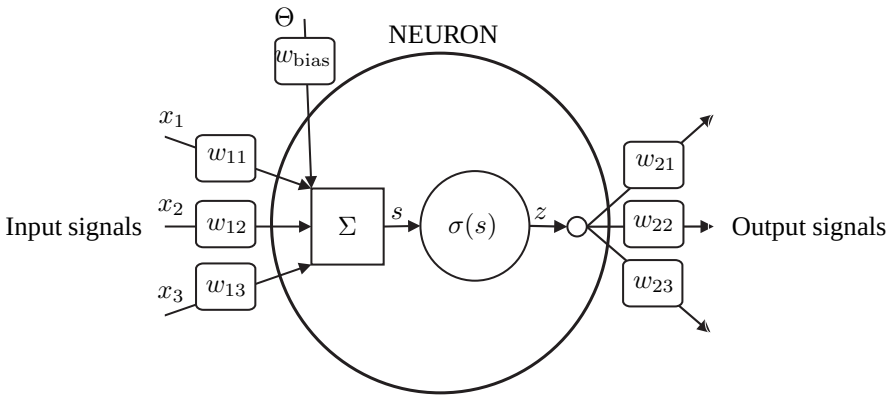


Fig. 4. A neuron

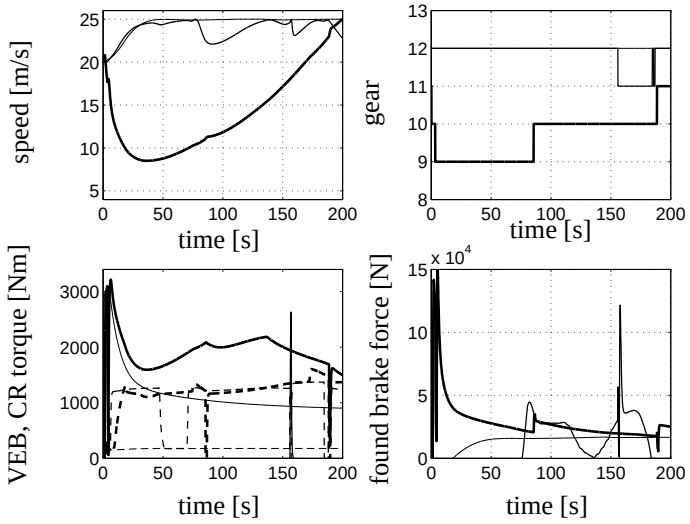


Fig. 5. Optimal blending for high mean speed on 3 roads. Thick solid line: 10% constant slope, Medium solid line: Isère 4 road, Thin solid line: 5% constant slope

the number of neurons in the middle layer) and the parameters of the network. While this feature has not been used in this paper, it is important for further work since, for the problem considered here, it is very difficult to specify an optimal network structure in advance.

For this paper, a fairly standard GA with tournament selection of individuals and

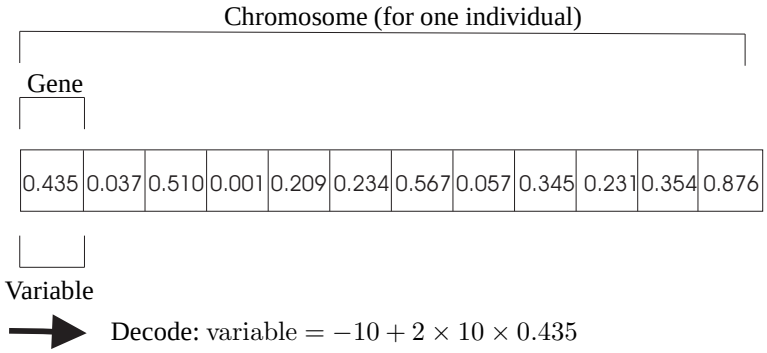


Fig. 6. Chromosome representing one individual. In this example real–number encoding is used

generational replacement has been used. The weights of the network are directly encoded (using real-number encoding) in strings called chromosomes, see Fig. 6. New individuals are generated using crossover and mutations. The crossover probability was set to a rather low value (0.3), since crossover often has a negative effect on neural networks. Mutations of two kinds were introduced: *ordinary mutations*, which change the value of a gene to a new, random value (within the allowed range), and *creep mutations* for which the new value of a gene is chosen with uniform probability in a narrow interval around the previous value. In Fig. 7 one iteration in the GA is shown.

4. RETARDATION ECONOMY

As discussed in [7], transport economy can be described differently depending on where the system boundary is set. If the system boundary is set around the vehicle and the owner of the vehicle, transport economy can be defined using:

- Income from transport mission
- Fuel cost
- Component wear (pads, tyres etc.)
- Other maintenance costs (engine service etc.)
- Utilization of vehicle (hours per day)
- Driver cost (salary etc.)
- Cost of vehicle purchase or rent
- Other costs (tax, road fee etc.)

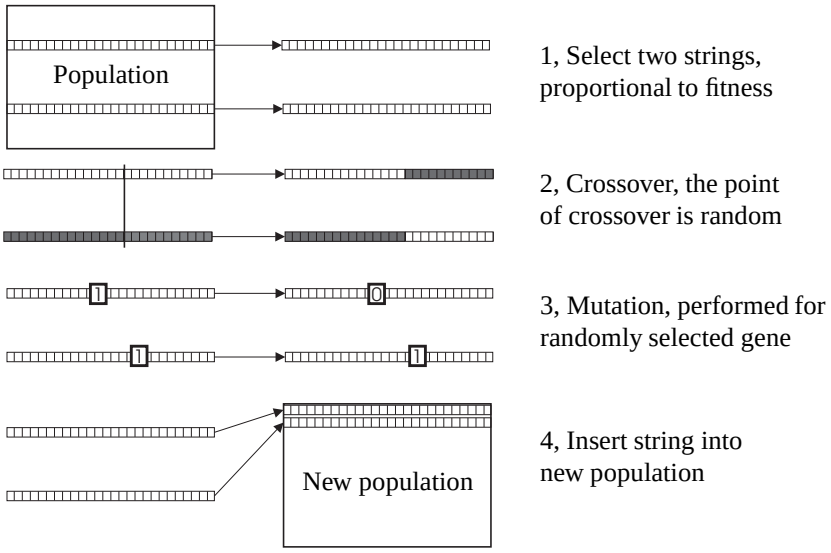


Fig. 7. One iteration in the GA

Looking at transport economy from a retardation control point of view, component wear for low maintenance cost and system performance for high mean speed to improve transport effectiveness are important. Formalizing this statement into a mathematical problem a criterion or fitness function (J) must be defined, e.g. as

$$\begin{aligned} J &= (\text{Cost of retardation})^{-1} \\ &= (\text{Maintenance cost})^{-1} + (\text{Cost of transport time})^{-1} \\ &= J_{\text{wear}} + J_{\text{speed}} \\ &= \frac{1}{C_{\text{tot}}} + \overline{v} \end{aligned} \tag{12}$$

Since it is difficult to quantify the cost of time (though not impossible) the criterion J was split into two different parts for maximization, as shown in Equ. 12 where mean speed ($\overline{v}$) and component wear cost C_{tot} represents cost of transport time and maintenance, respectively. It should also be mentioned that efficient use of the brake system will increase the mean speed on downhill slopes, but it will not necessarily affect the mean speed on the whole driving cycle. If large parts of the driving cycle consist of relatively flat road sections, the mean speed (transport time) is not much affected by an increase in the speed on single down hill slopes. However, high mean

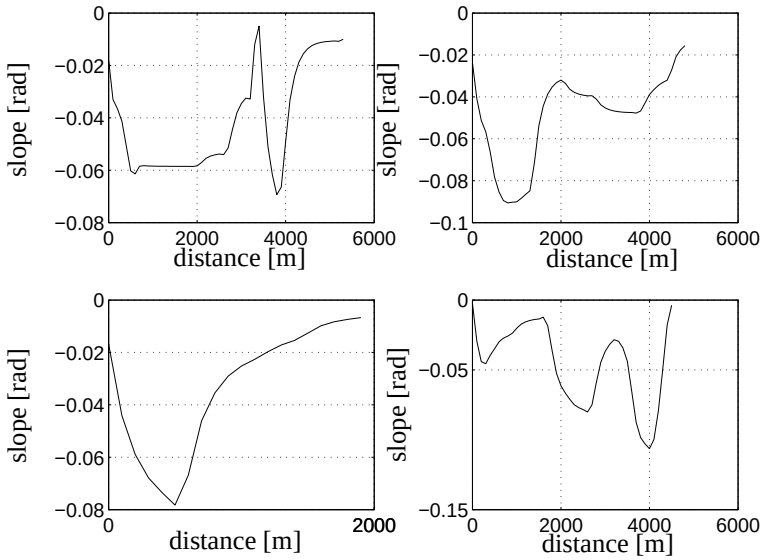


Fig. 8. Example of measured road profiles used. French alps, Isère1-4

speed on down hill slopes is often desirable also to increase the driver comfort and satisfaction (i.e. driveability). Here, a comparison with the development of engines can be made. Reducing emissions and fuel consumption are prime targets for engine developers but other customer demands like engine response, power reserve, top gear gradeability etc. also have to be considered. Stronger and more powerful engines are developed, even if the mean speed on the whole driving cycle does not necessarily increase with more powerful engines.

5. INCREASED MEAN SPEED – CASE 1

Case 1 is focused on optimal utilization of the complete retardation system including FBs, ABs, cooling system, and gearbox. Three examples are shown in Fig. 5 where the objective for the controller is to guide the vehicle down a 5% descent, a 10% descent, and a road profile (nr 4 in Fig. 8) with varying slope. The objective is to achieve highest possible average speed ($\overline{v}$) over a finite time interval without violating the constraints. For this task a 5-7-4 feedforward NN with sigmoid slope equal to 1, and with inputs $\{v, T_1, \alpha, T_{\rm coolant}, v_E\}$ and outputs $\{R_{\rm found_{req}}, {\rm Shift_{req}}, R_{\rm aux_{req}}, F_{\rm split_{req}}\}$ was trained with a GA. As mentioned earlier, both the input signals and the output signals were normalized to the interval $[0, 1]$, see Fig.9.

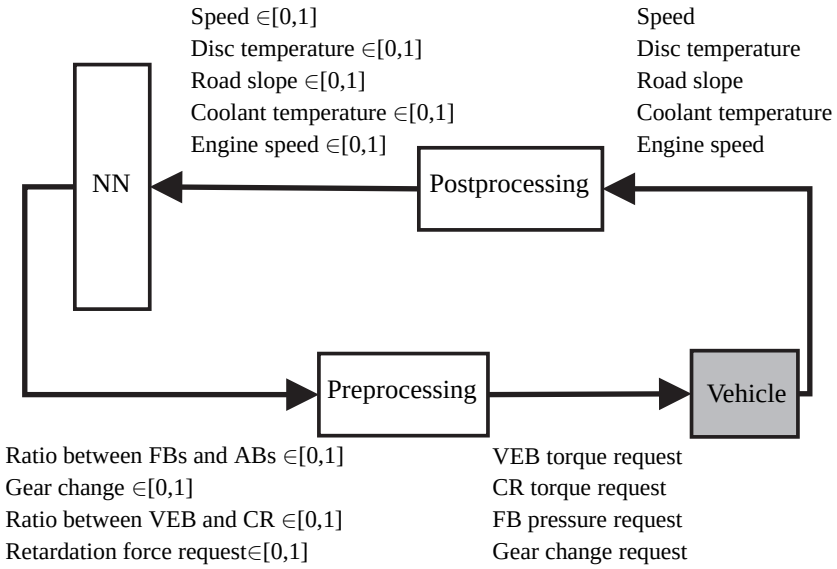


Fig. 9. Closed loop feedback

$F_{\mathrm{split}_{\mathrm{req}}}$ is the total retardation force request, $R_{\mathrm{found}_{\mathrm{req}}}$ splits the total retardation force request between FBs and ABs, from 1 for 100% FBs linearly down to 0 for 100% ABs. $R_{\mathrm{aux}_{\mathrm{req}}}$ splits the AB force requested between VEB and CR, and $\mathrm{Shift}_{\mathrm{req}}$ indicates gear shift (up) if larger than 0.7, gear shift (down) if smaller than 0.3, and no action between 0.3 to 0.7.

The fitness measure was defined simply as the distance travelled by the vehicle. Since the vehicle was only allowed to drive for a limited amount of time, the optimization procedure will strive to attain a high mean speed. The simulation was stopped if any of the following conditions was satisfied: $T_1 \geq 500\,^{\circ}\mathrm{C}$, $v \geq 25\,\mathrm{m/s}$, $v \leq 5\,\mathrm{m/s}$, $v_E \geq 2300\,\mathrm{rpm}$, $v_E \leq 600\,\mathrm{rpm}$, time $\geq 200\,\mathrm{s}$. The solution shown in Fig. 5 was obtained after 1000 generations using a population of 100 individuals. To set the structure (in this case only the number of neurons in the middle layer) some tests where performed. Seven neurons in the middle layer was found to be the minimum number required.

From the figure, it can be seen that the NN works very well with our choice of objective function. The mean speeds of the three runs are 24.9, 14.1, and 23.8 m/s, respectively.

The maximum mean speed that a skilled driver, driving as explained in section 1, can achieve on different slopes is shown in Fig. 11. Also the mean speed obtained from the NN controllers is shown. The NN controllers have the same structure and

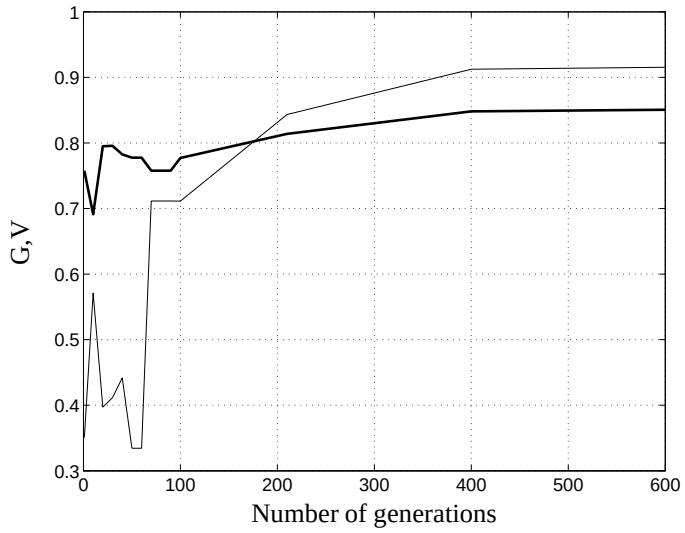


Fig. 10. Generalization measures, G and V. Thick solid line: V, Thin solid line: G

constraints as above, except that the simulation stop constraint is now time ≥ 2000 s, in order to make possible a comparison with the results for a skilled driver. For example, it is seen that the mean speed can be improved by 44% on the 10% slope and 22% on the 6% slope using the whole brake system optimally.

5.1. Generalization

When an NN obtained by training against a single road profile is tested on other roads, it is often found that performance is poor even for small changes in the road profile. It is thus clear that the training road must reflect a variety of obstacles in order for the NN to work in all situations found in reality. For this reason, we constructed a training road that reflects different obstacles encountered in a realistic driving situation. Measured road profiles from the French alps (Isère) and the Kassel hills in Germany were used for this purpose, see Fig. 8. These road sections were then discretized and randomly put together to make up a realistic training road. Also, sections of constant slope were put into this training road.

The main task now is to see to what extent the GA can find an NN that can perform well, i.e. attain a high mean speed, on different roads. Obviously, the ability of the GA to find such an NN depends on several things, such as e.g. the choice of the input and output signals, the shape of the training road, the definition of the fitness measure etc. In this paper we focus on the definition of a training road for generalization. To

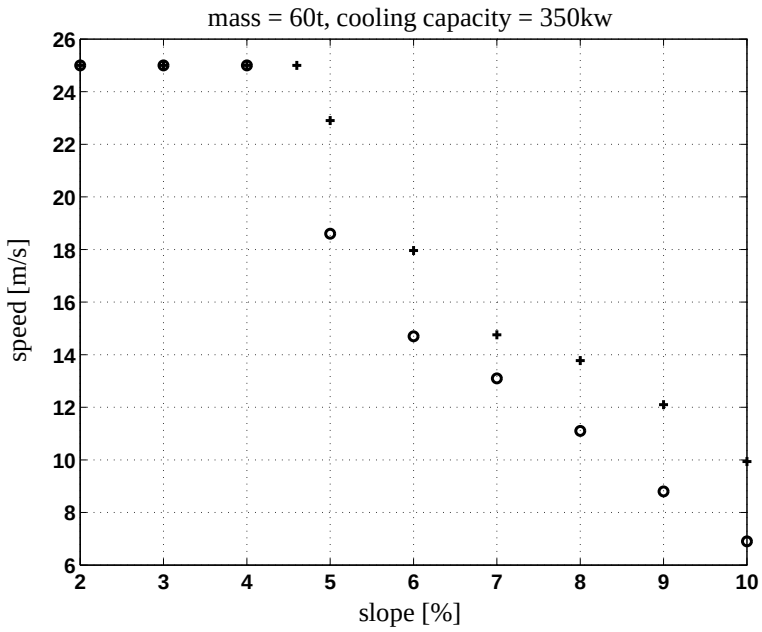


Fig. 11. Maximum stationary speed for different slopes. Circle: Normal driver using only ABs. Plus sign: optimal blending using ABs and FBs

quantify how well the NN is able to generalize we use the measure G and V, defined as

$$G = \frac{1}{N_{\mathrm{r}}} \sum_{i=1}^{N_{\mathrm{r}}} \frac{d_i}{L_i};\; V = \frac{1}{N_{\mathrm{r}} v_{\max}} \sum_{i=1}^{N_{\mathrm{r}}} \overline{v}_i \tag{13}$$

where d_i is the distance travelled by the vehicle on road i, L_i is the length of road i, and N_r is the number of roads. We used $N_r = 14$ different realistic road sections and the result is shown in Fig. 10. Since G increases with the number of iterations, it is clear that the training road reflects most of the obstacles found in the 14 test roads. When the NN can cope with most of the roads in the test set (high G) it is interesting to see that also the mean speed is improved (V is increases). These are two important indications that the generalization is successful.

In Fig. 12 the first half of the training road is shown together with the trajectories for the final NN. Clearly, the NN performs very well on this difficult training road and in Fig. 13 the speed trajectories for 5 other road profiles are shown. The performance is excellent on Isère 1, 2, and 3 (4 is not shown since it is rather short) whereas on the

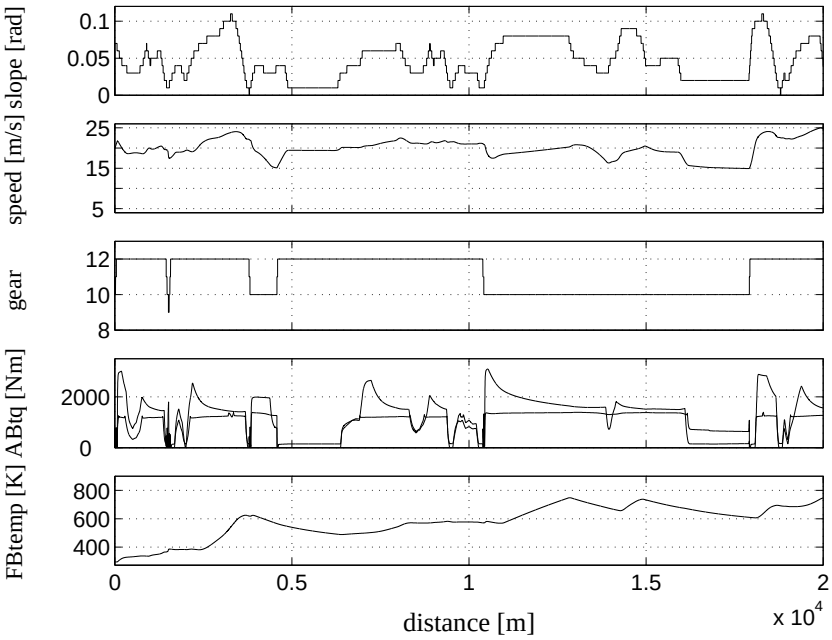


Fig. 12. First half of the training road used and the resulting NN trajectories

two flat roads the NN is somewhat conservative, as can be seen by comparing with Fig. 11.

6. DECREASED WEAR COST – CASE 2

When distributing a required retardation force between drive line and chassis it is interesting to look at wear characteristics of tyre, pad, and disc. The usual driver strategy is to keep the vehicle stationary using only ABs and, if necessary, lower the vehicle speed with the FBs to a level where the ABs can keep the vehicle stationary. However, one might ask oneself whether this is optimal from a wear cost point of view.

The economical model to calculate cost can be formulated in several different ways, and one of the main questions is whether or not to include the time of maintenance stops. As mentioned in section 4 we do not and the total cost ($C_{\rm tot}$) is formulated as in Eq. 14. The work cost ($C_{\rm w}$) and the material cost ($C_{\rm m}$) are what the truck owner had to pay for changing pads, discs, or tyres at a Volvo workshop in Sweden in 2002. Since the disc wear is not modelled we have used the rule of thumb often practiced at the workshops that every second time pads are changed, discs are also changed. δ

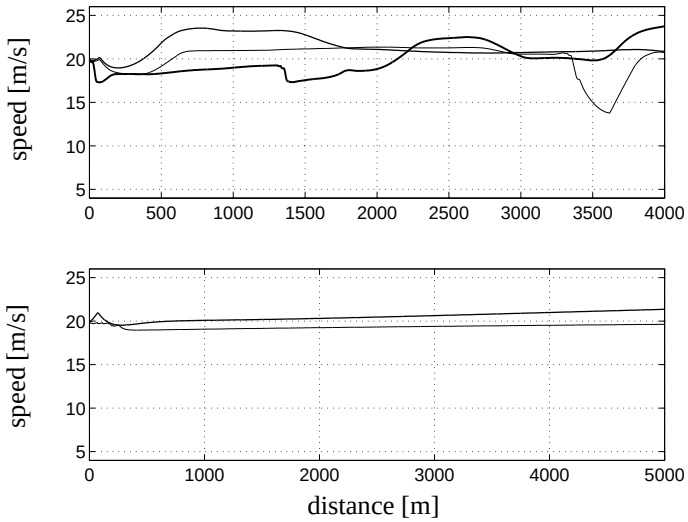


Fig. 13. Final NN tested on 5 road profiles. Isère 1, 2, 4 and flat 2% and 4%

Table 1. Components in equation 14

W_{disc}	Disc wear [mm]
C_{discm}	Material cost of disc [Euro]
C_{discw}	Work cost to change disc [Euro]
δ_{disc}	Available thickness of disc [mm]
W_{pad}	Pad wear [mm]
C_{padm}	Material cost of pad [Euro]
C_{padw}	Work cost to change pad [Euro]
δ_{pad}	Available thickness of pad [mm]
W_{tyre}	Tyre wear [mm]
C_{tyrem}	Material cost of tyre [Euro]
δ_{tyre}	Available tyre thread [mm]

denotes the original disc thickness, pad thickness, and tyre tread thickness. In tab. 6 all components of Equ. 14 are described.

$$C_{\mathrm{tot}} = W_{\mathrm{disc}} \frac{C_{\mathrm{discm}} + C_{\mathrm{discw}}}{\delta_{\mathrm{disc}}} + W_{\mathrm{pad}} \frac{C_{\mathrm{padm}} + C_{\mathrm{padw}}}{\delta_{\mathrm{pad}}} + W_{\mathrm{tyre}} \frac{C_{\mathrm{tyrem}} + C_{\mathrm{tyrew}}}{\delta_{\mathrm{tyre}}} \tag{14}$$

In Fig. 14 it is seen how the optimal distribution for minimum cost varies with retardation force demand. The vehicle is travelling with constant speed (15 m/s) on

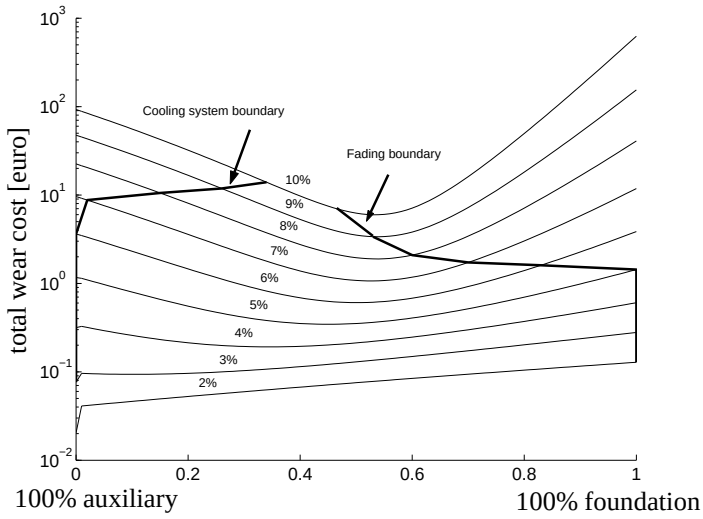


Fig. 14. Vehicle of 60 t and 6 axles travelling at constant speed 15 m/s down a 3000 m descent for 9 different slopes

9 different slopes of length 3000 m. Clearly, the optimal distribution is different for different slopes. This is due to the non-linear and time varying dynamics of Eqs. 11, 10, and 4. Thus, there is a need to design and investigate controllers that minimize wear cost. The idea is to have an NN that shifts gear and splits the force required for constant speed keeping in such a way that the wear cost is minimized. In reality, a PI controller can produce the force request, but here it is solved directly from Eq. 1 and, for this case, the inputs to the neural network are $\{F_{\rm req}, T_1, \alpha, T_{\rm coolant}, v_E\}$ and the outputs are $\{R_{\rm found_{req}}, {\rm Shift_{req}}\}$. For simplicity only two outputs are used, and the VEB has priority over the CR, i.e. the AB force demand is first requested from the VEB, and if the VEB is unable to deliver the demanded torque the remainder is requested from the CR.

Three examples where the NN guides the vehicle down a 2% descent, a 5% descent, and the varying road profile 4 from Fig. 8 are shown in Fig. 15. The fitness function is here defined as $e^{\alpha(t-T_{\rm stop})}/C_{\rm tot}$, where t is the time when the vehicle was stopped (either because a constraint was violated, in which case $t < T_{\rm stop}$, or because the stop time was reached). The constraints used are the same as earlier, except that the allowed range of variation for v is much narrower. It can be seen that the total wear cost for cruising down the 3000 m long 5% slope is approximately 0.38 euro in the optimal case. From Figs. 11 and 14 it can be concluded that a very experienced driver can keep 15 m/s by only using the ABs and the wear cost is then 1.1 euro. Thus, the wear cost

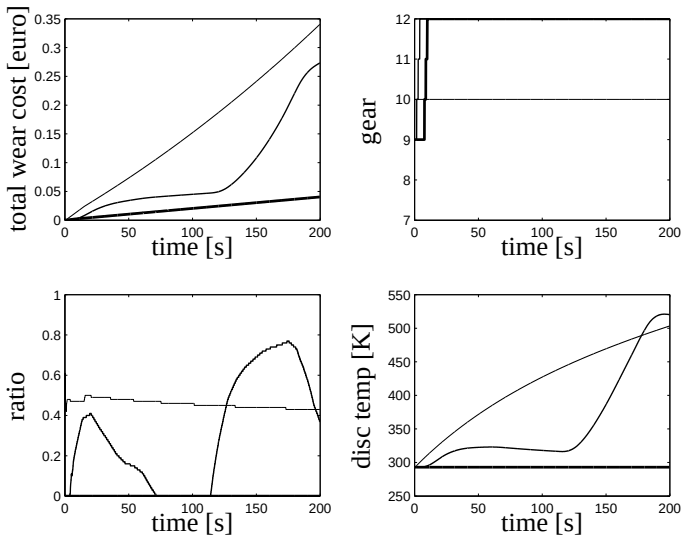


Fig. 15. Optimal blending for low wear cost at 15 m/s on 3 roads. Thick solid line: 2% constant slope, Medium solid line: Isère 4 road, Thin solid line: 5% constant slope. The *ratio* is the distribution between FBs and ABs

of the NN controller is almost 65% lower.

7. DISCUSSION

In this paper, optimization of driving strategies has been carried out using GAs only, and it has been shown that this method yields NNs that are able to cope with general road profiles in many cases.

A further improvement of the procedure would be to generate a set of basic situations, and to provide both input and output signals for these situations. In this case, the NN could first be trained, using e.g. backpropagation, to cope with the basic situations, and then be further trained against a road profile, using a GA.

A setback with the GA approach is that it does not guarantee optimality of the networks obtained. However, finding a provably optimal solution is not crucial in this problem. Instead, what is important is to find an NN that outperforms even an experienced truck driver.

A common objection to the use of NNs is the difficulty of interpretation. NN are sometimes considered to represent a black-box solution to the problem at hand. While interpretability is not necessarily relevant, it would be useful also to consider other

architectures, such as for instance fuzzy logic controllers (which can also be optimized using a GA). The issue is currently being investigated, and will be addressed in a forthcoming paper.

The input signals used here are a feasible choice, since all of the signals can be obtained in a real vehicle. The vehicle speed, engine speed, and coolant temperature are all signals measured in a modern heavy duty vehicle. The road slope and the vehicle mass (which is not used as an input signal here but can, in reality, vary significantly) are not as easy to access, but they be estimated, see [3]. The disc temperature is usually not measured or estimated today, but temperature sensors and estimation algorithms are under development.

For some road profiles, the NN trained for generalization is somewhat conservative (low mean speed) compared to real drivers. This can be explained by the fact that a real driver has road profile preview whereas the NN only has information about the current slope. In future work, preview will be included as well, since such a feature will probably become available in real vehicles in the near future (using e.g. GPS and maps).

8. CONCLUSION

Controllers have been obtained for several different road models (with different slopes). Their performance is shown to be much better even than that of an experienced driver. It is shown that there is a potential to improve mean speed in down hill cruising by optimal usage of the whole brake system, including FBs, ABs, gear box, and cooling system. It is also shown that it is possible to obtain a general NN that can handle a wide variety of different road profiles.

An additional advantage is that the resulting NN controllers are, in principle, directly implementable in an actual vehicle, as opposed to the infinite-dimensional trajectories obtained by optimal control methods. It should also be noted that the resulting optimization problem, when transforming the problem at hand into an optimal control problem, is very large (not suitable to solve using search methods) and non–convex due to the nonlinearities of the differential equations, [9], [8]. The problem must therefore be solved using gradient based optimization and, again, no guarantees can be given regarding optimality.

When considering wear cost of pad, disc, and tyres it is concluded that the strategy currently used by drivers is non-optimal. It is clear that the wear cost can be lowered by distributing the retardation force in an optimal way between FBs and ABs.

ACKNOWLEDGEMENTS

The authors are thankful to Dr. Johan Hulten at Volvo 3P for supplying the foundation brake model, to Dr Heiko Grünberg at Continental for supplying the tyre wear model, and to Mr Arne Andersson at Volvo technology for valuable discussions in the work of simplifying his original model of the cooling system.

REFERENCES

1. Andersson, A.: Transient multi disciplinary engine thermal management simulation, *Proc. of VTMS4 Conference*, 1999, London, England.
2. Kress, T. and Kress, R.: A tool for downhill accident analysis and brake design evaluation, it International Journal of Vehicle Design, 26 (2001) pp. 361–373.
3. Lingman, P. and Schmidtbauer, B.: Road slope and vehicle mass estimation using Kalman filtering, *supp. to Vehicle Systems Dynamics* 37 (2001), pp. 12–23.
4. Lingman, P.: Integrated retardation control, *thesis for the degree of licentiate of enginering, Chalmers University of Technology*. 2002, Göteborg, Sweden.
5. Yao, X.: Evolving artificial neural networks, *Proceedings of the IEEE* 87(9) (1999), pp. 1423–1447.
6. Hayking, S.: Neural Networks–A Comprehensive Foundation, *IEEE Computer Society Press. Macmillan Collage Publishing Company*, 1994
7. Lindkvist, A. and Gustavsson, B.: Lastbilskostnader, *Teknisk rapport, Transportforsknigskommissionen (TKF)*, 1980.
8. Boyd, S. and Vandenberghe L.: Convex optimization (2002), Draft from http://www.stanford.edu/class/ee364.
9. Betts, JT.: Practical Methods for Optimal Control Using Nonlinear Programming (2001), *SIAM, Society for Industrial and Applied Mathematics*

Paper 3

Utilizing Foundation Brakes in a Heavy Duty Vehicle Cruise Controller Application

Peter Lingman and Bengt Schmidtbauer

Submitted to Vehicle System Dynamics, 2004

Utilizing Foundation Brakes in a Heavy Duty Vehicle Cruise Controller Application

PETER LINGMAN[*,†,1] AND BENGT SCHMIDTBAUER[*]

SUMMARY

Two different longitudinal brake system controllers are derived to control foundation brakes, auxiliary brakes, and the gear box of a heavy duty vehicle. By altering driving strategies between the two modes, velocity mode (VM) and temperature mode (TM), transport efficiency and safety are increased compared to what is typically achieved by experienced drivers. In VM, the vehicle is controlled to constant speed using a linear quadratic controller with integral action. The TM controller aims at controlling the foundation brake temperature to a constant level, well below the fading temperature limit of the brakes. Gain scheduling combined with linear quadratic control is used to handle the nonlinearities of the process. Two different scheduling approaches are used. The first approach considers only stationary operating conditions and uses the road slope as scheduling variable. The second scheduling strategy is extended, using velocity based linearization, to also consider non–stationary operating conditions in the controller design. In the second case road slope and vehicle speed are used as scheduling variables. Both controllers perform well but when compared, the transient behavior of the case 2 controller is better. Robustness to parameter variations, measurements errors, and discontinuities due to the discrete gearbox are verified in simulations. The switching between VM and TM is realized in a state machine using input signals based on brake temperature, steady state performance, and vehicle speed.

1. INTRODUCTION

Cruising downhill with a heavy duty vehicle requires several actions from the driver. First of all a proper set speed must be decided upon. To obtain and keep this set speed the driver has to change gear and engage both foundation brakes (**FB**) and the auxiliary brake (**AB**). A problem is that the driver has almost no feed-back regarding vehicle mass and brake disc temperature. A common feature with most kinds of FBs is that they lose their brake capability when overheated (known as heat fading).

[1]Address correspondence to: Peter Lingman, Department of chassis and vehicle dynamics, Volvo 3P, Gropegårdsgatan, 405 08 Göteborg, SWEDEN. Phone +46 31 327 4342.
e–mail: peter.lingman@volvo.com
[†]Department of chassis and vehicle dynamics, Volvo 3P
[*]Div. of Mechatronics, Department of Machine and Vehicle Systems, Chalmers University of Technology, 412 96 Göteborg, SWEDEN.

Accidents have occurred in which a driver has chosen to drive to fast downhill, resulting in overheating of the FBs. Other more conservative drivers choose a low speed to ensure that the system is within safety limits. The drawback is of course a low mean speed. Thus there is a need to design an integrated retardation control system for high mean speed (high transport effectiveness and driver satisfaction) while keeping the system within certain boundaries (safety and law) [1], [2].

This paper suggests a way to include this complicated downhill cruising situation into the framework of ordinary constant speed cruise controllers (**CC**) for heavy duty vehicles. A mode switching cruise controller (**MSCC**) is proposed to switch between velocity mode (**VM**), and temperature mode (**TM**). The main idea is to follow a set speed (or set distance to the vehicle in front), $v_{\rm ss}$, by controlling engine, FBs, and ABs. This mode is called VM and is basically a standard cruise controller that can be found in all modern trucks. One important difference is however that also the FBs are used when there is a "high" brake demand from the controller. The usage of FBs will increase the performance of the controller since the available brake power is substantially increased compared to using only ABs (as in ordinary cruise controllers). However, one problem that arise with the introduction of FBs is downhill driving. In a downhill driving situation the FB may be highly utilized in order to keep the set speed $v_{\rm ss}$, potentially leading to heat fading and therefore a total loss of FB power. Therefore, when the system is approaching FB fading, $v_{\rm ss}$ cannot be tracked anymore and the speed of the vehicle must be lowered (if AB are fully utilized) in order to reduce the power input to the FB and to be able to shift down the gearbox to enable more drive axle AB brake torque. The brake system is utilized maximal when AB are engaged 100% and when the temperature of the FB is kept at a constant value, i.e. the heat dissipated from the FBs equals the FB brake power input (energy balance). A high FB temperature increases the steady state performance (higher downhill cruising speed) of the vehicle since the FB heat dissipation increases with temperature (more radiation), see Figure 4

Therefore, the controller is shifted to a one that tracks a reference in FB temperature instead of vehicle speed (TM). This approach results in both increased safety by avoiding fading and increased transport efficiency by using FBs, when needed, to increase the maximal downhill cruising speed. Furthermore, the interface between driver and vehicle can be kept similar to existing ones.

2. THE RETARDATION SYSTEM

The complete retardation system in a heavy duty vehicle is built up by several chassis, transmission, and engine systems. On the chassis there are two different type of FBs that dominate the market, drum brakes and disc brakes, each with different force and

thermal properties. In this paper, the focus is on the disc brake type, since it is the one that probably will dominate the truck market in the foreseeable future. During braking, the brake cylinder is filled with pressurized air creating a push force on the piston. This force transfers to the pads and creates a clamp force on the disc, and hence a brake torque is generated on the wheel. The kinetic energy of the vehicle is transformed into heat that transfers from the disc by convection and radiation to the surrounding air, and by conduction to the hub, bearing, axle etc. .

The characteristics of the FB can vary significantly due to high variations in the friction properties between pad and disc. Depending on, e.g. disc temperature, humidity of the air, temperature of the air, and non–wanted coatings like road salt and rust, the friction can vary strongly. Two phenomena called **fading** and **glazing** affect the friction between pad and disc. Fading is a loss of friction when the temperature exceeds a certain level. For disc brakes this begins around 600°C. Around 800°C there is only approximately 40% of the friction at 600°C left, and at 900°C the pads burn. These temperatures correspond to measurements a few mm under the disc surface. Glazing, on the other hand, occurs when the FB is used too little. If the driver never uses the brakes for high power retardation, a low friction layer will develop between pad and disc, resulting in bad brake performance. It is, in–fact, a problem for the industry that drivers tend to use the disc brakes too little, resulting in bad braking performance and low traffic safety. Low usage allows not only undesirable coatings to accumulate on the disc but also rust and dirt on the brake mechanism. When it comes to glazing, it is not easy to quantify exactly what ”too little braking” is, since it depends on factors such as the surrounding environment, disc and pad material, and disc temperature when braking. A rule of thumb used is that one or two high power brake manoeuvrers should be made each week to clean and condition the discs from unwanted surface coatings.

Several types of ABs (also called **non-wear brakes**) are on the market and they are all characterized by their non–wear properties. **Hydrodynamic retarders**, **engine compression retarders**, **engine exhaust retarders**, and **electromagnetic retarders (eddy-current brakes)** are some examples. In this paper, only the exhaust type of retarder (engine brake) is used to represent the AB since it is a standard component mounted to almost all Volvo trucks. The Volvo Engine Brake (**VEB**) is a combination of a compression and an exhaust retarder and is part of the engine. Since it is mounted in front of the gear box it is also called **primary retarder**, see figure 1. The exhaust retarder is a controllable valve mounted on the exhaust manifold of the engine. When engaged, the valve prevents the exhaust gases from flowing out from the engine, thereby increasing the pressure in the exhaust manifold, and therefore generating a negative (braking) torque on the driveline. The principle of the compression retarder is almost the same as for the exhaust retarder. An extra notch on each camshaft exhaust lobe is used to modify the intake and exhaust phases of the engine. Making the normal exhaust stroke a compression stroke, the exhaust gases are compressed.

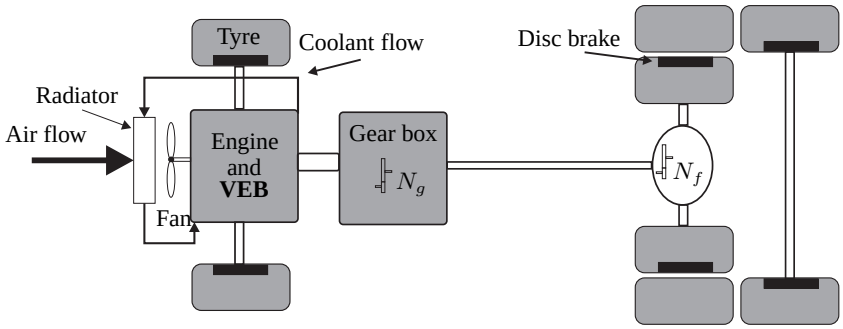


Fig. 1. Overview of a Volvo retardation system

When the piston is almost at its top dead center, the exhaust valves are opened, and the piston is therefore prevented from ”bouncing down”, and a net retardation force is generated on the piston. The VEB transfer vehicle kinetic energy into heat of wich approximately 40% is transfered to the which is then transferred to the engine cooling system and 60% via the exhaust gases.

In Figure 1 the main components of the retardation system used in this paper are presented. The foundation brake power is distributed on all wheels (including trailer) whereas the auxiliary brake works only on the drive axle. Trailers only have FBs and often of disc brake type. A typical combination for gross vehicle weight 60 tonnes has a draw bar trailer with 5 or 4 axles whereas a typical 40 ton combination has a trailer with 3 axles.

2.1. Vehicle model

In order to design and evaluate an integrated retardation controller a mathematical vehicle model was built. The model is very simple and therefore suitable for controller synthesis. The main equations of importance for the development of VM and TM controllers are presented below.

Longitudinal motion equation

$$m\dot{v} = F_{\text{eng}} - F_{\text{air}} - F_{\text{roll}} - F_{\text{grade}} - \\ -F_{\text{AB}} - F_{\text{FB}}$$

$$
\begin{aligned}
F_{\text{air}} &= C_{\text{air}} v^2 \\
F_{\text{roll}} &= C_{\text{roll}} m \\
F_{\text{grade}} &= mg \sin(\alpha)
\end{aligned} \tag{1}
$$

F_{eng} is the propulsion force, F_{air}, F_{roll}, and F_{grade} are resistance forces and F_{AB} and F_{FB} are AB and FB forces, respectively.

Foundation brake dynamics
First order dynamics with a time constant $T_{\text{FB}} = 0.4\text{s}$ is assumed for the FB. This corresponds to the pressure build up in the brake cylinder.

$$
\dot{F}_{\text{FB}} = \frac{1}{T_{\text{FB}}}(F_{\text{FBreq}} - F_{\text{FB}}) \tag{2}
$$

The temperature dynamics of the FB is a distributed process that, when represented on ODE form, has to be discretized. Here 2 masses are used, represented by T_1 and T_2. In Figure 2 the "position" of temperatures T_1 and T_2 are shown. T_1 represents a temperature 3 mm from the disc surface at radius of 170mm. T_2 represents a temperature 15 mm from the disc surface at a radius of 32mm, i.e. 3 mm from the spline that mounts the disc to the axle. Heat fading is represented by T_1 and it is therefore used as reference temperature for the TM controller.

$$
\begin{aligned}
\dot{T}_1 &= q_{\text{in}} - c_1(T_1 - T_2) - c_2(v)(T_1 - T_{\text{amb}}) - c_3(T_1^4 - T_{\text{amb}}^4) \\
\dot{T}_2 &= c_4(T_1 - T_2) - c_5(v)(T_2 - T_{\text{amb}}) - c_6(T_2^4 - T_{\text{amb}}^4)
\end{aligned} \tag{3}
$$

q_{in} is the FB brake power, i.e. $q_{in} = F_{\text{FB}} v$. c_1 and c_4 are heat conduction constants, $c_2(v)$ and $c_5(v)$ are vehicle speed dependent convection variables, and c_3 and c_6 are radiation constants.

Auxiliary brake dynamics
First order dynamics with a time constant $T_{\text{AB}} = 0.3\text{s}$ is assumed for the AB. This is the pressure build up for the exhaust and compression brakes.

$$
\dot{F}_{\text{AB}} = \frac{1}{T_{\text{AB}}}(F_{\text{ABreq}} - F_{\text{AB}}) \tag{4}
$$

Engine dynamics
First order dynamics with a time constant $T_{\text{eng}} = 0.4\text{s}$ is assumed for the engine.

$$
\dot{F}_{\text{eng}} = \frac{1}{T_{\text{eng}}}(F_{\text{engreq}} - F_{\text{eng}}) \tag{5}
$$

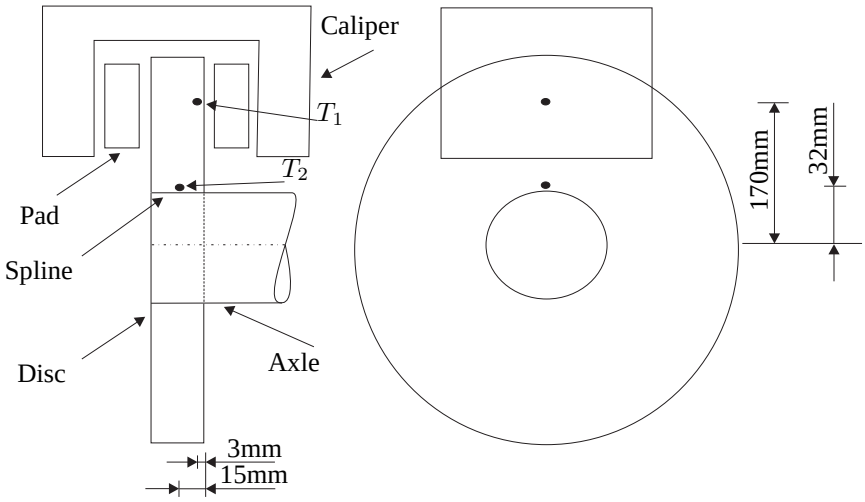
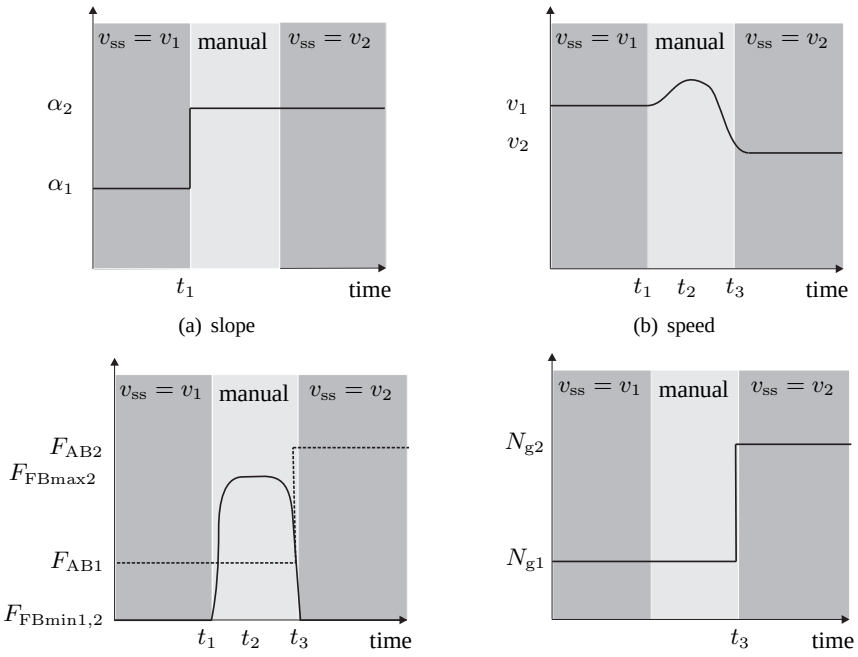


Fig. 2. Measurement points that correspond to foundation brake temperatures T_1 and T_2

3. PROBLEM FORMULATION AND MODE SWITCHING

As described earlier the main idea of the MSCC is to enable the driver to set a reference speed, as in a conventional CC systems (except for the FBs), and then assist the driver in situations where the chosen reference speed is to high and needs to be lowered in order to keep the system within safety limits (FB temperature). A typical "normal driving" sequence is shown in Figure 3. The driver has engaged the CC to track $v_{\rm ss} = v_1$. At time t_1 the downhill slope increases from α_1 to α_2. Since the available AB torque is not sufficient for the CC to track v_1, the vehicle will overspeed. Therefore, it is necessary for the driver to interact with the FB shortly after time t_1. The CC is now turned off and the vehicle is manually controlled by the driver until time t_3. At time t_3 the vehicle speed is v_2 and the driver can shift down to increase N_g (shift down) and engage the CC with the new lower set speed v_2. This new set speed, v_2, is not trivial to find and in practice the driver has to make an iterative procedure of lowering speed and shifting down to find a suitable v_2. By actively controlling the vehicle speed, clearly the overall driver and safety situation can be approved. The first obvious approach would be to automate this normal driver behavior by continuously calculating the new set speed v_2 and then, when the vehicle over speeds a predefined level, use the FBs and gear box to achieve the new set speed v_2. This approach does however not improve the retardation performance of the vehicle since it does not include the FBs more than previously. One way to achieve this is to also utilize the FBs to control the vehicle speed not only in situations where the speed is reducing but also in situations



(a) slope (b) speed

(c) brake force. Dashed AB, solid FB (d) gear ratio

Fig. 3. Typical manual (skilled driver) downhill driving sequence. Foundation brakes are engaged for a short time to reduce vehicle speed to enable gear down shift. The new vehicle speed is the stationary speed that can be held using only auxiliary brakes.

where the speed is kept constant as described in section 1. This approach introduced a need to also control FB temperature in order assure safety and the control problem can therfore be split into one mode where the vehicle speed is controlled (VM) and another where the FB temperature is controlled. In this section the overall algorithm that integrates (mode switcher) the VM and TM is presented and in the following sections the sublevel VM and TM controllers will be described and analyzed. Finally simulations of the complete controller will be presented.

To achieve a well performing mode switching (**MS**) algorithm several aspects must be covered. Most importantly the safety in terms of vehicle speed and FB temperature must always be ensured so that the vehicle is not traveling faster than the driver set speed and that the brakes do not fade. At the same time it is important to minimize the difference between driver chosen set speed and actual speed, i.e. to utilize the brake system to the fullest. For the MS, the most important signal to decide on when to switch between VM and TM is of–course the FB temperature (T_1 that indicates

fading). If the temperature of the FBs is greater than a predefined limit (or reference $T_{\rm ref}$ in TM) a switch from VM to TM should be done, i.e. the vehicle speed should be reduced in order to secure safety. The procedure should be vice versa if the FB temperature is under the reference. It is however not enough to only use the FB temperature but also the road slope, which is the "driving force" on downhill slopes, must be used in the switching. One example. Assume that the vehicle exits from a downhill slope with TM active and FB temperature $T_1(t)$. If only a temperature reference was used in the switching criterion, the rather slow cooling dynamics of the FB would keep the vehicle in TM for a rater long time (depends of-course on the original difference between $T_1(t)$ and $T_{\rm ref}$). This means that the vehicle could be traveling on a flat road, or even uphill slope, with TM active where in fact VM should be activated and the vehicle should accelerate to the driver set speed. On the other hand, since road topology preview is not assumed, the switching and activation of the VM must be done carefully since the downhill slope may continue after a short flat section, potentially leading to very high transient in FB temperature. If the TM controller would be active on flat or uphill road slopes, the vehicle speed would eventually be reduced to zero since the TM controller cannot produce a positive torque (it controlls the brakes).

Therefore the algorithm for switching between VM and TM is implemented as a state machine with three input signals, generated in three sub modes that each are described below

Sub mode 1
In sub mode 1 the FB temperature (T_1) is considered in order to keep control of FB temperature. The output is:

$$\begin{aligned} &0 \text{ if } \mathrm{T_1} < T_{\rm ref} \\ &1 \text{ if } \mathrm{T_1} \geq T_{\rm ref} \end{aligned} \tag{6}$$

Sub mode 2
In sub mode 2 the steady state downhill cruising performance of the vehicle is evaluated and compared to the actual slope (α). A maximal stationary vehicle speed ($v_{\rm statspeed}$) is calculated for every α. In Figure 4 $v_{\rm statspeed}$ as function of α is shown for three different maximum FB temperatures, 350^oC , 400^oC, and 500^oC). In this case the vehicle combination has 7 axles and a total weight of 60tonnes. Both the AB (Volvo engine brake) and FBs are fully utilized and the maximal allowed engine speed is 2250 rpm. The steps in Figure 4 are due to the discrete gearbox and engine speed limitation. Higher FB temperature results in higher stationary speed, much due the fact that the radiation increases with the temperature, see Equation 3. The output is:

$$\begin{aligned} &0 \text{ if } \alpha(\mathrm{t}) - \alpha_{\max}(\mathrm{v_{ss}})) < 0 \\ &1 \text{ if } \alpha(\mathrm{t}) - \alpha_{\max}(\mathrm{v_{ss}})) \geq 0 \end{aligned} \tag{7}$$

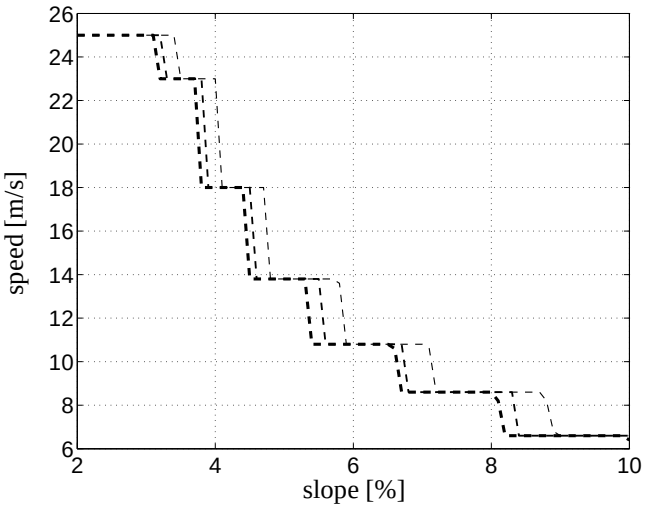


Fig. 4. Maximum stationary speed ($v_{\mathrm{statspeed}}$) as function of road slope (α) for vehicle with VEB and FB. Thick line: $T_1 = 350^0C$. Medium line $T_1 = 400^0C$. Thin line $T_1 = 500^0C$.

Sub mode 3

If the driver has chosen a low set speed, relative to the current road slope, the vehicle can over speed if the MSCC controller is in TM, i.e. the system can handle more brake power than requested in VM. This behavior is of course not acceptable and it is corrected in sub mode 3. The output is:

$$\begin{aligned} &1 \text{ if } \mathrm{v(t)} \leq v_{\mathrm{ss}}(1+c) \\ &0 \text{ if } \mathrm{v(t)} > v_{\mathrm{ss}}(1+c) \end{aligned} \tag{8}$$

Where c is a tunable parameter defining the acceptable overshoot in v.

The state machine used for the MS logic is represented by the truth table in Tab. 1. In Figure 5 one example of the MS, VM controller, and TM controller behavior is shown (the VM and TM controllers are presented in the following section). In this simulation the road slope is increased from a small slope to a large and back again. The vehicle starts in VM and row 2 (no overshoot in speed) where the driver selected set speed is 18m/s. The slope increases much and a transition to row 4 is made. The mode is unchanged and to keep v_{ss} the FBs are engaged. After a while the FB temperature is increased to T_{ref} and a switch to TM is made by a transition to row 8 directly followed by a jump to row 16. The vehicle is now in TM and the vehicle speed, v, is decreased.

Table 1. Truth table representing the state machine for mode switching. $\mathrm{Mode(t)} = 1$ gives TM and $\mathrm{mode(t)} = 0$ gives VM. The mode at the previous time sample is named Mode(t-dt)

Row	Mode(t-dt)	sub mode 1	sub mode 2	sub mode 3	Mode(t)
1	0	0	0	0	0
$2^{1,7}$	0	0	0	1	0
3	0	0	1	0	0
4^{2}	0	0	1	1	0
5	0	1	0	0	0
6^{6}	0	1	0	1	0
7	0	1	1	0	0
8^{3}	0	1	1	1	1
9	1	0	0	0	0
10	1	0	0	1	0
11	1	0	1	0	0
12	1	0	1	1	1
13	1	1	0	0	0
14^{5}	1	1	0	1	0
15	1	1	1	0	0
16^{4}	1	1	1	1	1

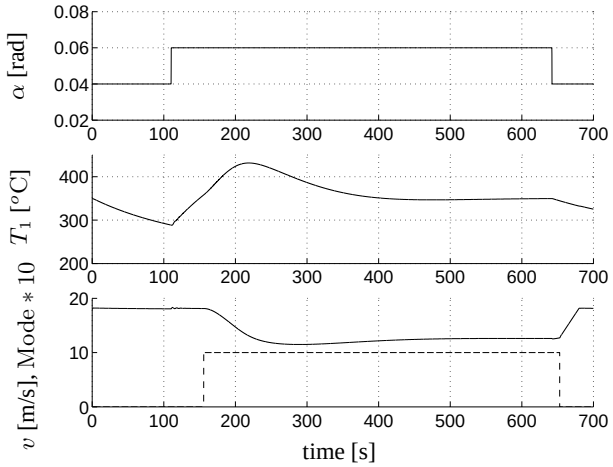


Fig. 5. One example of the mode switching strategy. By using the brakes, vehicle speed is kept at the driver selected set speed. Eventually, the foundation brakes become to warm and the speed is reduced by switching from velocity mode to temperature mode. Finally the slope is reduced. The foundation brakes are warm but the auxiliary brakes are enough to keep the driver selected set speed and therefore a switch back to velocity mode is made.

Finally the slope is decreased and a transition to row 14, followed by a transition to row 6, is made and the MSCC is now back in VM. Once the FB temperature is below T_{ref} a transition back to row 2 is made (no change in mode). This sequence of events is marked by numbers 1 to 7 in Tab. 1.

4. VELOCITY AND TEMPERATURE MODE CONTROLLERS

Once the framework of the MS controller is designed, sub level controllers for both modes must be designed and tuned to operate smoothly with the MS. In this paper the focus is on handling the integration of FB into the framework of a cruise controller. Therefore, the focus in this section is on the TM controller rather than on the VM controller. Constant speed or constant distance controllers are already well established on the market even if they usually do not engage the FB at all. However, we start by designing a VM controller using linear quadratic control (**LQ**) with integral action. LQ control is, as indicated by the name a linear model based control synthesis solving a quadratic criterion function. The control signal u (scalar here) is constructed by multiplying the system state vector (x) by a constant vector (L), i.e. the $u = -Lx$. The state feedback gain, L, is obtained by solving the algebraic Riccati equation, Equation 10, that minimizes the objective function, J, in Equation 11. The coefficient matrices R_1 and R_2 in the quadratic objective function, J, can be seen as tuning parameters for the controller. By increasing R_2 the control signal is punished, resulting in a slower convergence to the reference value and lower control signal variation, see [3] and [4].

$$0 = SA + A^T S - SBR_2^{-1}B^T S + R_1 \tag{9}$$

$$L = R_2^{-1}BS \tag{10}$$

$$J = E\{x^T R_1 x + u^T R_2 u\} \tag{11}$$

The model used to derive the VM LQ controller consists of the longitudinal force balance in Equation 12 below and a model for the road slope, α, according to Equation 14 below, see [5]. The control signal (u) is the longitudinal force, F, that is distributed between the engine, ABs, and FBs according to Equation 13 below where F_{ABmax} is the maximum available AB force, i.e. if the requested force (F) is positive propulsion force is requested from the engine and if the requested force is negative the AB is used and if that is not enough FBs are used in complement to the AB.

$$\begin{aligned} m\dot{v} = F_{\mathrm{eng}} - F_{\mathrm{air}} - F_{\mathrm{roll}} - F_{\mathrm{grade}} - \\ -F_{\mathrm{AB}} - F_{\mathrm{FB}} \end{aligned}$$

$$m\dot{v} = F - mg\alpha - F_{\text{roll}} - C_{\text{air}}v^2 = f_1 \tag{12}$$

$$\begin{aligned} F_{\text{eng}} \text{ if } & -\text{F} > 0 \\ -F_{\text{AB}} \text{ if } & -\text{F}_{\text{ABmax}} < \text{F} \leq 0 \\ -F_{\text{ABmax}} - F_{\text{FB}} \text{ if } & \text{F} \leq -\text{F}_{\text{ABmax}} \end{aligned} \tag{13}$$

$$\dot{\alpha} = -\alpha/T_\alpha + \nu_\alpha = f_2 \tag{14}$$

It is clear that the first state equation, f_1, in Equation 12 is non-linear. Therefore, linearization has to be performed in order to use LQ controller synthesis. The first order Taylor expansion is according to Equation 15 where x_0, u_0 define the linearization point. If the point of linearization is a stationary point ($F(x_0, u_0) = 0$) for the system, the non-linear state equation in Equation 12 can be directly written as a linear time invariant system $\dot{x} = Ax + Bu$ and $y = Cx + Du$, that can be used to derive the LQ controller.

$$F(x, u) \approx F(x_0, u_0) + \nabla_x F(x_0, u_0)(x - x_0) + \nabla_u F(x_0, u_0)(u - u_0) \tag{15}$$

Linearization at $v = v_{\text{ss}}$ including integral action yields:

$$A = \begin{bmatrix} -2C_{\text{air}}v_{\text{ss}}/m & -g & 0 \\ 0 & -1/T_\alpha & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{1}{m} & 0 & 0 \end{bmatrix}^T$$

$$x = \begin{bmatrix} v - v_{\text{ss}} & \alpha & \int(v_{\text{ss}} - v) \end{bmatrix}^T$$

$$u = \begin{bmatrix} F - F_{\text{roll}} \end{bmatrix}$$

In Section 6 simulations are shown for the VM LQ controller. Depending on loading conditions, environment, and reference speed the VM retardation strategy, that utilizes the FB without restrictions, may result in high FB temperatures, i.e. the available brake power is not enough. Therefore, the objective of the TM brake strategy is to restrict the FB temperature to a constant predefined and safe level (limit or reference). The choice of FB reference temperature is a trade–of between safety and speed. If a high reference is chosen the VM controller can be used for a longer time

before the speed must be reduced by the TM controller. Additionally also the continuous power input to the FB can be higher due to the increased energy dissipation at high temperatures, see Equation 3. On the other hand if, for example, the driver must suddenly engage the brakes hard to avoid a collision, a high reference may result in fading and low brake capability. Here we have chosen $T_1 = 350^0\mathrm{C}$ as reference (T_{ref}) which ensures full emergency brake capability. By also utilizing maximum available AB torque (M_{ABmax}) and by controlling the gear ratio so that the engine speed is maximal, F_{AB} becomes also maximal. Therefore, the difference between driver set speed (v_{ss}) and actual vehicle speed is minimized and the constraints on speed and temperature are held, yielding an optimal driving behavior. One can say that the level of conservativeness of this driving strategy is adjustable by altering the FB reference temperature (T_{ref}). As can be seen in Figure 4 higher temperature increase the stationary speed. Equation 1 and 3 can now, together with the road slope process, be rewritten into the following system of state equations where actuator dynamics are neglected due to much faster dynamics than the temperature dynamics of the FB's.

state equations

$$\begin{aligned}
\dot{T}_1 &= \tfrac{F_{\mathrm{FB}} v}{c_1} - c_2(T_1 - T_2) - (c_3 v - c_4)(T_1 - T_{\mathrm{amb}}) - \\
&\quad c_5(T_1^4 - T_{\mathrm{amb}}^4) && = f_1 \\
\dot{T}_2 &= c_6(T_1 - T_2) - (c_7 v - c_8)(T_2 - T_{\mathrm{amb}}) - \\
&\quad c_9(T_2^4 - T_{\mathrm{amb}}^4) && = f_2 \\
m\dot{v} &= mg\alpha - F_{\mathrm{FB}} - \tfrac{M_{\mathrm{ABmax}}\omega_{\mathrm{engmax}}}{v} - F_{\mathrm{roll}} - C_{\mathrm{air}} v^2 && = f_3 \\
\dot{\alpha} &= -\alpha/T_\alpha + \nu_\alpha && = f_4
\end{aligned} \tag{16}$$

The state vector is $x = \{T_1, T_2, v, \alpha\}$ and the control signal is $u = F_{\mathrm{FB}}$

Performing a linearization of Equation 16 in a stationary point yields the following system matrix and control signal vector that can be used in the LQ design. However, since A, B vary over the operating range of the system, it is not sufficient to design one controller, valid only in the neighborhood of one stationary point. This is also true for the VM LQ controller derived in Section 4 where the system matrix varies with the speed, v. One remedy for this problem is to utilize gain scheduling, see [6]. The main idea with gain scheduling is to calculate control laws for the system in several operating points. When this is performed a family of linear controllers, each valid around one specific operating point, is obtained. Scheduling variable/variables can then be used to choose the control law that best describes the actual operating condition of the process. Usually linear interpolation is used to extract a control law from the family of control laws. Gain scheduling results in a non linear controller. Linearization yields:

$$A = \begin{bmatrix} -c_2 - c_3 v_0 - c_4 - 4c_5 T_{10}^3 & -c_2 & \frac{F_{\mathrm{FB}0}}{c_1} - c_3(T_{10} - T_{\mathrm{amb}}) & 0 \\ c_6 & c_6 - c_7 v_0 - 4c_9 T_{20}^3 & -c_7(T_{20} - T_{\mathrm{amb}}) & 0 \\ 0 & 0 & \frac{M_{\mathrm{ABmax}}\omega_{\mathrm{engmax}}}{v_0^2} - 2C_{\mathrm{air}} v_0 & mg \\ 0 & 0 & 0 & -\frac{1}{T_\alpha} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{v_0}{c_1} & 0 & -\frac{1}{m} & 0 \end{bmatrix}$$

5. GAIN SCHEDULING AND SIMULATIONS FOR TM

The selection of scheduling variables is one of the main questions when it comes to gain scheduling, and the final solution is often found after some testing. In this paper two different scheduling strategies are proposed and evaluated. The first scheduling strategy, **case 1**, utilizes the road slope as scheduling variable and controllers are calculated at stationary points corresponding to different road slopes. The second scheduling strategy, **case 2**, follows the framework of velocity based linearization developed in [7]. In case 2 a transformation is performed to obtain a valid linear representation in every operating point of the system. By using speed and road slope as scheduling variables, controllers can be calculated both at stationary and non–stationary operating points.

5.1. Linearization only in stationary points, case 1

One guideline in the selection of scheduling variables is that the scheduling variables should vary slowly, i.e. it not vary faster than the process dynamics. Before proceeding with the selection of scheduling parameters a state transformation is performed to obtain an "energy based" description of the system. By multiplying f_3 in Equation 16 with the speed, v, a new state and control signal can be introduced. State v is replaced with the kinetic energy per unit mass, $v^2/2$, and the original control signal, F_{FB}, is replaced with the FB brake power, $F_{\mathrm{FB}}v$, see, Equation 17.

$$\begin{aligned} \dot{T}_1 &= \tfrac{F_{\mathrm{FB}}v}{c_1} - c_2(T_1 - T_2) - (c_3 v - c_4)(T_1 - T_{\mathrm{amb}}) - \\ &\quad c_5(T_1^4 - T_{\mathrm{amb}}^4) && = f_1 \\ \dot{T}_2 &= c_6(T_1 - T_2) - (c_7 v - c_8)(T_2 - T_{\mathrm{amb}}) - \\ &\quad c_9(T_2^4 - T_{\mathrm{amb}}^4) && = f_2 \\ mv\dot{v} &= mgv\alpha - F_{\mathrm{FB}}v - M_{\mathrm{ABmax}}\omega_{\mathrm{engmax}} - F_{\mathrm{roll}}v - C_{\mathrm{air}}v^3 && = f_3 \\ \dot{\alpha} &= -\alpha/T_\alpha + \nu_\alpha && = f_4 \end{aligned} \tag{17}$$

linearization now yields:

$$A = \begin{bmatrix} -c_2 - c_3 v_0 - c_4 - 4c_5 T_{10}^3 & -c_2 & \frac{-c_3(T_{10}-T_{\mathrm{amb}})}{v_0} & 0 \\ c_6 & c_6 - c_7 v_0 - 4c_9 T_{20}^3 & \frac{-c_7(T_{20}-T_{\mathrm{amb}})}{v_0} & 0 \\ 0 & 0 & \frac{g\alpha_0}{v_0} - \frac{F_{\mathrm{roll}}}{m v_0} - \frac{3C_{\mathrm{air}} v_0}{m} & g v_0 \\ 0 & 0 & 0 & -\frac{1}{T_\alpha} \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{1}{c_1} & 0 & -\frac{1}{m} & 0 \end{bmatrix}$$

The new linear representation of the system has now a system matrix that varies with T_1, T_2, v, and α. In case 1 the controller is scheduled for different operation points that correspond to stationary points for the system. The road slope is used as scheduling variable. Temperatures T_1 and T_2 are assumed to be constant (reference value) and v_0 is the corresponding stationary velocity of the vehicle. This means that the linear controller derived for each stationary situation has to work also when the vehicle is far away from stationarity. For example, as shown in Figure 3 where the controller should be shifted at time t_1 where the slope changes from α_1 to α_2. This means that the controller used to reach stationarity at time t_3 has to handle the transient situation between t_1 and t_3. This is one type of model error which is obvious during the transient phase. Another modeling error, though not due to the scheduling strategy, is the variation in the friction coefficient between brake pads and disc. In order to reduce steady–state errors due to model inaccuracies integral action or some kind of adaptation should be used. In this paper integral action is implemented in case 1 and therefore, in order to ensures integral action on the reference value, the design of the LQ controller is based on the augmented system in Equation 18. The control signal, u, is defined according to Equation 19 where K_f is the feed–forward gain for the incremental reference signal, Δr, and L^* is the state feed–back gain. A bias term, β, appears also in the control signal according to Equation 19. An overview of the proposed controller in case 1 is shown in Figure 6. The choice of scheduling points and the choice of penalty parameters in the criterion function are presented in Section 5.3.

$$\frac{\mathrm{d}}{\mathrm{dt}}(\Delta x^*) = \begin{bmatrix} A & 0 \\ -1 & 0 \end{bmatrix} \Delta x^* + \begin{bmatrix} B \\ 0 \end{bmatrix} \Delta u + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \Delta r \tag{18}$$

$$\begin{aligned} u &= u_0 + \Delta u \\ u &= u_0 + K_f \Delta r - L^*(x^* - x_0^*) \\ u &= \beta + K_f \Delta r - L^* x^* \end{aligned} \tag{19}$$

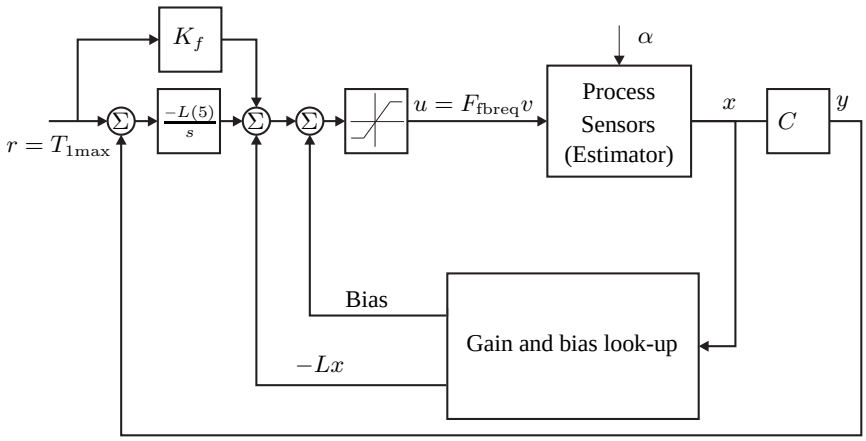


Fig. 6. Overview of gain schedule controller, case 1

5.2. Velocity based linearization, case 2

In case 2 the linearization procedure from case 1 is extended to non–stationary operating points by using both α and v as scheduling variables. In order to achieve this, a linear representation of the system in Equation 17, that is valid in the whole operating range of the system, must be found. The velocity based frame work in [7], offers a mapping of a non-linear set of equations to a linear set of equations that is *valid at every operating point of the system*, i.e. not only near stationary operating points. By taking the time derivative of the first order approximation in Equation 15 the constant term $F(x_0, u_0)$ is eliminated and a linear system is obtained, see [7].

By reformulating the first order approximation according to [7]:

$$\begin{aligned}\dot{x} = &\{F(x_n, u_n) - \nabla_x F(x_n, u_n)x_n - \nabla_u F(x_n, u_n)u_n\} + \\ &\nabla_x F(x_n, u_n)x + \nabla_u F(x_n, u_n)u \end{aligned} \tag{20}$$

where the state and input are the same at every operating point n.

Taking the time derivative of Equation 20 yields the velocity based linearization.

$$\dot{x} = w$$

$$\dot{w} = \nabla_x F(x_n, u_n)w + \nabla_u F(x_n, u_n)\dot{u} \tag{21}$$

The number of states is now twice the number of states in the original system. It should also be noted that the control signal now is $\dot{u}$. Using Equation 21 on the system in Equation 16 yields:

$$\dot{x}_{\text{tot}} = \begin{bmatrix} A_w & 0 \\ A_x & 0 \end{bmatrix} x_{\text{tot}} + \begin{bmatrix} B_w \\ 0 \end{bmatrix} \dot{u}$$

where

$$x_{\text{tot}} = \begin{bmatrix} \dot{T}_1 & \dot{T}_2 & v\dot{v} & \dot{\alpha} & T_1 & T_2 & v^2/2 & \alpha \end{bmatrix}^T$$

$$A_w = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$$

$$A_1 = \begin{bmatrix} -c_2 - c_3\sqrt{2x_{3\text{n}}} - c_4 - 4c_5x_{1\text{n}}^3 & -c_2 \\ c_6 & c_6 - c_7\sqrt{2x_{3\text{n}}} - 4c_9x_{2\text{n}}^3 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} \frac{-c_3(x_{1\text{n}} - T_{\text{amb}})}{\sqrt{2x_{3\text{n}}}} & 0 \\ \frac{-c_7(x_{2\text{n}} - T_{\text{amb}})}{\sqrt{2x_{3\text{n}}}} & 0 \\ \frac{gx_{4\text{n}}}{\sqrt{2x_{3\text{n}}}} - \frac{F_{\text{roll}}}{m\sqrt{2x_{3\text{n}}}} - 3C_{\text{air}}\sqrt{2x_{3\text{n}}} & g\sqrt{2x_{3\text{n}}} \\ 0 & -\frac{1}{T_\alpha} \end{bmatrix}$$

$$A_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

and

$$B_{\text{w}} = \begin{bmatrix} \frac{1}{c_1} & 0 & -\frac{1}{m} & 0 \end{bmatrix}$$

Since the states x do not appear on the right side of $\dot{w}$ in Equation 21, only w must be considered in the controller design. However, since the main output of the system is the disc temperature, T_1, the state vector, w, is augmented with T_1. The equation used in the design of the controller is given in Equation 22 which is a non-controllable (the rank of the observability matrix is 4) but stabilizable system, since the non-controllable road slope process is stable, see [3].

$$\dot{x}_{\text{tot}_{1\ldots5}} = \begin{bmatrix} A_w & 0 \\ 1 & 0 \end{bmatrix} x_{\text{tot}_{1\ldots5}} + \begin{bmatrix} B_w \\ 0 \end{bmatrix} \dot{u} \tag{22}$$

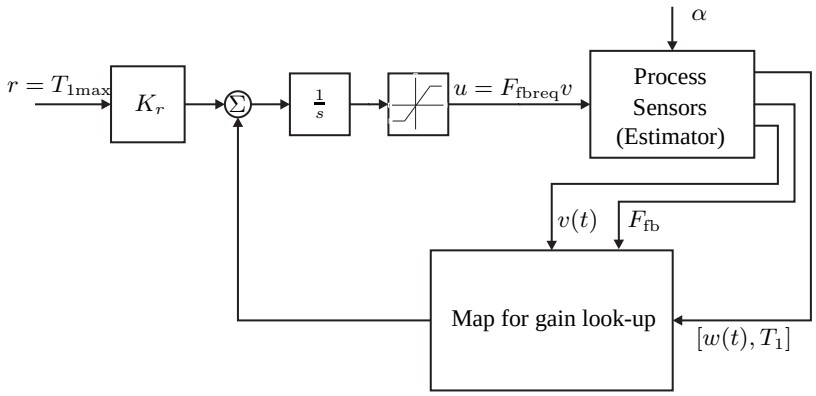


Fig. 7. Overview of gain schedule controller, case 2

An overview of the gain scheduled controller, proposed in case 2, is shown in Figure 7 where K_r $(= L(5))$ is the reference value gain which is designed to obtain zero steady state error from reference value to output. Vehicle speed and road slope are used to find the right control law, L. The reference value and feedback are then summarized and integrated to form the FB power request. The saturation after the integrator is required since the foundation brake force is not allowed to switch sign.

5.3. Simulations for temperature mode

The tuning of the penalties in the criterion function and number of scheduling points should be done based on analysis in both the time domain and frequency domain. Transient responses must be analyzed in the time domain and gain margin, phase margin, internal stability, sensitivity function, and complementary sensitivity function should be analyzed in the frequency domain. However, since this paper only considers the feed–back design without observers, the frequency domain is not considered in the controller tuning process. The stability and disturbance sensitivity properties of the closed loop system are in fact very good (even unrealistic) when using LQ synthesis without observers. For example, the gain margin for reduced gain is never lower than 6dB and the gain margin for increased gain is always infinity.

In systems that have constraints in, for example, actuator amplitudes (saturation) the time domain analysis is very important and exposes problems that are not shown in the frequency domain. From Equation 17, it is clear that, when the process dynamics are formulated in TM, one unstable pole is introduced through the road slope in Equation 17. This open loop unstable pole is always moved to a closed loop stable pole when the feed–back is designed with LQ methodology. However, since the con-

Table 2. Seven stationary points that are used as scheduling points for case 1

α [rad]	T_1 [^{o}C]	$F_{\rm fb}$ [N]	v [m/s]
0.05	350	3384	15.8
0.06	350	4157	12.6
0.07	350	4897	10.5
0.08	350	5649	9.1
0.09	350	6361	7.9
0.10	350	7100	7.1
0.11	350	7842	6.4

trol signal is restricted to only positive values (saturation), stability may not always be achieved. When the condition, $F_{\rm AB} > F_{\rm grade} - F_{\rm roll} - F_{\rm air}$, is fulfilled the vehicle speed is decreased to zero. To avoid this, a stabilizing action is implemented by decreasing $F_{\rm AB}$ when $F_{\rm FBreq} < 0$, see Equation 23. Where $r_{\rm wheel}$, N_g, and N_f are wheel radius, gear ratio, and final gear ratio respectively.

stabilizing action:

$$\begin{aligned} &if \;\; F_{\rm FBreq} < 0 \\ &\quad M_{\rm ABreq} = M_{\rm ABmax} - (F_{\rm FBreq} r_{\rm wheel})/(N_g N_f) \end{aligned} \tag{23}$$

To be able to compare the two scheduling techniques, step responses for a small step in the road slope around steady–state operating conditions were studied. The controller in case 1 was scheduled in seven different stationary points according to Tab. 5.3. In case 2 scheduling was performed in all the combinations of $\{\alpha, v\}$ as shown in Tab. 5.3. In order for both the controller to span the same working domain the case 2 controller was scheduled also for the speed $v = 25 \;\; {\rm m/s}$, i.e. 56 scheduling points in total. The tuning was then performed so that the controllers in the scheduling points of case 1 had the same step responses (10% increase in the road slope) as the corresponding controllers in case 2.

For the simulations in TM the following setup was used:

Penalty parameters for case 1 (same in all scheduling points):

$$R_1 = \begin{bmatrix} 9*10^5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5*10^3 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 1 \end{bmatrix}$$

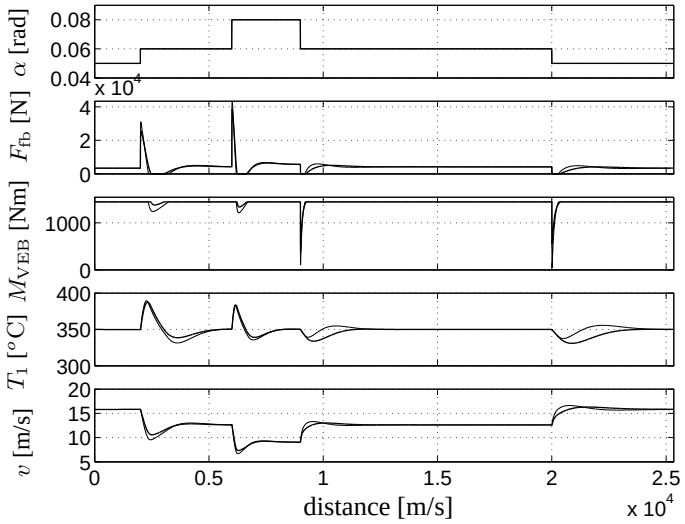


Fig. 8. Step responses for temperature mode with CVT. The maximum temperature transient is approximately 8.6% over the set temperature. Stabilizing action as active (dip in $M_{\rm veb}$) at every slope change to ensure stability. Thin lines are for case 1 and thick lines are for case 2.

Penalty parameters for case 2 (same in all scheduling points):

$$R_1 = \begin{bmatrix} 0\ 0\ 0\ 0\ 0 & 0 \\ 0\ 0\ 0\ 0\ 0 & 0 \\ 0\ 0\ 0\ 0\ 0 & 0 \\ 0\ 0\ 0\ 0\ 0 & 0 \\ 0\ 0\ 0\ 0\ 5 * 10^3 \end{bmatrix}$$

$$R_2 = \begin{bmatrix} 1 \end{bmatrix}$$

In Figure 8 the vehicle drives on a 5% road slope with warm discs. At distance 2500m the road slope increases to 6%. The CC in TM then increase the FB force resulting in a temperature transient peak value of around 380^0C and a reduction in speed. After this transient the disc temperature is controlled down to the reference value of 350^0C. At 6000m, 9000m, and 20000m the slope is increased to 8%, decreased to 6%, and 5% respectively, and the FB and AB forces are adjusted to achieve temperature control. Note that the stabilization is activated (reduction of the VEB torque ($M_{\rm veb}$)) at all the transients. Thin lines are for case 1 and thick lines are for case 2.

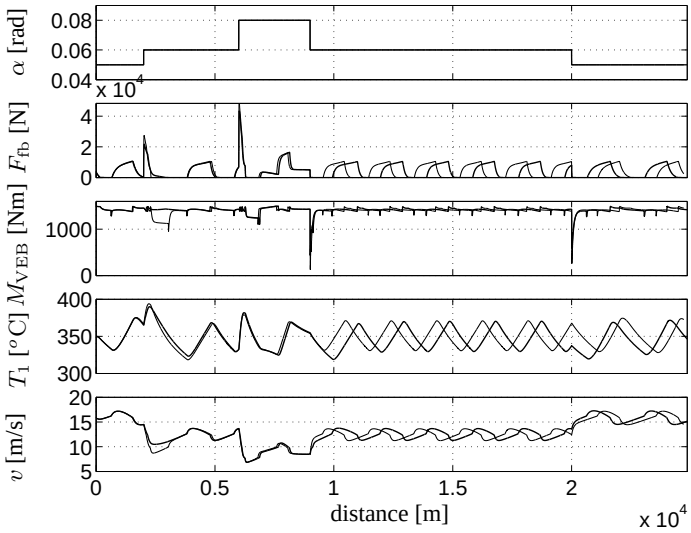


Fig. 9. Step responses for temperature mode with discrete gear box. Without feed–forward and hysteresis. A limit cycle arise due to the discrete gear box. Thin lines are for case 1 and thick lines are for case 2.

5.4. Simulations for full power mode and discrete gear box

In the previous sections the controller was derived based on a idealized, continuously controlled gear box. In reality there are gear boxes with continuously variable ratio, (**CVT**). CVTs are however not used in heavy duty vehicle applications. Therefore, since all gearboxes on heavy duty vehicles are of discrete type, the control of the engine speed must be reformulated. By calculating the continuous gear ratio, required to keep maximal engine speed, a discrete gear ratio is obtained by choosing the gear level closest to the continuous ratio. Simulation results, using the same vehicle and controller setup as in Section 5.3, and a discrete gearbox, are shown in Figure 9. The controller performance is deteriorated by the ratio error. When the road slope is constant at 5% a limit cycle arise where, for example, the FB temperature varies between 330^0C to 370^0C. Further, the controller reacts rather slowly to the disturbance in gear ratio. A remedy to this could be to feed forward the measurable ratio (force) error to the controller. This is done by adding the AB force difference between continuous and discrete gear box, ΔF_{gc}, to the requested FB force, F_{FBreq}, see Equation 24. Additionally, hysteresis was implemented in the sampling of N_g to N_{gd}. In Figure 11 a 5% hysteresis in the gear sampling is shown. The solid line maps N_g to N_{gd} if $\dot{N}_g > 0$ and the dashed line maps N_g to N_{gd} if $\dot{N}_g < 0$. In Figure 10 the step response for a

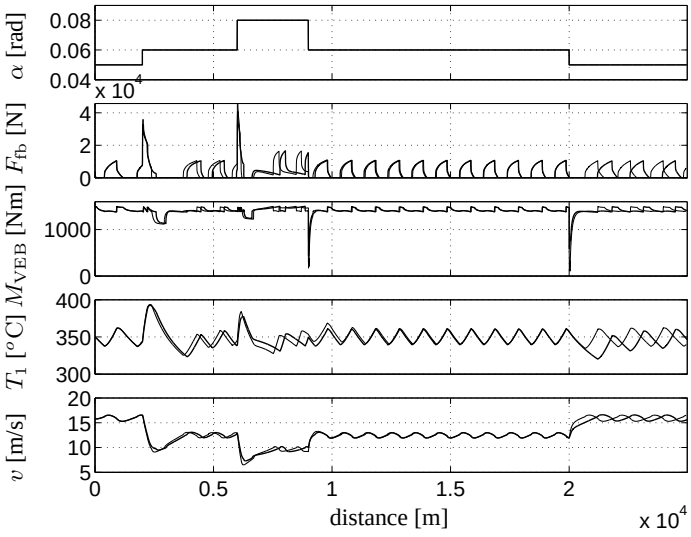


Fig. 10. Step responses for temperature mode with discrete gear box. Feed–forward and hysteresis are implemented. The temperature variations are reduced compared to when feed–forward and hysteresis where not used. Thin lines are for case 1 and thick lines are for case 2.

discrete gearbox using 1% hysteresis is shown. On the 5% constant slope section there is still a limit cycle present. The period distance for the FB temperature is around 1000 m and the variation is between 340^0C to 360^0C, i.e. a clear improvement compared to the case where no feed forward and no hysteresis were used.

$$F_{\mathrm{FBreq}} = F_{\mathrm{FBreq}} + \Delta F_{\mathrm{gc}}$$
$$\Delta F_{\mathrm{gc}} = \frac{(N_{\mathrm{g}} - N_{\mathrm{gd}}) M_{\mathrm{veb}} N_f}{R_w} \tag{24}$$

A more accurate temperature control can of course be achieved by setting the hysteresis in the gear shifting to 0 but then gear position is shifted to often. There is clearly a trade–off between accuracy in the temperature control and how often the gear is allowed to be shifted.

5.5. Robustness

As mentioned earlier controller robustness must be studied in both the time domain and frequency domain. The time domain analysis is very important in systems that, for example, have constraints in actuator amplitudes. Vehicle mass, road slope, and FB

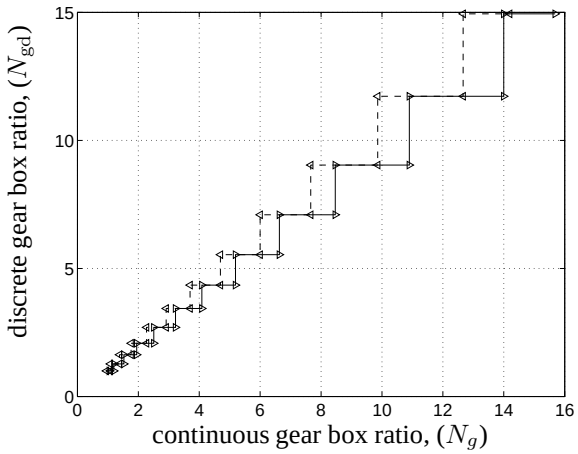


Fig. 11. N_g mapped to $N_{\rm gd}$ with 5% hysteresis. Solid line: increasing continuous ratio. Dashed line: decreasing continuous ratio

temperature are measured or estimated quantities used in the controller design and as feed–back signals. The controller must therefore handle perturbations in these signals.

Vehicle mass
In modern heavy duty trucks the mass of the whole combination is estimated in software. The accuracy of the estimation is affected by several variables like the quality of engine torque estimation and vehicle speed measurement. Onother problem in the estimation of vehicle mass is the possible lack of persistent excitation, i.e. the longitudinal forces acting on the vehicle do not have enough intensity to yield a perfect vehicle mass estimate [5]. Therefore, the TM controllers must robustly handle the modelling error introduced by the vehicle mass parameter. In Figure 12 simulations are shown for the same setup as shown in Figure 8. The only difference is that the vehicle mass in the controller design process has been varied (affecting the controller gains). The real vehicle mass is 60 tonnes and controllers where designed for 60, 50 and 70 tonnes ($\pm 17\%$). Clearly the control of the FB temperature is affected by the modeling error in the vehicle mass. However, the performance is rather good considering that the perturbation in vehicle mass (i.e. vehicle mass estimation error) is large (17%).

Road slope
The perturbed road slope , $\alpha_{\rm pert}$, is modelled as filtered white noise, α_ν, added to a filtered road slope process, $\alpha_{\rm f}$, see Equation 25 were $T_\nu = 10^3$s and $T_\alpha = 5s$.

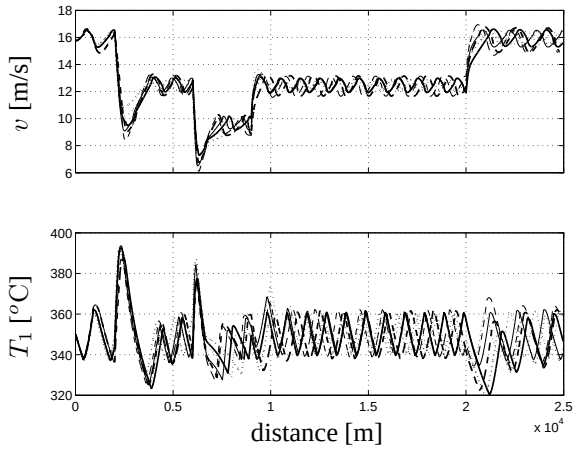


Fig. 12. Robustness to variations in the vehicle mass. Thin lines are for case 1 and thick lines are for case 2. Solid line: 60 tonnes, Dashed line: 50 tonnes, and Dotted line: 70 tonnes. Real vehicle mass is 60 tonnes.

$$\begin{aligned} \alpha_{\rm pert} &= \alpha_\nu + \alpha_{\rm f} \\ \alpha_\nu &= -T_\nu \frac{{\rm d}\alpha_\nu}{{\rm ds}} + \nu \\ \alpha_{\rm f} &= -T_\alpha \frac{{\rm d}\alpha_{\rm f}}{{\rm ds}} + \alpha \end{aligned} \tag{25}$$

In Figure 13 the effect of measurement errors in the road slope are shown. The accuracy in the control of the FB temperature is deteriorated for both case 1 and case 2. There is, for example, a rather large overshoot in the the FB temperature and a corresponding undershoot in the vehicle speed at 6000m. During this transient (also at 3000 m) the stabilizing action is active in case 1. The stabilizing action introduces a modeling error into the state equations used in case 1. This modeling error does not appear directly through the AB force but rather indirectly from changing stationary scheduling points. In case 2, the modeling error caused by the stabilizing action does not, as seen in simulations, affect the control as much as in case 1. When the AB torque is changed, an error is introduced into the control signal in case 2 since the time derivative of $M_{\rm ABmax}\omega_{\rm engmax}$ is not zero. The result is acceptable considering that the perturbation in the road slope is rather large with peaks of $\pm 50\%$ and long periods of time around $\pm 20\%$.

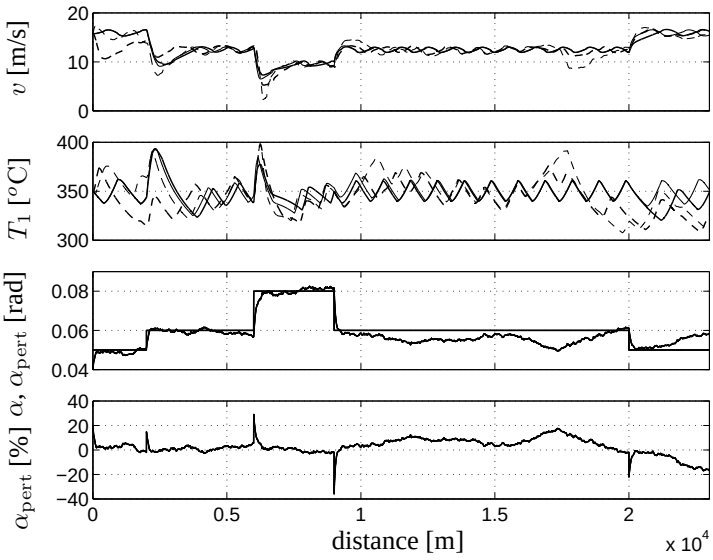


Fig. 13. Robustness to perturbation in the road slope. Thick line: case 2, Thin line: case 1. Solid line: No perturbation, Dashed line: Perturbed road slope.

FB temperature

Typically the FB temperature is an estimated system state like the vehicle mass, and therefore it is important that the TM controllers can handle errors in the foundation brake temperature. By altering parameters in Equation 3 a new FB temperature model is derived. This new model is then used to create a perturbation in the state vector $x_{\mathrm{tot}_{1\ldots5}}$. In Figure 14 the effect of a perturbation on the FB temperature dynamics is shown. The thin solid line is for the non–perturbed case and the thick solid line and dashed line are for the perturbed case. Clearly, a bias in the FB temperature measurement results in a bias in the FB temperature. Also here the stabilizing action is active at 3000 m and 6000 m.

Friction between brake pad and brake disc

Slow variations in the friction between brake pad and brake are caused by several factors. Typically, a surface coating that reduces the friction is created if the power input to the brake has been low for long periods of time. Another surface coating that lowers the friction is rust on the disc due to road salt, high humidity etc. . Therefore, the brake controller must perform in a robust way even when there is a disturbance in friction. In Figure 15 the friction has been reduced by 20% for case 1 and case 2. Also

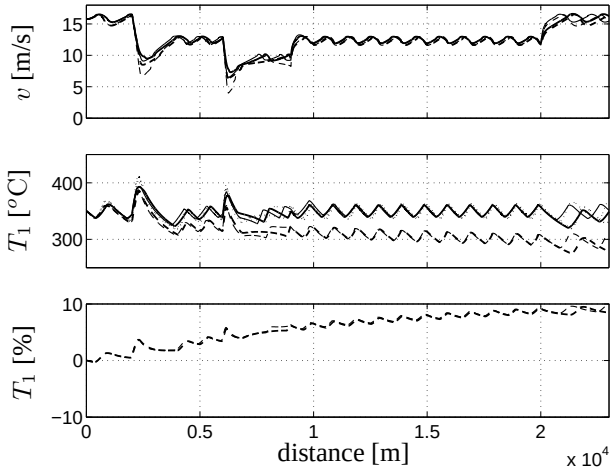


Fig. 14. Robustness to perturbation in the FB temperature dynamics.Thick line: case 2. Thin line: case 1. Solid line: No perturbation. Dashed line: Real FB temperature when perturbed. Dotted line: Perturbed FB temperature dynamics.

the non perturbed system is shown. Both controllers are stable but there is a rather large undershoot in the vehicle speed at 6000 m for case 1. Also here the stabilizing action is active at 3000m and 6000m

6. SIMULATIONS USING REAL ROAD PROFILES

Measured road profiles from the french alps (Isère) are used to verify the functionality of the MSCC. In Figure 16, 17, 18, and 19 simulations are shown for the MSCC on four different road profiles. These road profiles are very challenging due to sections of very steep decents combined with a high rate of change in slope. All simulations are done with the mode switching algorithm in Section 3, case 2 TM controller (velocity based scheduling), driver set speed $v_{\rm ss} = 15{\rm m/s}$, and FB reference temperature $T_{1{\rm ref}} = 350^{\rm o}{\rm C}$. On the first road, Figure 16, the VM is active all the time, i.e. there is no need to reduce the speed under the driver selected set speed. In Figure 17 the TM is activated two times. The first activation is due to high utilization of FBs on the downhill slope between 2000 m to 3500 m causing the vehicle speed down from 15m/s to approximately 9m/s. There is also a transient in FB temperature that is rather high (30% overshoot) due to a very fast change in road profile combined with a high set speed and heavy combination (60 tonnes). In the second activation of TM there is almost no change in vehicle speed, i.e. in this case the reference temperature cor-

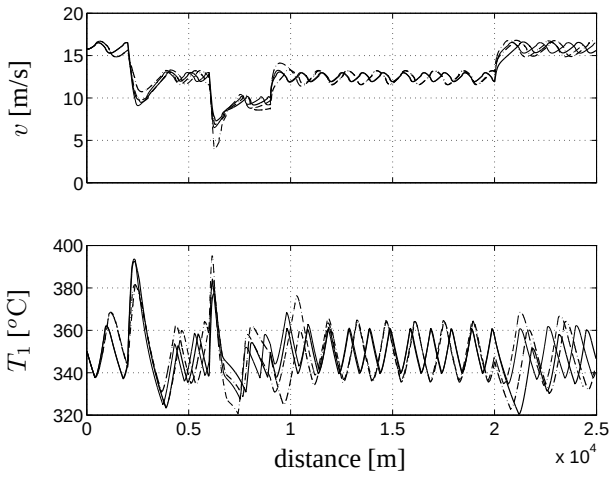


Fig. 15. Robustness to variations in the friction between brake pad and brake disc. Thin lines: case 1. Thick lines: case 2. Solid lines: No perturbation. Dashed lines: 20% decrease in friction.

responds to the set speed for this particular road profile. Infact, due to the decrease in downhill slope starting at 5800 m, the TM controller increases the vehicle speed. This causes a switch to VM (sub mode 3 in Section 3) at 5950 m, producing a small overshoot in vehicle velocity. On the third road, Figure 18, once again the TM is never activated and the driver set speed can be kept at all time using the complete brake system, even if the initial FB temperature is rather high. The road profile in Figure 19 is very demanding causing a 30 % overshoot in FB temeprature. At 4600 m the TM is activated but since the downhill slope decreases fast directly after this switch no reduction in vehicle speed made. Instead, like in Figure 17, the VM is activated at 4800 m to avoid overspeeding. At approximately 5500 m a very steep downhill slope starts, causing a switch to TM and a rather high overshoot in FB temeprature. Overall the performance is very good for the MSCC algorithm. Switching between VM and TM ensures that FB fading is avoided with good marginal (fading starts at $600^0\mathrm{rmC}$) and that the same time the complete brake system is utilized to the fullest to ensure good driver comfort and transport efficiency.

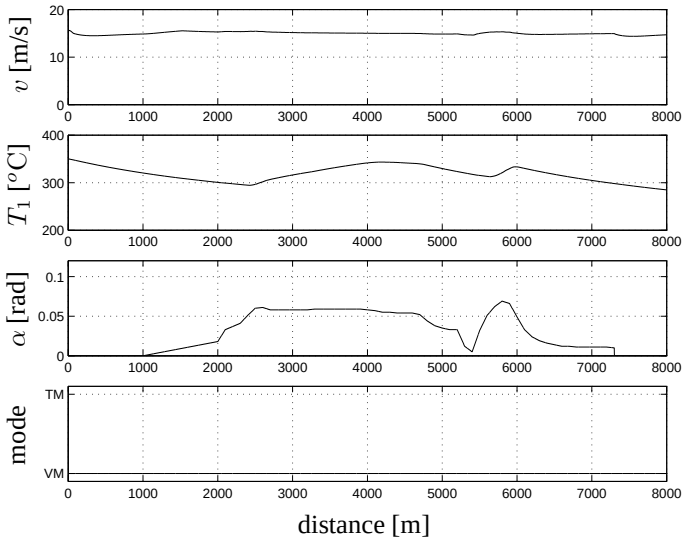


Fig. 16. MSCC on Isère 1 road. Set speed is 15m/s. The VM mode is always active and the driver selected set speed can be kept during the whole slope even if the foundation brake initial temperature is rather high.

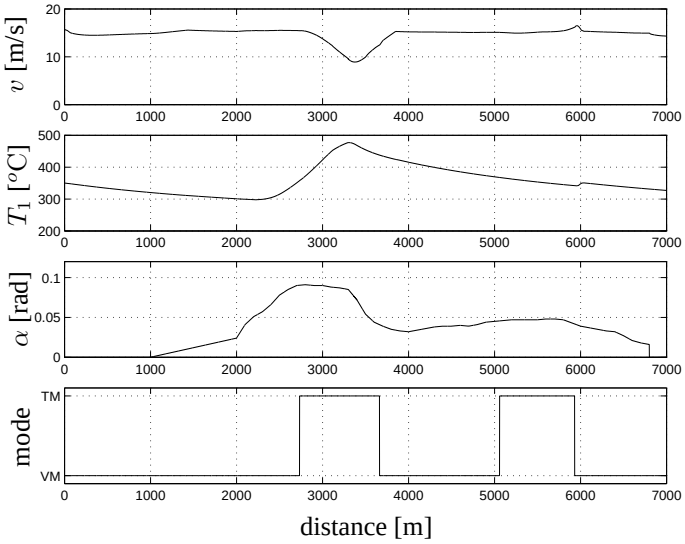


Fig. 17. MSCC on Isère 2 road. Set speed is 15m/s. When the slope increases drastically a switch to temperature mode is made. The maximum temperature transient is 30% over the reference temperature.

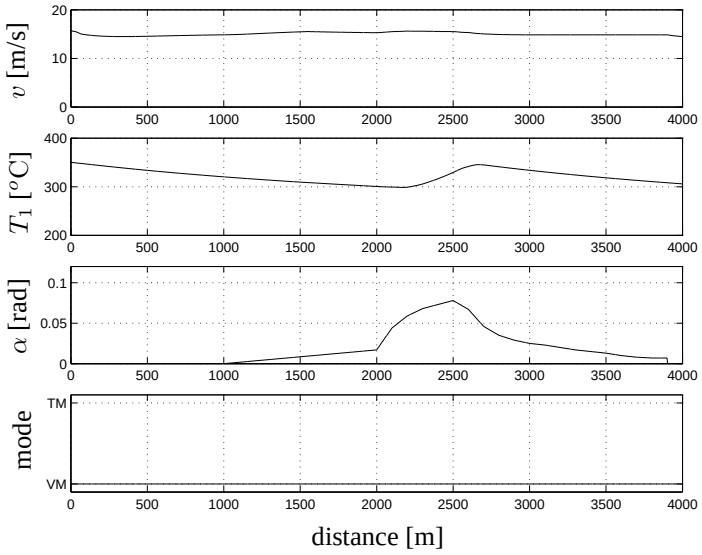


Fig. 18. MSCC on Isère 3 road. Set speed is 15m/s. The VM mode is always active and the driver set speed can be kept during the whole slope even if the initial foundation brake temperature is high.

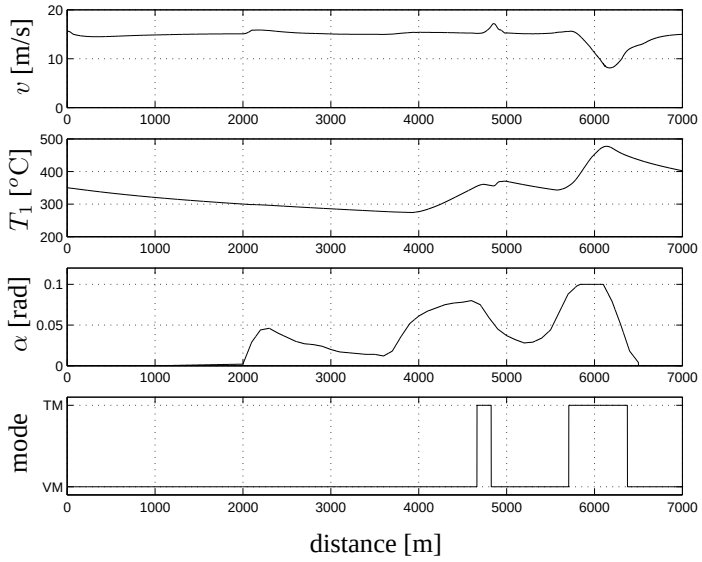


Fig. 19. MSCC on Isère 4 road. Set speed is 15m/s. The road slope varies much causing a switch to temperature mode two times. The maximum temperature overshot is 30% over the reference temperature.

7. DISCUSSION

The main objective of this paper has been to include the FBs into the framework of constant speed controllers in order to increase the continuous braking capacity of heavy duty vehicles. Safety and transport efficiency are increased by automatically switching control strategy between constant speed and constant FB temperature. Additionally, also the complexity of the drivers task is reduced. Special attention is given the constant temperature controller and the switching strategy. Linear quadratic control (LQ) is a natural candidate for the temperature controller synthesis since it offers a way of weighting between control signal activity (air consumption) and deviation from reference value (FB temperature). Building a family of linear controllers using gain scheduling, is a well known strategy to handle non–linearities that has proven to be very efficient in many practical applications. Another, very important feature for the future, is that LQ control provides a straight forward way of including road profile preview by expanding the state vector with samples of the upcoming road slopes. Road profile preview will probably be available in the future using e.g. GPS and road topology maps. The switching strategy will also benefit from preview information which, for example, can be included into submode 2 to further improve the judgment of the vehicles steady state capacity.

A comparison between using only road slope (case 1) and using road slope together with vehicle speed (case 2) as scheduling variables is made. The transient performance of case 1 is lower than that of case 2. This performance difference is believed to arise from the scheduling strategy rather than from the tuning of the individual LQ controllers. The model used in case 1 is perturbed in transient manoeuvrers whereas the model used in case 2 accurately describes the process also during transients. The stabilizing action handles the instability well but causes some problems in case 1. When the AB torque (power) is reduced, the stationary values that are used in the scheduling of case 1, are altered. This is not compensated for in the scheduling. For example, in Figure 15, the stabilizing action is active between 6200m and 6800m. In this interval there is a large undershoot in the vehicle speed and an overshoot in the FB temperature. Compensating for the stabilizing action is not easy since it would reintroduce a fast scheduling variable (VEB power) into case 1.

Since the payload varies in real applications, also the mass must be considered as an additional scheduling variable. This is rather trivial to achieve since accurate estimations and measurements (air suspension) of the total mass are available in real time.

In a real truck application the temperature dynamics of the different FBs of the vehicle are not the same. Different types of FBs (larger or smaller disc brakes) can be mounted on different axles, and the ability to dissipate energy can vary depending on the installation of the FB (e.g. air flow). Typically, the truck is equipped with the same type of FBs on all axles and the rear axle has the lowest cooling capacity due to the

installation (less air flow etc.). One way to solve this is to expand the current SISO controller to a MIMO controller where each FB force and temperature is controlled. Another solution could be to control the temperature of only one axle (master), and then distribute the FB force according to some predefined values. The master axle and the FB force distribution should be chosen in a conservative way to ensure that no axles are overheated.

It should also be mentioned that the framework can be expanded to also cover more advanced cruise controllers. One example of this are adaptive cruise controllers (ACCs), that utilizes ABs and FBs to keep a predefined constant distance to the vehicle in front. One common feature of most of theses systems is that the FBs are utilize conservatively, only for short durations of time (without FB temperature feed–back) to ensure that no heat fading occur. By instead utilizing the mode switching strategy presented in this paper, both performance and safety are improved.

8. CONCLUSIONS

Controllers and switching strategies have been developed to coordinate auxiliary brakes, foundation brakes, and gear box. The controller synthesis, based on a non–realistic continuously variable transmission assumption, is shown to work well also for a realistic discrete transmission. It can be concluded that the performance of the retardation system can be increased, even compared to skilled drivers, by utilizing the complete brake system of the vehicle. Additionally, safety critical fading situations can be avoided by controlling the foundation brake temperature. Usage of real measured road profiles and extensive simulations show that the algorithm is robust to realistic disturbances.

REFERENCES

1. P. Lingman. Integrated retardation control-thesis for the degree of licentiate of engineering. Chalmers University of Technology, 2002.
2. P. Lingman and M. Wahde. Integrated retardation control using neural networks with genetic algorithms. In *Proceedings of the 6th International Symposium of Advance Vehicle Control, AVEC, Hiroshima*, volume 6, pages 751–756, 2002.
3. B. Schmidtbauer. *Modellbaserade Reglersystem*. Studentlitteratur, Lund, Sweden, 1999.
4. L. Ljung and T. Glad. *Control Theory - Multivariable and Nonlinear Methods*. Taylor and Francis, 2000.
5. P. Lingman and B. Schmidtbauer. Road slope and vehicle mass esimation using kalman filtering. In *Supp. to Vehicle Systems Dynamics*, volume 37, pages 12–23, 2001.
6. KJ Astrom and B Wittenmark. *Adaptive Control-Second Edition*. Addison-Wesley Publishing Company, Inc, 1995.
7. D.J. Leith and W.E. Leithead. Gain–schedule and nonlinear systems: Dynamic analysis by velocity–based linearisation families. *International Journal of Control*, 70:289–317, 1998.

Paper 4

Open Loop Optimal Downhill Driving Brake Strategies Using Nonlinear Programming

Peter Lingman, Antóni Duarte, and Ricardo Oliveira

Submitted to International Journal of Heavy Vehicle Systems, 2005

Open Loop Optimal Downhill Driving Brake Strategies Using Nonlinear Programming

Peter Lingman*
Antóni Duarte
Ricardo Oliveira

Department of chassis and vehicle dynamics, Volvo 3P,
Gropegårdsgatan, 405 08 Göteborg, SWEDEN
Phone +46 31 327 4342. E-mail: peter.lingman@volvo.com
*Corresponding author

Abstract: In this paper the development of open loop optimal trajectories for downhill driving is described. Problem formulation including process modeling, model simplification, transformation into a finite optimization problem, and implementation into the TOMLAB optimization package are presented. Non–linearities introduced by the vehicle model, environment model (road slope), and the discontinuous behaviour of the gearbox are overcome by carefully smoothening and scaling the problem functions, and by supplying TOMLAB with analytical derivatives. Results show that there is a large potential in controlling the complete brake system of a heavy duty truck and thereby simultaneously improve both mean speed (transport efficiency) and component wear cost. The resulting optimal trajectories define the upper limit for what is theoretically achievable in a real, closed loop, controller implementation and can be used to both inspire and verify the development of such algorithms.

Keywords: cruise control, downhill driving, fading, auxiliary brakes, foundation brakes, nonlinear programming, optimal control

1 Introduction

Downhill driving with a heavy duty vehicle requires several actions from the driver. First of all a proper set speed must be decided upon. To obtain and keep this set speed the driver has to change gear and engage both foundation brakes (**FB**) and the auxiliary brake (**AB**). A problem is that the driver has almost no feed-back regarding vehicle mass and brake disc temperature. A common feature with most kinds of FBs is that they lose their brake capability when overheated (a phenomenon known as heat fading). Accidents have occurred in which a driver has chosen too drive to fast downhill, resulting in overheating of the FBs. Other more

conservative drivers choose a low speed to ensure that the system is within safety limits. The drawback is of course a low mean speed. Another important aspect of strategies for optimal retardation force distribution is how the tyre, pad, and disc wear are affected. FBs distribute the total retardation force demand on all vehicle axles whereas ABs use the drive axle only, see Fig. 1. The standard strategy is to utilize ABs (also known as *non-wear brakes*) in order to save brake pads and discs. However, some drivers have noticed a very high drive tyre wear causing high maintenance cost when using this strategy. Therefore the trade-off between tyre, pad, and disc wear cost is investigated here.

Earlier work in this area has been reported in [1], [2], and [3]. In [2] an optimization of a feed-forward neural network using evolutionary algorithms was performed. It was concluded that it was possible to reduce substantially the wear cost and increase the cruising speed when compared to a normal driver. However, the trade–off between mean speed and component wear cost was not studied. Due to the nature of both the optimization algorithm and the rather complex model, it was not possible to conclude that the results obtained where globally optimal. The potential in improving mean speed and wear cost by further integrating the total brake system was, however, revealed when comparing the results to the strategy used by skilled drivers.

This paper focuses on open loop optimal behaviour (optimal control) rather than on achieving a closed loop feed-back like in [2]. The objective here is to state what is theoretically possible to achieve when all states and controls (speed, acceleration, brake force, road slope etc.) are known a priori. Measured road profiles are used for the generation of driving strategies that are wear cost optimal and transport efficient (high mean speed) and Pareto curves are used to explain how wear cost is affected by mean speed, i.e. multi–objective optimization is used. Furthermore, the different optimization results presented are always compared to the behaviour of real drivers. The results presented in this paper can be used to both compare and develop implementable closed loop control strategies for heavy duty vehicle downhill driving.

2 Brake system functionality and components

Figure 1 shows the various components on a Volvo truck's retardation system: foundation brakes (**FB**), Volvo Engine Brake (**VEB**) and Compact Retarder (**CR**). The foundation brakes are able to handle high braking power demands such as emergency stops as well as low power demands. FBs on a truck are commonly of the disc brake type. Trailers are usually equipped with drum brakes. However, disc brakes are believed to replace completely drum brakes in the future also for trailers. Therefore only FBs of disc brake type are considered here. When FB are applied the brake pads are pressed against a disc that rotates with the wheels. The friction between the pads and discs generates a braking torque. The energy dissipated by friction is converted into heat and the temperatures of the pads and disc increases. When the disc temperature a few mm under the disc surface is around 600°C the friction coefficient between pads and disc begins to decrease and consequently the braking capability is reduced. For temperatures around 800°C the braking force becomes dangerously low and around 900°C the pads burn. This

reduction of friction is usually referred to as heat fading. Almost all new Volvo Trucks are equipped with a VEB. It is an auxiliary braking (**AB**) system intended for low power retardation, like downhill cruising. Since it is mounted in front of the gearbox and behind the engine it is often called primary retarder. When the VEB is engaged an exhaust valve permits the exhaust gases from exiting the engine and thereby creating a brake torque. Additionally also the valve timing is changed to further improve the brake performance, which then, via the gearbox and final gear, is applied on the driven wheels. The energy is converted into heat and is dissipated to the exhaust gases and the engine cooling system. Approximately 40% of the heat is transferred to the engine cooling system and 60% is transmitted to the exhaust gases. The CR is also an AB and it is called secondary retarder (or driveline retarder) because it is mounted behind the gearbox. Note that both ABs only have the ability to brake the driving axles. The CR used in Volvo trucks is of the hydrodynamic type. The braking torque provided by this type of retarder results from the oil friction between the stator and rotor (connected to the driveshaft). The output torque is pneumatically controlled by the amount of oil inside the retarder housing. The heat generated by the oil friction is dissipated in a heat exchanger connected to the engine cooling system. It is assumed that all the braking energy dissipated in the CR is transferred to the coolant. Since both ABs dissipate energy to the coolant their usage is limited by the cooling power available. Thus the cooling system must also be taken into account as part of the integrated retardation strategy. The cooling power depends on the coolant flow and air flow through the radiator. The coolant flow is generated by a pump connected to the crankshaft and therefore it is proportional to the engine speed, and the air flow depends on the vehicle speed and fan speed. The fan is connected to the engine via a viscous coupling (**VC**). Engine speed and the controllable slip in the viscous coupling determine the fan speed. The continuous generation of high braking forces by the CR can overload the cooling system, i.e. the CR can generate more heat than can be handled in the cooling system. Therefore the force must be controlled to prevent the coolant from reaching the boiling temperature. In reality this controller acts on both coolant temperature and oil temperature inside the CR. However, in order to reduce the number of variables in the optimization, a simplified control strategy is used here. Figure 4 (right) shows the fraction of maximum force that can be provided as function of the coolant temperature at the output of the retarder heat exchanger (T_{cr}). In reality this means the CR can only be used at full load for a short period of time ($\approx$ 10–20 seconds) before it starts to be down–regulated.

One further requirement of the cooling system is to keep the temperature of the coolant in the engine between narrow limits, around the best operating point. When the load placed on the cooling system is low it is possible that the coolant temperature at the radiator outlet drops significantly. To prevent the coolant from going into the engine block with a low temperature there is a thermostat that allows some fraction of the hot coolant to bypass the radiator. The thermostat is adjusted in such a way that after the junction of the two coolant fractions the temperature is within the desired limits. A more extensive description of state-of-the-art retardation systems of modern trucks is presented in [3].

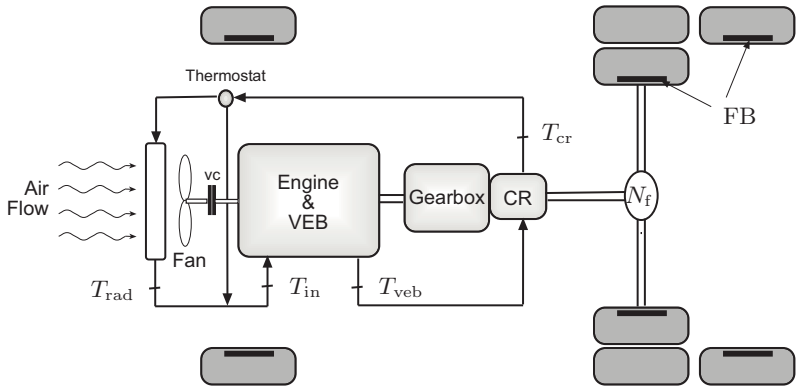


Figure 1 Volvo Retardation System (6x2 tractor)

2.1 Vehicle model used for optimization

In order to calculate optimal control trajectories a mathematical model of both the vehicle and environment must be made. It is also important to keep the model simple, e.g. to avoid hard non–linearities and to keep the number of states to a minimum. We start with the longitudinal force balance of the vehicle.

Longitudinal motion equation

$$
\begin{aligned}
m\dot{v} &= F_{\text{eng}} - F_{\text{air}} - F_{\text{roll}} - F_{\text{grade}} - \\
&\quad -F_{\text{VEB}} - F_{\text{CR}} - F_{\text{FB}} \\
F_{\text{air}} &= C_{\text{air}} v^2 \\
F_{\text{roll}} &= C_{\text{roll}} m \\
F_{\text{grade}} &= mg \sin\alpha
\end{aligned} \tag{1}
$$

F_{eng} is the propulsion force (assumed to be zero here), F_{air}, F_{roll}, and F_{grade} are resistance forces and F_{VEB} and F_{CR} are the AB forces. F_{FB} is the FB force. The dynamics of the brake actuators are assumed to be very fast (static) compared to the other system dynamics. This means that, for example, the pressure build–up time in the FBs is neglected, i.e. a request is immediately executed.

Foundation brake model

The temperature dynamics of the FB is a distributed process that, when represented on ODE form, has to be discretized. Here two masses are used, represented by T_1 and T_2. In Figure 2 the position of temperatures T_1 and T_2 are shown. T_1 represents a temperature 3 mm under the disc surface at radius of 170mm. T_2 represents a temperature 15 mm under the disc surface at a radius of 32mm, i.e. 3 mm from the spline that mounts the disc to the axle. Heat fading is represented by T_1 which also is the temperature used in the parametrization of the pad wear model in Equation

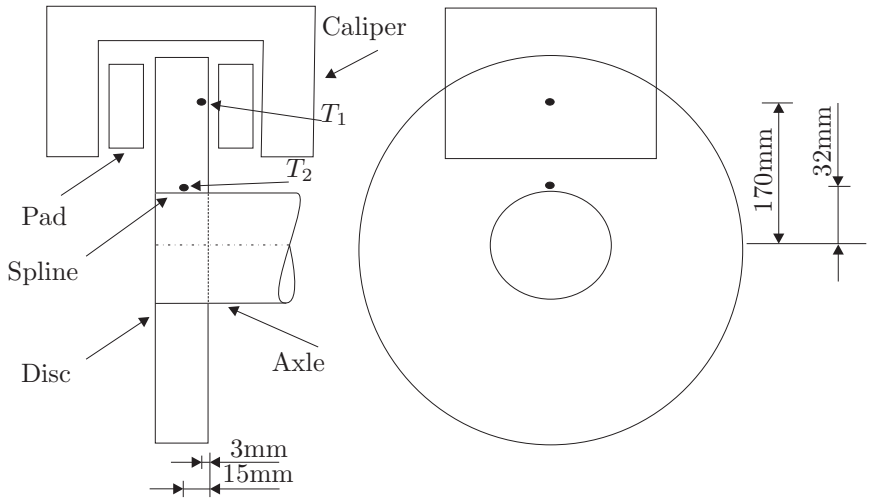


Figure 2 Foundation brake temperatures T_1 and T_2. T_1 is placed close to the surface at the efficient brake radius, i.e. almost in the middle of the pad and disc contact surface. T_2 is placed close to the spline where the disc mounts to the axle

14. The temperature dynamics of the FB is shown in Equations 2 and 3.

$$\dot{T}_1 = q_{\text{in}} - c_1(T_1 - T_2) - c_2(v)(T_1 - T_{\text{amb}}) - c_3(T_1^4 - T_{\text{amb}}^4) \tag{2}$$

$$\dot{T}_2 = c_4(T_1 - T_2) - c_5(v)(T_2 - T_{\text{amb}}) - c_6(T_2^4 - T_{\text{amb}}^4) \tag{3}$$

$q_{in} = F_{\text{FB}}v$ is the FB brake power, c_1 and c_4 are heat conduction constants, $c_2(v)$ and $c_5(v)$ are vehicle speed dependent convection variables, and c_3 and c_6 are radiation constants. Different temperature dynamics can be assigned to different axle installations on the truck. For example, due to differences in air drag, the FBs on the rear axle can dissipate less energy compared to the front axle. However, in order to simplify the optimization (reduce the number of states and controlls) the FB force is assumed to be distributed evenly between the axles and it is also assumed that all axles of the combination have the same temperature dynamics.

Auxiliary brake model
Since the fast dynamics of the actuators is neglected, the only dynamics to be handled for the ABs is the temperature regulation of the CR as seen to the right in Figure 4. For this a model of the cooling system is needed.

$$\dot{T}_{\text{veb}} = \frac{1}{\tau_{\text{veb}}}\left(\frac{\gamma F_{\text{veb}} v}{q_{\text{water}} Cp_{\text{water}}} + T_{\text{rad}} - T_{\text{veb}}\right) \tag{4}$$

$$\dot{T}_{\text{cr}} = \frac{1}{\tau_{\text{cr}}}\left(\frac{F_{\text{cr}} v}{q_{\text{water}} Cp_{\text{water}}} + T_{\text{veb}} - T_{\text{cr}}\right) \tag{5}$$

$$\dot{T}_{\text{rad}} = \frac{1}{\tau_{\text{rad}}}\left(q_{\text{water}} Cp_{\text{water}}(T_{\text{cr}} - T_{\text{rad}}) + \varepsilon(T_{\text{eb}} - T_{\text{cr}})\right) \tag{6}$$

$$\varepsilon = av + bw_{\text{eng}} + c \tag{7}$$

T_{veb}, T_{cr} and T_{rad} are temperatures in three different points of the cooling system, as represented in Figure 1. T_{eb} is the air temperature in the engine bay, assumed constant. The coolant flow is proportional to the engine speed: $q_{\text{water}} = kW_{\text{eng}}$. In equation (6), ε is a heat transfer coefficient modeled according to the ε–NTU method, see [4] and [5]. Although it depends on the fan speed, vehicle speed and coolant flow it can be, in fact, reduced to a function of only engine speed and vehicle speed. The engine propels the fan via a viscous coupling. It is assumed that the fan speed is always at the maximum speed allowed by the viscous coupling (minimum slip) as this results in a higher cooling capability. In order to further reduce computational load the heat transfer coefficient is modeled as a plane according to Equation 7. Parameters a, b, and c are chosen conservatively rather than by minimizing the least square error. This guarantees that the cooling power is estimated on the side of caution and the maximum error (largest underestimation of cooling power) is very low ($\approx 3\%$) in the speed range relevant for downhill driving (10 to 25 m/s). The thermostat prevents the coolant to enter the engine with a temperature lower than a set value (T_{ts}). When the thermostat is fully open all the coolant flows through the radiator and thus the temperature at the engine coolant inlet (T_{in}) is the output of the radiator (T_{rad}). When some of the coolant flow is bypassed the engine temperature would, ideally, be kept at the desired level. Assuming an ideal behaviour and neglecting the dynamics of the thermostat the differential equation for the coolant temperature going into the engine is given in branches:

$$\dot{T}_{\text{in}} = \begin{cases} \dot{T}_{\text{rad}} & \text{if } T_{\text{rad}} > T_{\text{ts}} \text{ (Fully open, all water through radiator)} \\ 0 & \text{if } T_{\text{rad}} < T_{\text{ts}} \text{ (Otherwise)} \end{cases} \tag{8}$$

To avoid the discontinuity imposed by such an equation the effect of the thermostat was replicated with a different approach. Equation (6) was rewritten into an inequality constraint:

$$\dot{T}_{\text{rad}} \geq \frac{1}{\tau_{\text{rad}}} \left(q_{\text{water}} Cp_{\text{water}} (T_{\text{cr}} - T_{\text{rad}}) + \varepsilon (T_{\text{eb}} - T_{\text{cr}}) \right) \tag{9}$$

and a lower bound on T_{rad} was set, with the value of T_{ts}. The effect of the inequality is that T_{rad} can take values greater than those given by (6) and so its value can be artificially kept at T_{ts} when required. Obviously when relaxing the equality in equation (6) the trajectory of T_{rad} can go up to arbitrarily high values and this effect will propagate to the temperatures downstream the engine coolant outlet, and thus limiting the usage of the auxiliary brakes. However, this does not occur. If the cooling system becomes saturated the optimal behaviour is to keep the cooling system temperatures as low as possible, i.e. satisfying (9) as equality. Otherwise the real trajectories for the cooling system temperatures can only be obtained by validation of the results.

The principal torque characteristics of the VEB and CR are given in Figures 3 and 4 (left) respectively. To reduce computational load the VEB torque as function of engine speed was approximated with a second order polynomial and engine

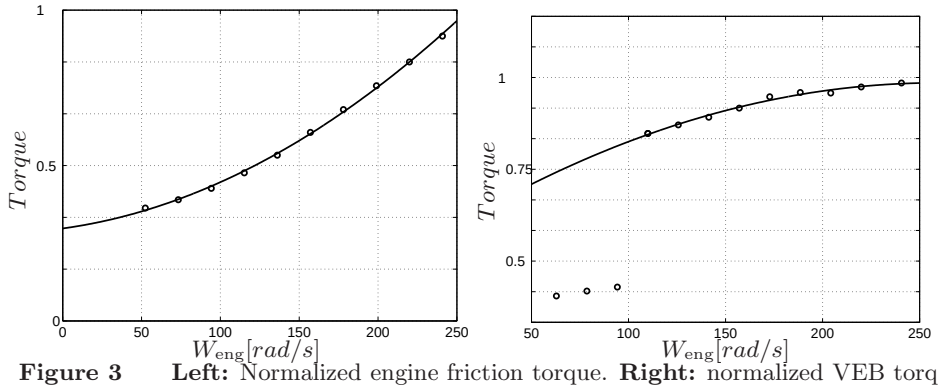


Figure 3 **Left:** Normalized engine friction torque. **Right:** normalized VEB torque

speeds below 100 rad/s were not considered since this is not a feasible engine operating range. The CR upper torque limit characteristics were approximated as the intersection between two semi planes in the intervals [5,16] m/s and [16,32] m/s, i.e. vehicle speed above 32m/s was neglected. Note also that in Figure 4 (left) four different torque characteristics are shown for the CR. This is only to illustrate how different air pressure levels (0.65, 1.2, 1.75, and 2.65 bar) in the retarder oil compartment alter the brake torque, 2.65 bar produce always the maximum torque.

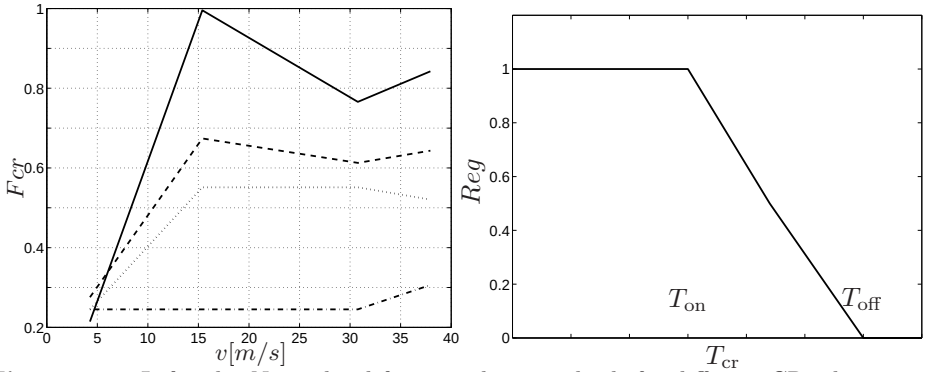


Figure 4 Left side: Normalized force on driving wheels for different CR oil pressures (0.65, 1.2, 1.75, 2.65 bar). Right side: Regulation of the CR force based on its temperature output. When the output temperature (T_{cr}) reach T_{on} the regulation starts and it the temperature reach T_{off} the CR is turned off

As can be seen in Figure 1 the VEB brake torque is transmitted to the driving wheels via the gearbox and rear axle, i.e. the resulting drive wheel torque will depend on the gearbox ratio (N_{g}) and final ratio (N_{f}). The maximum AB brake power is achieved at maximum engine speed since the VEB peak torque occurs at the upper limit of the engine speed as seen in Figure 3. Furthermore, both coolant flow and fan speed increases with engine speed, resulting in higher cooling performance and therefore improved AB performance. In Figure 5 maximum VEB force is shown as a function of vehicle speed. It can be seen in the left figure that the gearbox introduces a hard non–linearity into the problem since only a finite number of gearbox ratios are available. To model this behaviour a variable that can only take discrete values can be introduced. This is, however, not possible

from an optimization point of view since it would require mixed–integer solvers, which are not designed for large scale NLP problems and can therefore typically not be used to solve optimal control problems, which always involve a large number of parameters. Therefore the discontinuous upper bound on the VEB force must be removed. One simplification to overcome this problem is to consider the gear ratio as continuously variable, i.e. a continuously variable transmission (**CVT**). This means that the engine speed would be kept at a constant value where the ABs performance is maximum. The gear ratio is therefore a function of vehicle speed and the VEB force would be like in Equation 10 where the engine speed, W_{eng}, is constant. This would result in an upper bound for the VEB force according to the thin solid line in the right figure in Figure 5. Additionally, a lower bound must be set in order to avoid propulsion from the VEB. Also the heat transfer coefficient (Equation 7) is changed to $\varepsilon = dv$ (d constant) which is a very good approximation in the speed range from 10 m/s to 25 m/s. For speeds below 10m/s a small underestimation of the cooling performance is made (5% at 5m/s).

$$F_{\mathrm{vebmax}} = \frac{N_{\mathrm{g}} N_{\mathrm{f}}}{R_{\mathrm{w}}} T q_{\mathrm{veb}}(W_{\mathrm{eng}}), \quad \text{with} \quad W_{\mathrm{eng}} = \frac{v N_{\mathrm{g}} N_{\mathrm{f}}}{R_{\mathrm{w}}} \tag{10}$$

$$N_{\mathrm{g}}(v) = -k_{\mathrm{i}}(v - \bar{v}_{\mathrm{i}})^r + N_{\mathrm{g_i}}, \quad \textit{for} \quad v \in [v_{\mathrm{b_{i-1}}}, v_{\mathrm{b_i}}] \tag{11}$$

Even if the CVT approach simplifies the problem significantly it can be seen to the right in Figure 5 that an overestimation of achievable VEB force occurs. To overcome this problem the the VEB force upper bound discontinuous function (left in Figure 5) were approximated with a smooth function (thick line to the right in Figure 5) according to Equation 11 . Where i is the index of the approximated gear, r is an odd integer that controls the non–linearity of the approximation, $\bar{v}$ is the mean value of two speeds that give maximum engine speed in consecutive gears, N_{gi} is the gear ratio, v_{bi} are the speeds that correspond to the break points caused by gear shifts in Figure 5, and finally k_{i} is a scalar that chosen together with v_{bi} and N_{gi} give continuous first order derivatives. Even if this ensures a smooth first order continuous approximation the hard non–linearity still give serious problems for the optimization routine (see next chapter). Increasing the parameter r to improve the approximation forces the derivative around gearshifts to approach infinity and hence the convergence to optimality becomes very slow. Note also that the approximation is non–convex. It is not possible to improve the smoothness of the VEB force upper bound without overestimating the AB performance. The approach taken here was to perform one optimization with a relatively low value on r to generate a feasible trajectory for the gear shifting. Then a multi-phase problem is solved using the gear ratios for the gear trajectory computed before (see next chapter). For each phase the gear ratio is a constant value and the maximum VEB force becomes:

$$F_{\mathrm{vebmax}} = \frac{N_{\mathrm{g_k}} N_{\mathrm{f}}}{R_{\mathrm{w}}} T q_{\mathrm{veb}}(W_{\mathrm{eng}}), \quad \text{with} \quad W_{\mathrm{eng}} = \frac{v N_{\mathrm{g_k}} N_{\mathrm{f}}}{R_{\mathrm{w}}} \tag{12}$$

where $N_{\mathrm{g_k}}$ is the gear ratio for the k-th phase. Note that the resulting constraint is the equation for the dashed line in Figure 5 corresponding to the gear in that phase. Since the gear ratio is fixed in each phase it is also necessary to introduce

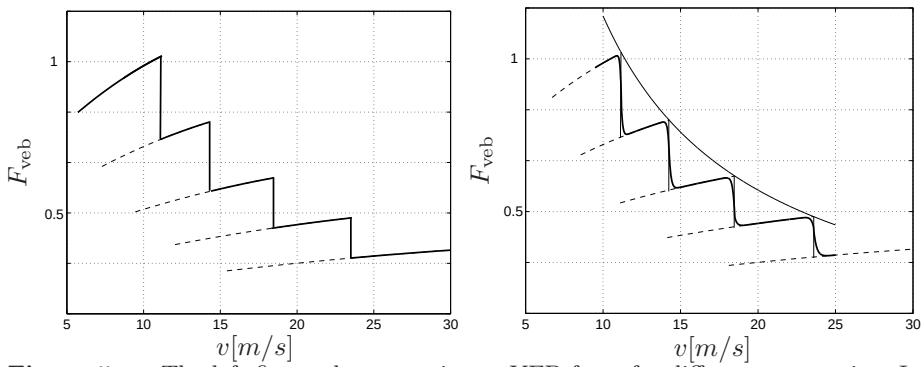


Figure 5 The left figure shows maximum VEB force for different gear ratios. In the right figure a smooth approximation for the maximum VEB force is shown (thick solid line). Also the maximum VEB force achievable with the CVT is shown (thin solid line). It is clear that the CVT approach over estimates the VEB force.

a constraint to ensure that the vehicle speed allowed in that gear is not exceeded. In other words the engine speed must be limited:

$$W_{\text{eMIN}} \leq W_{\text{eng}} = \frac{vN_{\text{g}_\text{k}}N_{\text{f}}}{R_{\text{w}}} \leq W_{\text{eMAX}} \tag{13}$$

This second problem does not involve particularly difficult functions and using the solution of the first as initial guess usually leads to a very short number of iterations. It is just a question of accommodating the small differences between the real VEB force limit and the approximation.

The main vehicle dynamics are now modeled and simplified in a, for optimization suitable way, without significantly reducing the model accuracy.

Component wear cost model

In order to study the effect on component wear when using ABs and FBs a mathematical model is required. As mentioned earlier usage of FBs creates disc, pad, and tyre wear whereas usage of AB only creates tyre wear. The pad wear is modeled according to Equation 14.

$$\dot{S}_{\text{pad}} = q_{\text{in}}(v, F_{\text{fb}})S_0 e^{cT_1^{k_0}} \tag{14}$$

q_{in} is the brake power and c, S_0 and k_0 are constants that are chosen in such a way that a Volvo pad is represented by Equation 14. Note that the pad wear rate increases exponentially with FB temperature. It is assumed that all FBs of the combination have the same dynamics. The tyre wear rate is modeled according to Equation 15.

$$\dot{S}_{\text{tyre}} = v\left(a + b\tau_{\text{tyre}}^2 + c\tau_{\text{tyre}}^4 + d\tau_{\text{tyre}}^6\right) \tag{15}$$

The tyre wear rate, $\dot{S}_{\text{tyre}}$, is a function of tyre torque, τ_{tyre}, and the constants a, b, c, and d are chosen in such a way that Equation 15 represents the tyre wear of a Continental truck tyre of dimension 315/80 R22.5. It is assumed that the tyre wear

dynamics are the same for all axles of the combination. It should also be noted that the tyre wear model only takes into account tyre wear caused by longitudinal forces, i.e. the tyre caused by lateral motion (curves etc.) is not modeled.

The total cost rate is given in Equation 16 where C_{fb} is the pad and disc wear cost and C_{tyre} is the tyre wear cost. Since the disc wear is very hard to model due to large variations depending on environment (rust, humidity etc.) a rule of thumb was used to model the disc wear: in practice the discs are replaced every second time pads are changed. So, their wear was simply taken as being half that of the pads.

$$\dot{C}_{\mathrm{total}} = C_{\mathrm{fb}}\dot{S}_{\mathrm{pad}} + C_{\mathrm{tyre}}\dot{S}_{\mathrm{tyre}} \tag{16}$$

In Equation 17 the cost of pad and disc wear (C_{fb}) is calculated from the cost of parts and labor associated with replacing them.

$$C_{\mathrm{fb}} = \frac{C_{\mathrm{padw}} + C_{\mathrm{padm}}}{\delta_{\mathrm{pad}}} + \frac{C_{\mathrm{discw}} + C_{\mathrm{discm}}}{2\,\delta_{\mathrm{pad}}} \tag{17}$$

where

W_{padw} is the cost of labour to replace one pad
W_{padm} is the cost of one pad
W_{discw} is the cost of labour to replace one disc
W_{discm} is the cost of one disc
δ_{pad} is the thickness of the pads

The cost for tyres (C_{tyre}) is calculated in the same way, Equation 18

$$C_{\mathrm{tyre}} = \frac{C_{\mathrm{tyrew}} + C_{\mathrm{tyrem}}}{\delta_{\mathrm{tyre}}} \tag{18}$$

where

C_{tyrew} is the cost of labour to replace one tyre
C_{tyrem} is the cost of one tyre
δ_{tyre} is the tyre thread thickness

Numerically the costs referred to here are those charged in a Volvo workshop in Sweden.

3 Method used to obtain optimal control trajectories

The optimal control problem consists of the determination of the control signals to be applied in order to make a system evolve in time according to a predefined objective. To achieve this the dynamic system must first be represented by a set of differential equations that describe the state, $\mathbf{y}(t)$, of the system when subjected to control inputs, $\mathbf{u}(t)$, see Equation 19. The system equations were presented in the previous chapter. Secondly, this set of equations must be transformed in order to be implementable into a Non–Linear Programming (**NLP**) software. The aim

of this chapter is to describe the different steps, starting after the mathematical model of the process and ending at a solution to the optimal control problem.

$$\dot{\mathbf{y}} = \begin{bmatrix} \mathbf{f}_1(y_1(t), y_2(t), \ldots, y_{\mathrm{ns}}(t), u_1(t), \ldots, u_{\mathrm{nc}}(t), t) \\ \vdots \\ \mathbf{f}_{\mathrm{ns}}(y_1(t), y_2(t), \ldots, y_{\mathrm{ns}}(t), u_1(t), \ldots, u_{\mathrm{nc}}(t), t) \end{bmatrix} \tag{19}$$

where n_{s} and n_{c} are the number of state and control variables respectively. The total number of variables will be denoted as $n_{\mathrm{Var}} = ns + nc$.

The mathematical formulation of the objective can be defined as:

$$\min \quad \Psi\left[\mathbf{y}(t_{\mathrm{I}}), \mathbf{y}(t_{\mathrm{F}})\right] + \int_{t_{\mathrm{I}}}^{t_{\mathrm{F}}} \Phi\left[\mathbf{y}(t), \mathbf{u}(t)\right] \quad dt \tag{20}$$

It means that the objective can be defined as the minimization of a function over the time horizon as well as functions defined at specific time instants.

In reality it is likely that the control signals are in some way limited by physical properties of the systems. These limitations can be written in the form:

$$\mathbf{g}_{\mathrm{L}} \leq \mathbf{g}\left(\mathbf{y}(t), \mathbf{u}(t), t\right) \leq \mathbf{g}_{\mathrm{U}} \tag{21}$$

3.1 Obtaining a finite optimization problem–large scale NLP

The optimal control problem involves continuous functions of state and control variables ($\mathbf{y}(t)$ and $\mathbf{u}(t)$) whereas a NLP requires a finite number of variables and constraints (x and C). This section describes the technique used to convert the dynamic system into a problem with a finite set of variables. The solution of the optimal control problem comes as the result of merging methods for optimization and numerical solution of differential equations. The method used here is commonly referred to as *multiple shooting*. Scalar variables are used to represent the values of the continuous functions in a finite number of time instants within the time range of the problem. A numerical integration method is then used to propagate the differential equations between consecutive time instants, as suggested in Figure 6. The points corresponding to the time instants are called *nodes* or *grid points*. If the system characteristics change along the problem horizon the concept of phases must be introduced. Each phase consists of unique set of ODEs and boundary conditions in order for the trajectories to be continuous.

If the grid has M points the resulting NLP will have Mn_{Var} variables. These are usually aligned in the vector x as:

$$x^{\top} = (y_1^{\top}, y_2^{\top}, \ldots, y_k^{\top}, \ldots, y_M^{\top})$$

where y_k is the n_{Var} component vector with states and controls values at the time instant t_k (k-th grid point). Using this method to solve for optimal control trajectories the resulting NLP will always be very large.

In order to ensure continuity of the trajectories it is necessary to impose that the function values at the end of an interval are the same as the starting values of the following interval (i.e. $\xi = 0$ in Figure 6). In NLP terms this means that $M - 1$ equality constraints must be imposed for each state variable (*defect*

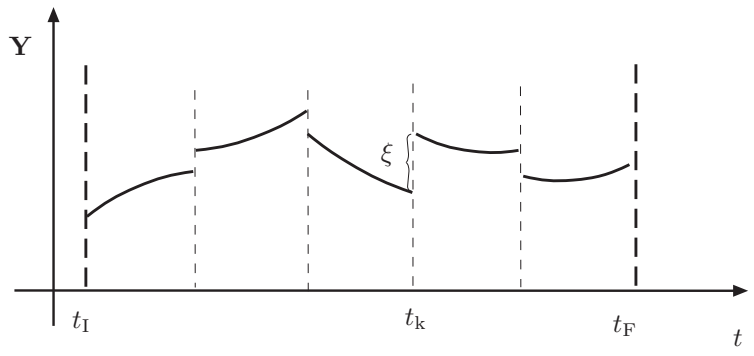


Figure 6 Multiple shooting

constraints). The constraint equations are derived from the numerical integration method used. Any numerical integration method can be employed but the most frequently used are Euler, Trapezoidal and Hermite Simpson. These methods vary in ease of implementation, accuracy and computational cost associated with the constraint evaluations. Using the trapezoidal method results in defect constraints of the form:

$$0 = \mathbf{y}_{k+1} - \mathbf{y}_k - \frac{h_k}{2}\left(\dot{\mathbf{y}}_{k+1} + \dot{\mathbf{y}}_k\right) \tag{22}$$

The formulas for other methods and a detailed description of the procedure just described can be found in [6].

The discrete form of the path constraints, Equation (21), is obtained by applying them at each grid point. Each one produces M constraints of the form

$$\mathbf{g}_L \leq \mathbf{g}\left(\mathbf{y}_k, \mathbf{u}_k, t_k\right) \leq \mathbf{g}_U \tag{23}$$

3.2 Numerical solution

A Nonlinear Program (NLP) is a problem that can be put into the (canonical) form:

$$\begin{array}{lll} \text{Minimize:} & F(x) & (\Re^n \to \Re) \\ \\ \text{subject to:} & C_{\mathrm{e}}(x) = 0 & (\Re^n \to \Re^{m_{\mathrm{e}}}) \\ & C_{\mathrm{i}}(x) \geq 0 & (\Re^n \to \Re^{m_{\mathrm{i}}}) \end{array}$$

The function F is called the *objective function* and C_{e}, C_{i} are equality and inequality *constraints*, respectively. All together, these functions are referred to as the *problem functions*. The vector $x \in \Re^n$ is the vector of variables and is said to be *feasible* if it satisfies the constraints. The set of all feasible solutions is called the *feasibility set*. In the case if no feasible solution exists the problem is said to be *infeasible*. The solution of these type of problems is almost always obtained through an iterative sequence of vectors x_{i} (where i is an iteration counter). The process starts with a vector x_0 frequently called *initial guess*. The (global) optimal solution

is the feasible solution that minimizes the objective function. Because NLPs are difficult to solve much attention has been given to special cases of problems for which successful methods have been developed. Important examples are Linearly Constrained (LC), Quadratic Programming (QP) and Linear Programming (LP) problems.

One of the most successful algorithms for NLP is based on solving a sequence of QPs, called Sequential Quadratic Programming (SQP). A QP is characterized by a quadratic objective function and linear constraints. In a simplified manner the SQP algorithms follow the general steps (see also [6], [7], [8], and [9] for more information):

1. Check convergence of current solution vector. Terminate if optimal conditions are satisfied.

2. Compute a linear approximation to the constraints (1^{st} order Taylor series expansion). Requires evaluation of first order derivatives of the constraints, i.e. the Jacobian matrix.

3. Compute a quadratic approximation to the objective function (2^{nd} order Taylor series expansion). Requires evaluation of first and second order derivatives, i.e. the gradient and the Hessian of the objective function.

4. Solve the QP subproblem constructed from the approximations derived in the two previous steps. The vector $p = x_{\mathrm{k}} - x_{\mathrm{k}-1}$ gives the search direction. (Because the approximations are only valid near the previous solution $x_{\mathrm{k}-1}$ it may not be safe to take the solution of the QP subproblem x_{k} to the next iteration, line search is therefore applied).

5. Line search: Find a point along the search direction p that gives a sufficient decrease in a *merit function*, i.e. take $x_{\mathrm{k}} = x_{\mathrm{k}-1} + \alpha p$ with $0 \leq \alpha \leq 1$. The merit function is a measure of optimality that involves the value of the objective function as well as some measure of the constraints violations. Note that the line search procedure is a minimization subproblem in a single variable (α). Each line search iteration requires the evaluation of both the objective function and the constraints.

6. Update vector x and all quantities and return to first step.

In the implementation of a SQP algorithm several other issues in addition to those mentioned here must be handled. For instance, it is possible that the QP subproblem is infeasible or unbounded as the result of linearization of nonlinear constraints. Other difficulties can arise from rank deficient constraints or indefinite Hessian. A thorough discussion on methods for NLP is presented in [10], including efficient methods to solve QPs, discussion of line search procedures and definition of merit functions. Strategies to deal with the difficulties described above are also discussed. Merit functions and line search methods were developed in order to improve the global convergence properties of optimization algorithms. These and other alternative globalization strategies such as *Trust Region Methods* and *Filters* are discussed in detail in [6]. Depending on non–convex criterion function or, like in this paper, non–convex constraints, there is no proof that a globally optimal solution is in fact obtained. However, in order to reduce the possibility for convergence to

local optimality much work was done on scaling variables, defect constraints, path constraints, and the objective functions. Variables where scaled to order 1 and the Jacobian was scaled in such a way that the row and column norms where approaching the same order of magnitude, and thereby decreasing the condition number of the Jacobian. Objective function scaling was a little more difficult. Merit functions (which are used as a measure of optimality) are constructed from objective function values and the sum of constraint violations. At optimality the constraint violations tend to zero but if the objective function is disproportional the globalization properties of the merit function can deteriorate. Scaling should therefore produce an objective function value close to unity at the solution. This scaling does, however, also affect the gradient and Hessian and may possibly result in poor convergence. Ideally Hessian row and gradient norms should be scaled to unity. Usually these three problem function properties are hard to achieve simultaneously and the scaling of the objective function was therefore a trade–off between function values and derivative values. Problem function scaling proved to be very efficient in order to avoid local optimality and also reduced the computational effort very much. The computational effort is also very dependent on how derivatives are calculated and providing analytically calculated derivatives was necessary to achieve reasonable computational times (minute range). Relaying on internal NLP software finite difference methods does generally not work well for optimal control problems since they do not utilize the special structure of optimal control problems like described in [4]. Therefore, it is essential, that the problem functions are *smooth* i.e. they should be at least twice continuously differentiable. Careless use of discontinuous functions such as: *min, max, round, absolute value, if conditions, and linear interpolation on sampled data* will most likely result in failure in the convergence to the solution, and therefore, smooth approximations are necessary whenever relevant data is given in tabular form. This is usually achieved by least squares methods or cubic spline interpolation. In this paper the discontinuity of the gearbox provided several difficulties. Mixed Integer Nonlinear Programming (MINLP) made it possible to introduce the gear position as an integer, see for example [11]. This did, however, not improve the result since the computational effort was very large even for very small problems. Instead the smooth VEB force approximation followed by a multi phase optimization, where the gear ratio is locked in every phase, was the best choice, see Chapter 2.1.

Another very important issue is providing an initial guess. Here a very easy method was used to build an initial trajectory by simulating constant speed driving at the initial speed level under the assumption that maximal available AB forces must be used before engaging the FBs. One drawback with this approach is that the initial guess may be infeasible if a too high initial speed is chosen, i.e. violating the upper bound on the FB temperature. Even so, this proved to be much more efficient than using a random initial guess or relaying on solver internal methods to find suitable initial trajectories.

For the solution of the NLP, a state–of–the–art optimization package called TOMLAB, see [9], has been used in this work. The solution obtained from the NLP solver requires some form of validation. This is because the control and state trajectories computed by the optimization routine might not be in accordance with each other, that is, when the control signals are applied to the system the states behaviour will deviate from their calculated trajectories. This occurs mainly for two

reasons, constraint violation and numerical integration error. Constraint violations where minimized by the scaling of the problem functions, and therefore, numerical integrator errors where the main error source. The validation was performed by simulating the system response to the optimal control signal. If the result was not satisfactory, grid refinement was typically used to improve the result.

4 Results and discussion

In this section, downhill strategies generated from the optimization are presented and discussed. Optimal control strategies to obtain maximum mean speed on different road profiles are presented first, followed by a study, using Pareto curves, of the trade–off between mean speed and component wear cost, i.e. multi–objective optimization. Special attention is given to the trade–off between different levels of constant speed travel and component wear cost. The constant speed scenario is especially interesting since ordinary (constant speed) cruise controllers are often fitted on modern trucks. The vehicle environment in terms of downhill slopes is represented by both constant road slopes and one real measured road profile. Constant slope road profiles are used to clearly demonstrate optimal driving strategies, and the measured road profile (French alps) is used for the comparison between driving strategies for different vehicle combinations (VEB, VEB+FB, VEB+CR, etc.) and initial states (FB temperature). Additionally, throughout this section, similarities and differences between optimization results and what is achieved by real drivers are discussed.

4.1 Mean speed maximization

Mean speed optimal trajectories that were obtained using Tomlab are shown in Figure 7. The truck weight is 60 tonnes and it is equipped with VEB and FBs. Thick lines represent a configuration with a CVT whereas thin lines represent a discrete gearbox. Due to the high initial FB temperature the FB are not used very much the first 2000 m of the slope, instead the VBE is fully utilized. This enables the FB to cool down. Therefore they can, when the downhill road slope increases, be applied to reduce the vehicle speed. In the CVT case, the brake system is fully utilized from 2000 m and forward. This can be seen from the fact that the VEB is fully utilized (not directly shown in the figure) and that the FB temperature is at the upper limit of 350°C which ensures maximum energy dissipation. Between 3000m and 5000m the system is stationary. It is interesting to see that approximately 800 m before the downhill slope increases from 6% to 8% the speed is slowly reduced from approximately 13 m/s to 10 m/s ensuring that the upper limit on FB temperature is not violated. An optimal driver would therefore need to have a road slope preview of 800 m to achieve the same. If the preview was shorter than 800 m a transient in FB force would be required to slow the vehicle down, resulting in a FB temperature transient over 350 °C. In the case a discrete gear box was used the brake system is not fully utilized and the FB temperature only reach 350 °C for short periods of time. There is also a difference in mean speed between CVT (13.4 m/s) and discrete gear box (11.9 m/s) configurations. It can also be seen that the discrete gear box configuration does not reach stationarity like the CVT configuration does.

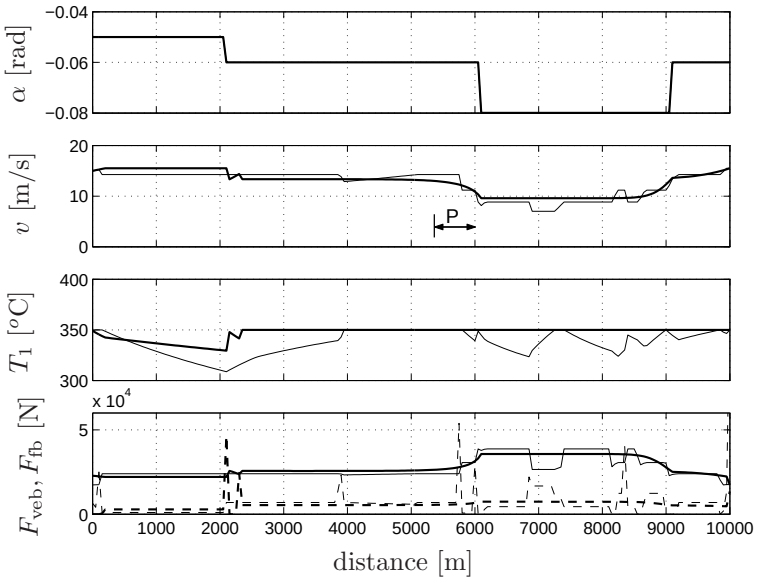


Figure 7 Optimal trajectories for a 60 tonnes truck with VEB+FB. Both CVT (thick lines) and discrete gear box (thin lines) are shown. F_{veb}: solid line and F_{fb}: dashed line. The preview distance (P), indicates the distance from where the speed reduction starts (to handle the upcoming increase in slope) to when the slope actually increases.

It must also be noted that solution for the discrete gearbox is feasible but it is not possible to prove that a global optimality is reached due to the rather difficult non–linearity introduced by the finite number of ratios. The typical behaviour for a discrete gearbox is shown in Figure 8 (same configuration as in Figure 7) where the road slope is constant at 5% the whole time. Again the CVT configuration utilizes the brake system to the fullest (maximum VEB and T_1). For the discrete gearbox a cyclic behaviour can be seen now when no road slope transients are present (initial and endpoint effects are present, though). Between 900 m and 1200 m two FB pulses are applied to reduce the vehicle speed. Almost the same two pulses (in terms of level and frequency) are again applied between 1650 m and 1950 m. The mean duration of these four pulses is 13 s and the FB temperature is varying between 330 °C and 350 °C. *This behaviour can very much be compared to the behaviour of a very skilled driver. One fundamental difference is, however, that the driver does not know the FB temperature at which the vehicle is operating at, i.e. in manual driving safety is not secured.*

In Figure 9 trajectories for the configurations FB+VEB+CR (thin lines) and FB+VEB (thick lines) on a real road sections (French alps) are shown. It is clear that the addition of a CR improves the performance in terms of improved mean speed. It can also be seen that the FB+VEB+CR configuration uses FBs shortly and only on section of the road that are very steep. This behaviour of course depends on the initial FB temperature when entering the downhill slope. In Figure 10 the maximum average speed for different brake combinations (all with CVT) are shown as function of initial brake temperature. When all available brake components are used (FB+VEB+CR) maximum allowed mean speed is always achieved

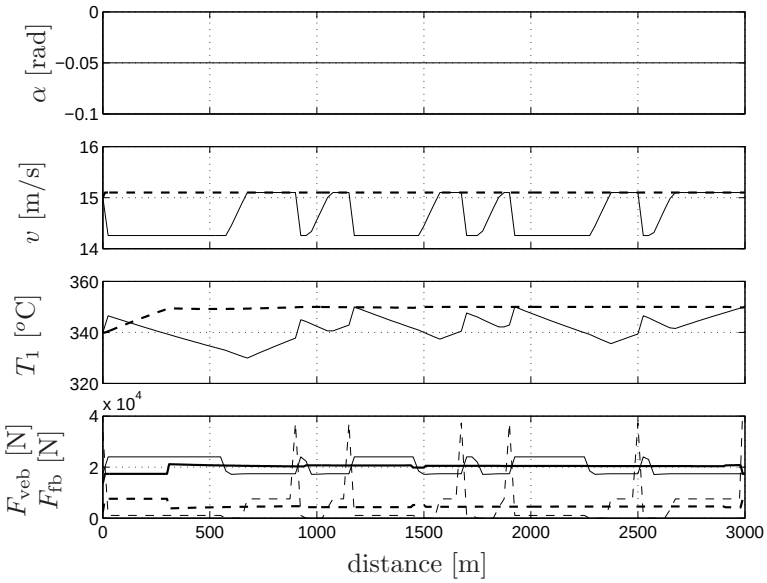


Figure 8 Optimal trajectories for a 60 tonnes truck with VEB+FB. Both CVT (thick lines) and discrete gear box (thin lines) are shown. F_{veb}: solid line and F_{fb}: dashed line. A cyclic behaviour can be seen for the FB usage in the discrete gearbox configuration.

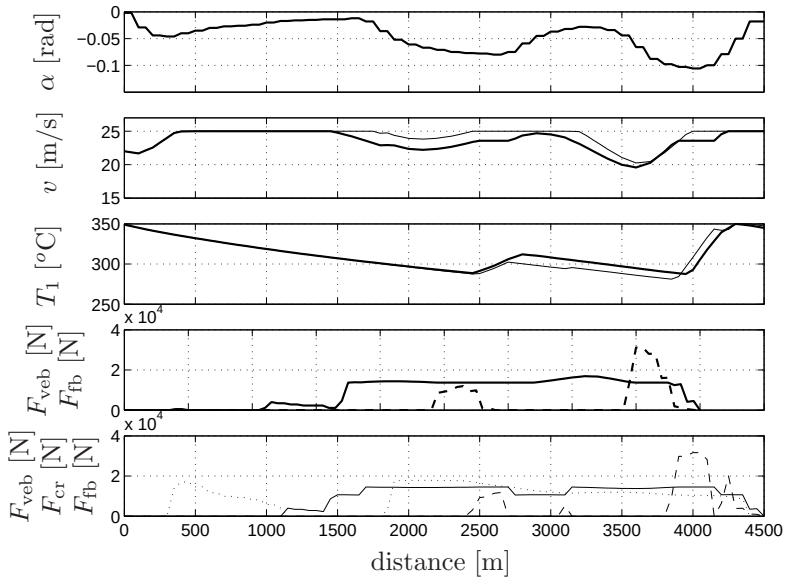


Figure 9 Optimal trajectories for a 60 tonnes truck on real road profile (French alps) with FB+VEB and FB+VEB+CR. Only CVT configuration. Thick lines: FB+VEB. Thin lines: FB+VEB+CR. Solid line: F_{veb}, dashed line: F_{fb}, and dotted line: F_{cr}.

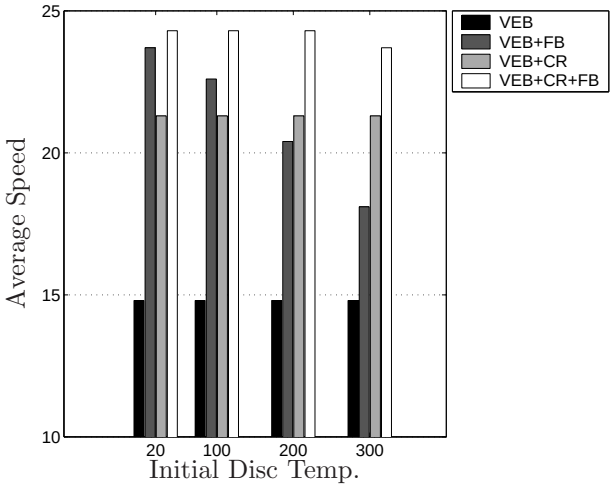


Figure 10 60 tonnes truck with CVT and 4 different brake systems. Maximum average speed for different initial disc brake temperatures on real downhill slope (French alps).

except if the initial brake temperature is the highest. Comparing the performance of VEB+FB and VEB+CR show that using FB is superior and increases the performance if the initial brake temperature is rather low (below $100 - 150\,^{\circ}\mathrm{C}$). This is especially interesting since it indicates that there is a potential in replacing the relatively expensive, heavy, and dragging ($\approx$ 2 kW always) CR with the already existing FBs, without a reduction in performance.

4.2 Multi–objective optimization, speed vs. wear

As discussed earlier the component wear cost was modeled to enable generation of wear cost optimal control strategies. Since the most wear optimal strategies is not to drive at all, constraints on speed have to be introduced. Here the vehicle speed is either constrained to a constant level (ordinary cruise controller) or to a certain average speed. In Figure 11 the trade–off between average speed and component wear cost for a realistic road profile is shown. Four different initial brake temperatures are shown (20, 100, 200, and 300 °C). As expected, the lowest temperature results in the best performance: lowest cost for all different average speeds. As the temperature increases the difference in performance between FB+VEB and FB+VEB+CR increases, i.e. a vehicle equipped with both VEB and CR can obtain a higher mean speed without increasing the wear cost. It can also be seen that using AB only compared to using AB+FB result in very bad performance in terms of wear cost. Constant speed traveling is usually achieved by engaging a cruise controller. It is therefore interesting to see how component wear cost is affected by constant speed driving and optimized speed driving, i.e. in both cases the average speed is required to be the same, see Figure 12 (left). The initial FB temperature is 100 °C and the initial speed is always set to the corresponding average speed, i.e. not the same as in Figure 11 where only one initial speed was used. It can be seen that the variable speed (thick lines) produce a lower cost compared to the

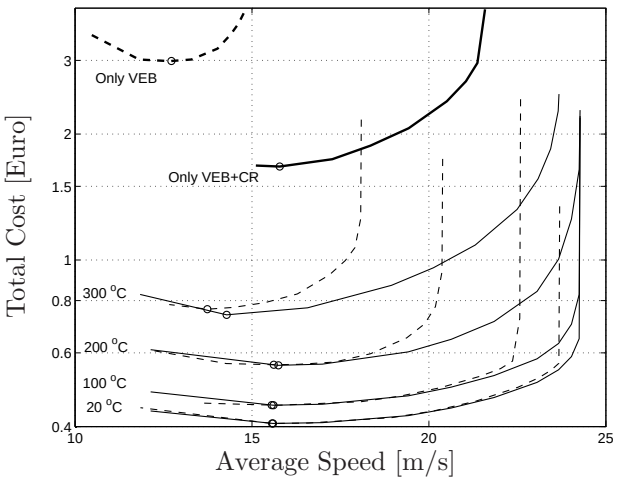


Figure 11 Pareto curves for 60 tonnes truck on French alp road. Dashed lines represent VEB and solid lines VEB+CR. Thick lines represent ABs only. Thin lines represent ABs+FBs. The scale on the cost axis is logarithmic.

constant speed driving (thin lines). There is also a possibility to achieve a higher mean speed by driving with a variable speed compared to constant speed driving. In Figure 12 (right) trajectories for point C and V (left figure) with VEB+FB are shown with thin solid line and thick solid line respectively. It can be seen that the variable speed case (point V) has a very large speed variation between 10 and 25m/s. Even if this is the most wear cost optimal way of driving is may not be realistic to achieve in real traffic situations. The dashed line represent more realistic cost optimal driving with only 15 m/s ±10% maximum allowed speed variation. Even if the speed variation is modest, the wear cost is reduced 23% compared to constant speed wear optimal driving.

In Figure 13 wear cost optimal brake force distribution trajectories are shown for a vehicle of 60 tonnes on a 5 % downhill slope ranging from 1 km to 20 km. Additionally the wear cost reduction when using optimal blending compared to when only using ABs is shown. For example, on a 1 km long slope the optimal distribution is to use approximately 47 % AB and 53 % FB and the cost reduction is 73% compared to using 100%ABs. Using 100% ABs is the normal driver strategi to minimize the FB wear. When also the tyre wear is taken into consideration, the normal driving strategy of using only ABs is not cost efficient.

5 Conclusion

Increasing payload and reducing air and rolling resistances are extremely important features in the truck industry and will result in increased requirements on the brake system both in manoeuvers that require high brake power (collision avoidance) and manoeuvers that require low power (downhill cruising). In this paper the development of a software tool to allow evaluation of different brake configurations and downhill driving situations (slopes) has been described. The main result shows that both downhill driving mean speed and component wear cost can be im-

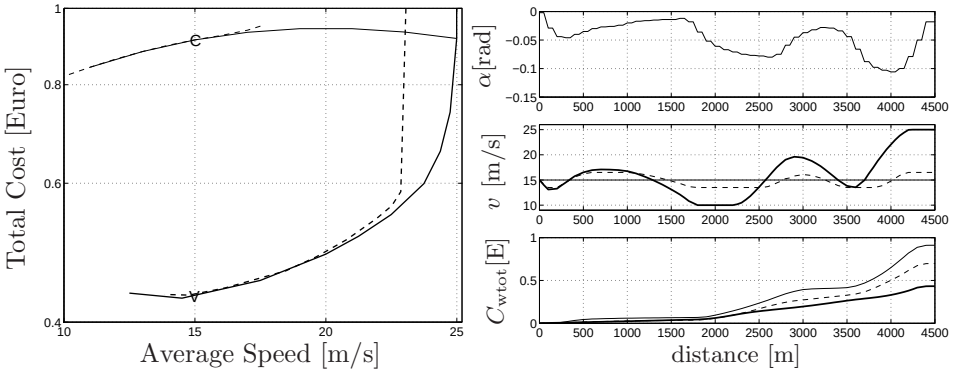


Figure 12 **Left figure:** Pareto curves for 60 tonnes truck on French alp road. Dashed lines represent FB+VEB and solid lines FB+VEB+CR. Thick lines represent wear optimal variable speed driving. Thin lines represent wear optimal constant speed driving. The scale on the cost axis is logarithmic. **Right figure:** The trajectories for point C and V with FB+VEB from the left figure are shown with thin solid line and thick solid line respectively. It is clear that the speed variation for point V is rather high (between 10 and 25 m/s). Thin dashed line represent wear optimal trajectories but the speed is constrained to ±10% of 15 m/s. Even small speed variations allow the wear cost to be reuduced significantly when compared to the constant speed driving.

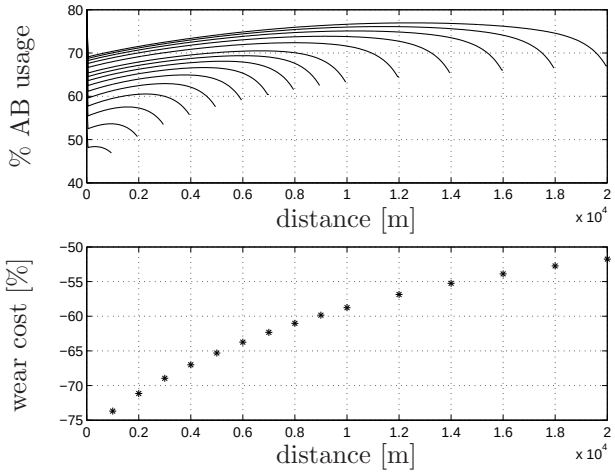


Figure 13 60 tonnes vehicle equipped with FB+VEB on a 5% downhill slope of different length ranging from 1 km to 20 km. Constant speed 40 km/h and the initial FB temeprature is 100 °C . The upper figure shows the optimal brake force distribution between ABs and FBs (100% means that no FBs are used) that minimizes the component wear cost. The lower figure present the cost reduction when comparing optimal blending and normal driver behaviour, i.e. optimal blending vs. only ABs. Clearly the FB should be used in order to reduce cost. FBs can be used more on shorter slopes.

proved simultaneously compared to the normal driving strategy of only using ABs, see Figure 11. The possibility to have variable speed (constraint on mean speed!) clearly improves the wear situation significantly but variable speed might not be applicable in all situations where constant speed keeping is instead preferred. In this paper both the possible maximal constant speed and the distribution of brake force between ABs and FBs in order to minimize wear cost under such driving is presented. The main conclusion is that even for constant speed driving, which is a desirable feature, the FB should be applied under controlled conditions to reduce and distribute the tyre wear to all axles of the combination. It is also shown that even a moderate variation of speed (constraint on mean speed!) allows the wear cost to be reduced significantly compared to constant speed driving, see Figure 12. To achieve wear optimal driving, with either constant or variable speed, fully in a practical implementation is not easy and would clearly require road profile preview. A feasible implementable strategy to increase the mean speed (and improve the driver comfort) would be to control the vehicle speed to a predefined constant value (cruise control) by using ABs and if necessary also the FBs. Then, when the FB temperature is approaching a limit, one would switch strategy to control temperature of the FBs and hereby reduce the vehicle speed. This would ensure maximal usage of the brake system since AB are fully utilized (highest possible gear ratio) and the energy dissipation of the FBs is maximal when the temperature is at the highest acceptable level. One drawback would be that the upper bound on the FB temperature would be relaxed and transients in FB temperature would be allowed (another is that it is not wear optimal). In fact the upper bound would possibly be a reference instead. For example, assume that the FB temperature is stabilized at the reference temperature and the road inclination increases. This would require a temporary increase in FB brake power (and therefore temperature) in order to reduce the vehicle speed to the new suitable level. One way to solve this is to choose the FB temperature reference conservatively to avoid fading also in transients conditions. Another approach would be to include road profile preview to obtain the the behaviour described in Figure 7, where the vehicle speed is reduced before the downhill slope increases. It can also be seen from Figure 8 that the introduction of a discrete gearbox (like in all production trucks) results in a cyclic behaviour where the vehicle speed is alternating between two levels. This is in fact similar to a strategy utilized by some very skilled drivers. The cruising speed is chosen in such a way that using only ABs accelerates the vehicle slowly and then, when the speed becomes too high, the driver applies FB to reduce the speed. This procedure is repeated until the end of the slope, making the vehicle speed vary in a similar way as seen in Figure 8. It is, however, important to remember that the driver may not have chosen an optimal speed level and, combined with the fact that the driver does not know the actual FB temperature, the driving may either be conservative or dangerous (overheating).

References and Notes

1 Lingman, P. and Wahde, M. (2002) 'Integrated Retardation Control Using Neural Networks with Genetic Algorithms', *Proceedings of the 6th International Symposium of Advance Vehicle Control, AVEC, Hiroshima*, Vol. 6, pp.751–756.

2 Lingman, P. and Wahde, M. (2004) 'Transport and Maintenance Effective Retardation Control Using Neural Networks With Genetic Algorithms', *Journal of Vehicle System Dynamics*, Vol. 42, pp.89–107.

3 Lingman, P. (2002) 'Integrated Retardation Control-Thesis for the degree of licentiate of engineering', *Chalmers University of Technology*.

4 Andersson, A. (1995) 'A Model for Engine Thermal Management Simulation, *Nordic Matlab Conference*, pp.84–89.

5 Eckert, E.R.G and Drake, R.M. (1959) 'Heat and Mass Transfer, *McGraw–Hill*,

6 Betts, J.T. (2001) 'Practical Methods for Optimal Control Using Nonlinear Programming', SIAM Philadelphia.

7 Boyd, s. and Vandenberghe, L. (2004) 'Convex Optimization', Cambridge University Press.

8 Gill, P.E., et.al. (2002) 'User's Guide For Snopt Version 6, A Fortran Package for Large-Sacle Nonlinear Programming', *Systems Optimization Laboratory, Stanford university, California and Department of Mathematics, University of California, San Diego La Jolla*.

9 Holmström, K., Göran, A. (2003) 'User's Guide for Tomlab 4.0.5', *Tomlab Optimization, Vasteras, Sweden*.

10 Gill, P.E., et.al. (1981) 'Practical Optimization', *Academic Press, London*.

11 Floudas, C.A. (1995) 'Nonlinear and Mixed-Integer Optimization', *Oxford University Press*.

Errata to:

Integrated Brake Control
Downhill Driving Strategies

Peter Lingman

Location	**Is**	**Should be**
p5:r-3	non–ventilated	ventilated
p5:r-2	ventilated	non–ventilated
p13:Figure 2.10		line from retarder to thermostat
p28:r8	21	12
p29:r3	a an	an
p35:r-4	$FP(t)$	$L_{\mathrm{fp}}(t)$
p65:Equation 6.2	q_{in}	cq_{in}, same for all FB temperature Equations
p66:r4	F_{FBd}	F_{FBf}, f for front
p84:r-6	(a)	ä
Paper 3, p24:r-3	M_{ABmax}	M_{AB}
Paper 4, p8:r18	the the	the

Note: row y from top of page x (px:ry) and row y from bottom of page x (px:r-y)