2 HISTORICAL AND LITERATURE OVERVIEW

2.1 Preamble

In this chapter, an overview is given of semi-active dampers, hydro-pneumatic springs and hydraulic oil flow. Since a large amount of research has been done on adjustable dampers, this overview only covers discretely variable dampers, with a fast valve response (fast enough to control body resonance modes up to 2*Hz*). This chapter focuses on literature concerned with large off-road vehicles, but in cases where the applicable technology has not yet be demonstrated on heavy vehicles, reference is made to commercial and passenger vehicles. Systems similar to the one investigated in this study are also discussed.

2.2 Suspension classification

Before semi-active dampers are discussed, it is necessary to define the term: semi-active. There exist many different opinions on the definition of semi-active suspensions. Some authors generalise the word "active" to any suspension system employing an external power supply and signal processing. This definition would however make it difficult to distinguish between a suspension powered by an external hydraulic pump and a suspension using only a small electric current to switch a valve. The difference between passive, adaptive, semi-active and fully active suspension systems are explained in the following paragraphs.

a.) Passive suspension

If a graph of suspension displacement or velocity is plotted against suspension (spring or damper) force, the workspace of a passive suspension is in the first and third quadrants, since both spring and damper forces oppose the direction of displacement and velocity. The force elements in a passive suspension are not adjustable and cannot be controlled. Figure 2-1 is a graphical representation of a passive suspension workspace. The shaded area indicates the workspace, while the line indicates typical force element characteristics.

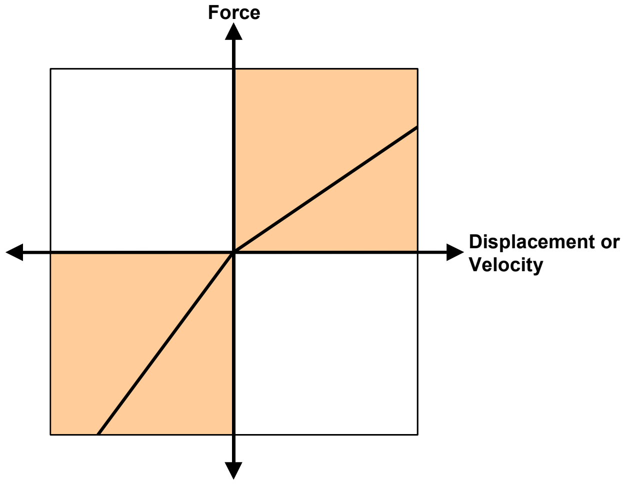

Figure 2-1: Passive suspension workspace

b.) Adaptive or slow active suspension

The workspace of an adaptive or slow active suspension is the same as for a passive suspension, but the force element characteristics can be altered. The main difference between adaptive and semi-active suspensions is the rate at which the characteristics can be changed. For an adaptive suspension, the switching time is slower than the sprung mass natural frequency and requires minimal energy input to switch (see Figure 2-2).

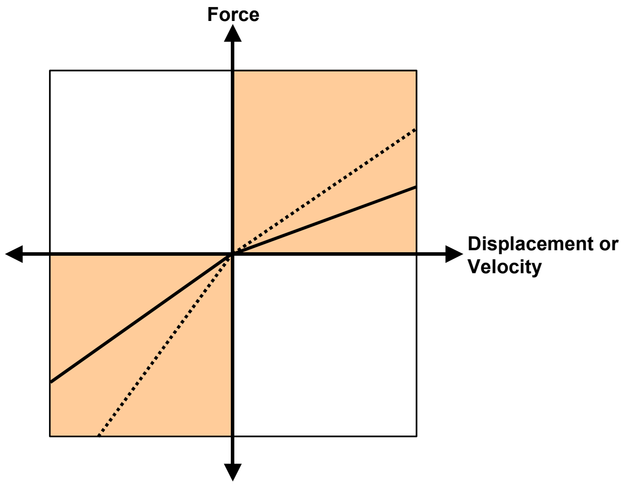

Figure 2-2: Adaptive suspension workspace

c.) Semi-active suspension

The semi-active suspension workspace is the same as the passive and adaptive suspensions and like the adaptive suspension, the force element characteristics can be altered. Spring and/or damper characteristics of a semi-active suspension can be altered rapidly (faster than the sprung mass natural frequency). The energy required to switch between characteristics is still low, but generally higher than an adaptive suspension (see Figure 2-3). Other than the switching signal, no energy is added to the system from an external source.

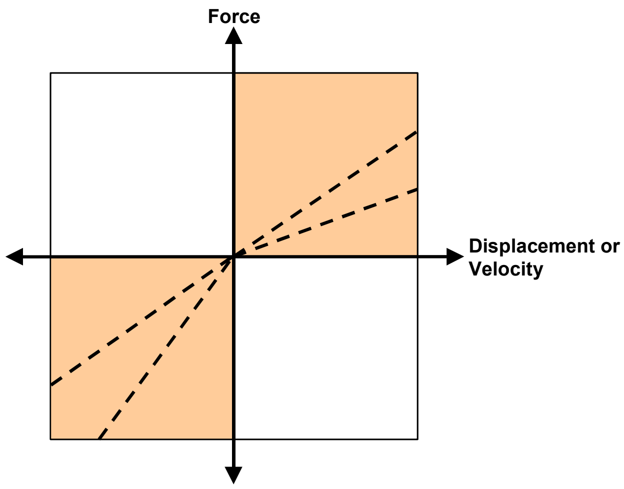

Figure 2-3: Semi-active suspension workspace

d.) Active suspension

The workspace of an active suspension is in all four quadrants, because a positive force can be exerted for negative velocities or displacements and vice versa. The bandwidth of an active suspension is similar to that of a semi-active suspension, but the energy consumption is considerably higher. An external power source is required for this type of suspension.

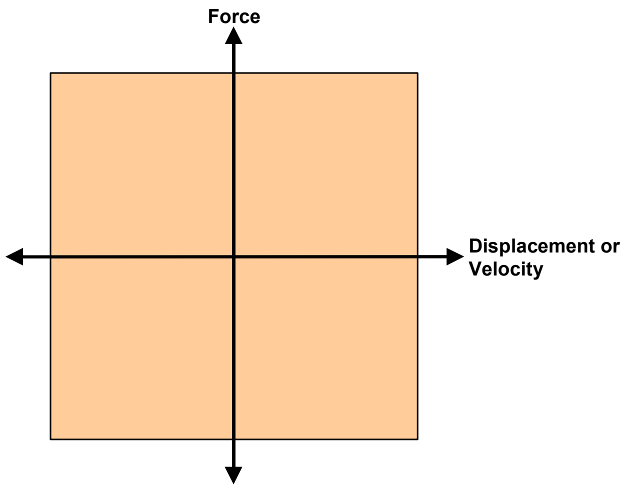

Figure 2-4: Active suspension workspace

2.3 Semi-active dampers

2.3.1 Background

Semi-active dampers were conceptualised in the 1970's and numerous configurations and control strategies were simulated and tested since then. Semi-active dampers greatly influence the vehicle dynamics (ride comfort and handling). This is also the main reason for developing semi-active dampers, namely to improve ride comfort without compromising handling and stability, by switching between hard and soft damper characteristics.

Most semi-active damper studies are conducted on passenger car sized vehicles. Experimental test rigs mostly consist of quarter car models with a sprung mass of ±250kg and an unsprung mass of ±50kg. Not many papers describe the development, or modelling, of semi-active dampers for heavy off-road vehicles (sprung mass of 2500kg to 3000kg).

Numerous so-called semi-active suspension systems were fitted to production vehicles, but most of these suspensions can be classified as adaptive. The reason for the confusion is that most of these suspensions are fast acting (as is semi-active suspensions), but they are employed in an adaptive manner. Examples of such suspension systems are:

- TEMS Toyota Electronic Modulated Suspension Toyota Soarer 1983 (Yokoya et al 1984).
- ASC Adaptive Suspension Control by Armstrong 1989 (CAR July 1989).
- 1987 Thunderbird Turbo Coupe Programmed Ride Control (PRC) Suspension (Soltis 1987).
- 1984 Continental Mark VII/Lincoln Continental Electronically-Controlled Air Suspension (EAS) System (Chance 1984).

Semi-active suspension development in the past mainly focussed on semi-active dampers, although some semi-active roll control devices and semi-active springs were also developed in more recent years.

2.3.2 Semi-active damper control

Modern and classical control theory accounts for very little of the control strategies successfully implemented on heavy off-road vehicles. Other control strategies similar to these were developed for implementation on vehicle platforms where not all the control parameters can be measured.

Many studies have been done to compare theory and experiments. In most of these studies, it was found that simulations are optimistic and often do not include all the physical phenomena and limitations.

Although the aim of this study is not to develop or test new control laws for semi-active suspension elements, some of the well-known strategies were used in the development and testing of the spring/damper unit in this study. The control strategies of Karnopp (Barak 1989), Hölscher and Huang (Nell 1993; Nell & Steyn 1994), and Rakheja and Sankar (1985:398-403) were used to determine the performance potential of the semi-active spring/damper system. The detail of these control strategies is described by Nell (1993).

2.4 Hydro-pneumatic springs

A hydro-pneumatic spring consists of two fluids acting upon each other, usually gas over oil. A compressible gas, such as Nitrogen is used as the springing medium, while a hydraulic fluid is used to convert pressure to force. In a pneumatic or air spring the external force directly compresses the gas and in a hydro-pneumatic suspension hydraulic fluid is used.

2.4.1 Historical overview

Hydro-pneumatic suspensions have been introduced on battle tanks in the 1950's. The first hydro-pneumatic struts were fitted to a prototype tracked vehicle, as a result of research done by two German companies, Frieseke and Höpfner from Erlangen and Borgwald from Bremen into the use of compressible fluids in suspension systems (Hilmes 1982). Since then, several other military vehicles were fitted with hydro-pneumatic suspensions, but most of them did not go into production due to reliability problems and short life span of the mechanical components. Initially, confidence in this type of suspension was low, due to sealing and design problems. These problems were later solved, but ride height change due to heat transfer to the compressed gas, still proved to be cumbersome, especially on tracked vehicles, where track tension is important.

The first production tracked vehicle fitted with a hydro-pneumatic suspension was the Swiss Strv-103 Main Battle Tank (MBT). This vehicle was fitted with a rigidly mounted main weapon and the height adjustable hydro-pneumatic suspension was used to tilt the vehicle upward or downward (Hilmes 1982). Several other military vehicles have since been fitted with hydro-pneumatic suspensions. These include vehicles such as the Swiss Mowag Piranha (Figure 2-5), the British Challenger MBT and the French Giat Vextra (Figure 2-6).

Figure 2-5: Mowag Piranha

Figure 2-6: Giat Vextra

Since the introduction of more reliable sealing techniques, hydro-pneumatic springs have become more popular and are occasionally used in passenger cars, as well as in some large off-road vehicles. This type of suspension system is popular due to its non-linear characteristic and versatility. The non-linear characteristic causes the spring rate to increase as the load is increased. It also reduces body roll and pitching, results in more constant wheel loads and usually eliminates the necessity for a sophisticated bumpstop. Many controllable suspension systems make use of hydro-pneumatic springs because the hydraulic fluid can easily be channelled through ducts, orifices and valves. By adding, or removing, hydraulic fluid, the vehicle dynamics and ride height can be altered.

Hydro-pneumatic suspensions are not commonly used on commercial vehicles due to the high capital cost involved. Instead, pneumatic suspensions consisting of air bellows are mostly used on freight carrying vehicles. Hydro-pneumatic suspensions are found on passenger vehicles, where the design is simplified to minimise manufacturing costs.

Numerous hydro-pneumatic suspensions or suspension components are available on the world market. The internal working of these units all differ, but the basic principal, i.e. compressing a gas, is the same. Technical details of some of these units are supplied in Appendix A.

2.4.2 Modelling of hydro-pneumatic springs

Depending on the degree of complexity and accuracy required from the mathematical model, various mathematical models of hydro-pneumatic and pneumatic springs are available. Some of these models are discussed in more detail in the following paragraphs:

a.) Polytropic process

Hydro-pneumatic springs are often approximated as a politropic process, which is easy to model. In a politropic process the following pressure-volume relationship governs:

$$PV^n = \text{constant} \tag{2-1}$$

with

P – Gas pressure
V – Gas volume
n – Polytropic constant

The following processes can be modelled as a polytropic process:

$n = 0$ Isobaric (Pressure stays constant)
$n = 1$ Isothermal (Temperature stays constant)
$n = k$ Isentropic (Entropy stays constant)
$n = \infty$ Isochroic

A reversible adiabatic process is isentropic, therefore a hydro-pneumatic spring can be modelled by using a polytropic constant between isothermal (1) and adiabatic (k), which is dependent upon the specific heat capacity of the gas. The value of k for Nitrogen (ideal gas) is 1,4 at 300K.

Since hydro-pneumatic accumulators usually have thick walls to handle the high pressures, it invariably results in a high thermal capacity. This means that the gas compression and expansion process in a practical hydopneumatic spring, at realistic excitation frequencies, is close to adiabatic. A politropic constant of 1,35 is often used in mathematical models of hydro-pneumatic springs (Meller 1987). The following equation, proposed by Meller (1987), can be used to determine the hydro-pneumatic spring rate in the static position.

$$c = \frac{npA^2}{V} \tag{2-2}$$

with
c – Gas spring rate
n – polytropic exponent (1.35)
p – effective pressure
A – pressure loaded area
V – gas volume

Figure 2-7 shows the ideal gas, hydro-pneumatic spring characteristics for an isothermal an adiabatic process.

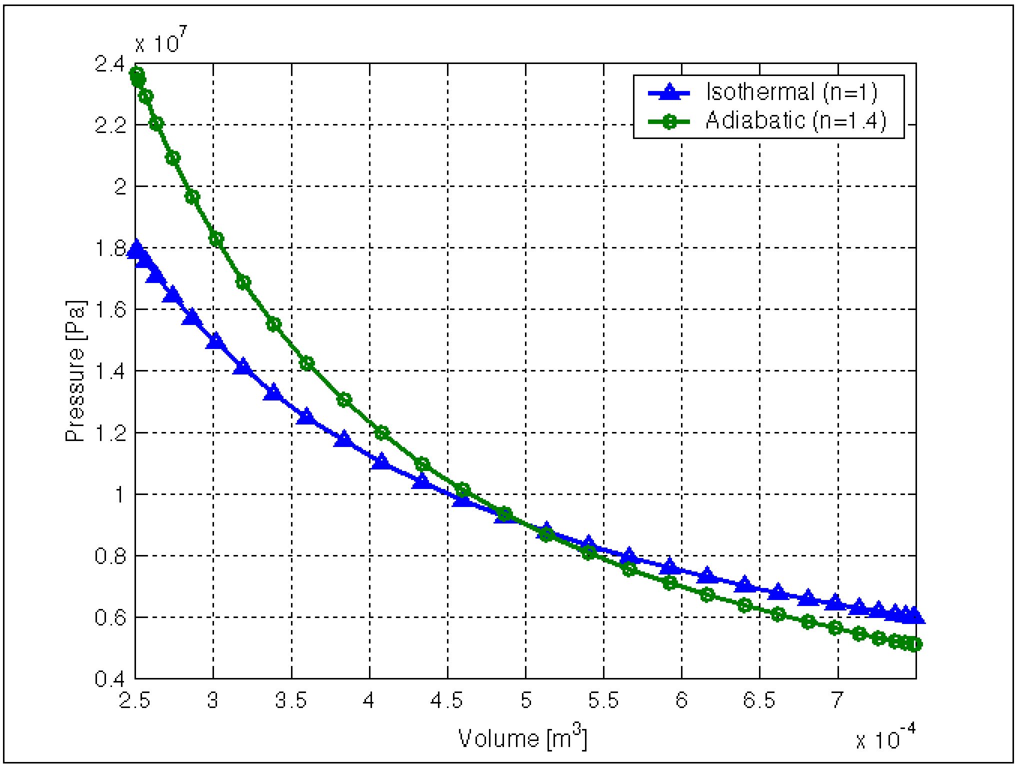

Figure 2-7: Isothermal and adiabatic spring rates (ideal gas)

In this approach, no provision is made for heat transfer to the surroundings. This approach was used by Horton and Crolla (1986), TACOM (1975), Félez and Vera (1987) and Meller (1987), amongst others.

b.) Ideal gas approach

The ideal gas equation of state can also be used to determine the pressure volume relationship of a compressible medium like Nitrogen. Ideal gas assumptions are however only useful at low densities. The ideal gas equation of state can be written as follows:

$$
\begin{aligned}
&PV = mRT \\
&\text{with} \\
&P - \text{Gas pressure} \\
&V - \text{Gas volume} \\
&m - \text{Gas mass} \\
&R - \text{Specific gas constant (296.8 J/kgK for Nitrogen @ 300K)} \\
&T - \text{Gas temperature [K]}
\end{aligned}
\tag{2-3}
$$

Applying the conservation of mass theorem in a closed system and assuming that R stays constant (ideal gas assumption), the ideal gas equation can also be written as:

$$\frac{P_1V_1}{T_1}=\frac{P_2V_2}{T_2}$$

where (2-4)

P_1,V_1,T_1 – Properties at first state

P_2,V_2,T_2 – Properties at second state

Because of its simplicity, this equation is very convenient to use for thermodynamic calculations.

c.) Real gas approach

For pressures and temperatures above the critical point the ideal gas approach may result in significant errors, therefore a real gas approach has to be used. The critical temperature (T_c) of Nitrogen is 126,2K and critical pressure (P_c) 3,39MPa (Van Wylen and Sonntag 1985). Figure 2-8 indicates the compressibility factor (Z) of Nitrogen, as a function of both temperature and pressure. From this figure it is clear that the compressibility factor is very sensitive to the pressure and that the ideal gas approach will only hold for pressures lower than those normally found in hydro-pneumatic suspension systems (Els 1993) (see Appendix A).

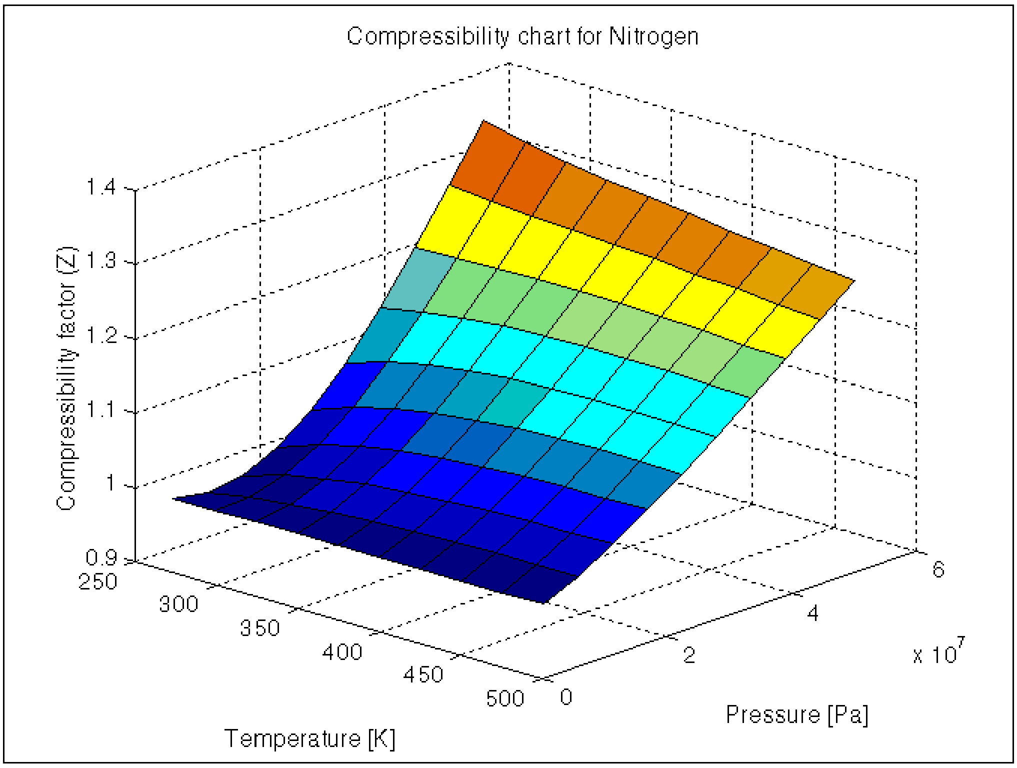

Figure 2-8: Nitrogen compressibility

An accurate equation of state, which is an analytical representation of P-v-T behaviour, is often required for computational models. Several different equations of state have been used. Most of these are accurate only up to some density less than the critical density, though few are reasonably accurate to approximately 2,5 times the critical density. All equations of state fail badly when the density exceeds the maximum density for which the equation was developed.

The best-known and oldest equation of state is the Van der Waals equations first proposed in 1873. A simple equation of state namely the Redlich-Kwong equation was developed in 1949 and is considerably more accurate than the Van der Waals equation (Van Wylen & Sonntag 1985).

The Beattie-Bridgeman equation of state is an empirical equation first proposed in 1928 (Van Wylen & Sonntag 1985). This equation is reasonably accurate for densities lower than 0.8 times the critical density. A more complex equation of state, that is suitable for higher densities, is the Benedict-Webb-Rubin (BWR) equation of state developed in 1940. This equation has eight empirical constants and is essentially an extension of the Beattie-Bridgeman equation through the addition of the high density terms. The BWR equation can be written as follows:

$$P = \frac{RT_g}{v} + \left(\frac{B_0 RT_g - A_0 - C_0 / T_g^2}{v^2} \right) + \left(\frac{bRT_g - a}{v^3} \right) + \frac{a\alpha}{v^6} + \left(\frac{c\left(1 + \gamma / v^2\right) e^{-\gamma / v^2}}{v^3 T_g^2} \right) \tag{2-5}$$

with

T_a - Ambient temperature

T_g - Gas temperature

C_v - Specific heat capacity of the gas

P - Gas pressure

V - Gas volume

v - Gas specific volume

$a, A_0, b, B_0, c, C_0, \alpha, \gamma$ - constants for nitrogen gas

The first term of this equation can be recognised as the ideal gas term, while the rest of the terms are correction terms, compensating for the non-ideal behavior. The BWR equation was used with great success by Pourmovahed and Otis (1990), as well as by Els (1993). The gas pressures for the study conducted by Pourmovahed and Otis (1990) varied between 1 and 19.5MPa, while static pressures of between 6 and 10MPa and maximum pressure of 40Mpa were used in the study of Els (1993).

d.) Thermal time constant, real gas approach

In 1993 the real gas, time constant model, described by Pourmovahed and Otis (1990) and Otis and Pourmovahed (1985), was adapted and applied to hydro-pneumatic springs by Els (1993). This model takes into consideration the heat transfer effects between the gas and the surroundings. The following equations describe the heat transfer between the gas and the surroundings (Els & Grobbelaar 1999):

$$\dot{T} = \frac{(T_a - T_g)}{\tau} - \frac{T_g}{C_v}\left(\frac{\partial P}{\partial T_g}\right)_v \dot{v} \tag{2-6}$$

with
T_a - Ambient temperature
T_g - Gas temperature
τ - Thermal time constant
C_v - Specific heat capacity of the gas
P - Gas pressure
v - Gas specific volume

According to Pourmovahed and Otis (1990), the thermal time constant can either be determined through calculations or through experimental testing. The thermal time constant can be determined experimentally by observing the gas pressure for a step change in the gas volume. During the step change in gas volume, the gas is compressed and the temperature rises. As the gas cools down the pressure reduces. The thermal time constant, τ, is the time it takes the gas pressure or temperature to drop by 63,2% to the final equilibrium pressure or temperature (see Figure 2-9).

Although the thermal time constant is not constant (varies with the heat transfer coefficient), it was found that a constant value could be assumed, if extreme accuracy is not required (Pourmovahed & Otis 1990). The thermal time constant for the accumulators used in this study, as determined by Els (1993), is approximately 6s.

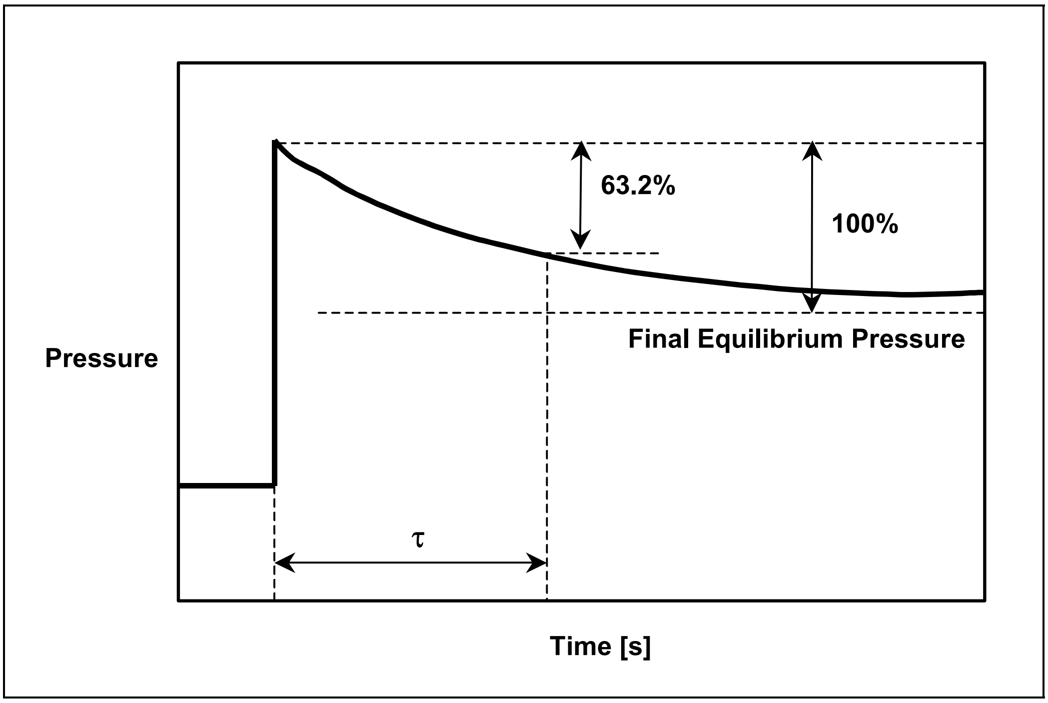

Figure 2-9: Experimental determination of the thermal time constant

Heat transfer may account for up to 30% of the thermal losses (at specific excitation frequencies), which results in the characteristic hysteresis loop of a hydro-pneumatic spring. Figure 2-10 shows the hydro-pneumatic spring characteristic when heat transfer effects are taken into consideration.

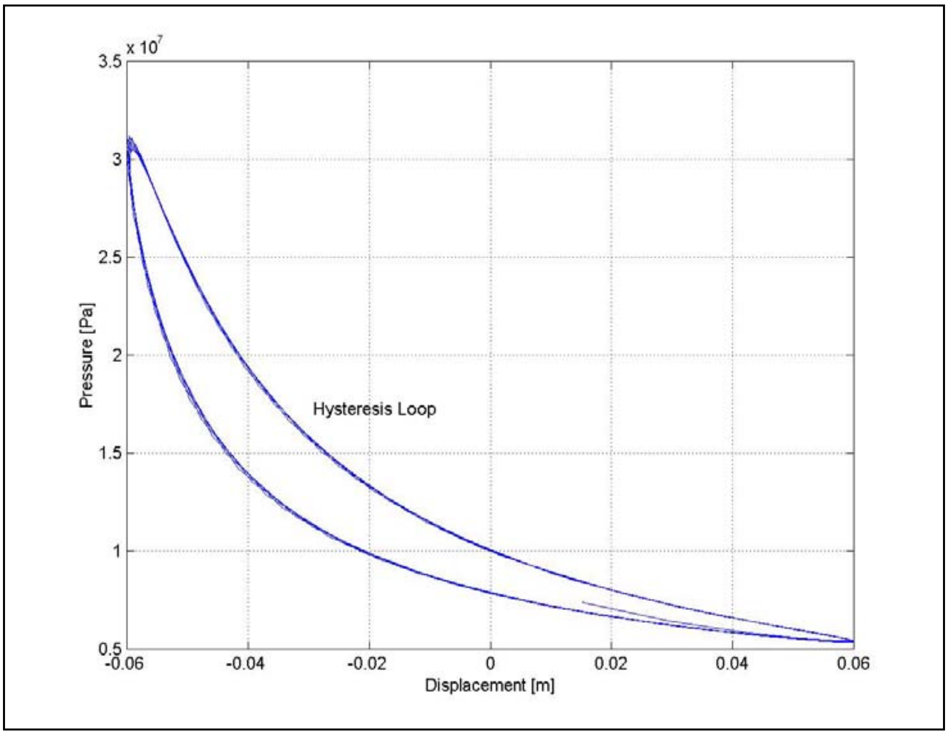

Figure 2-10: Characteristic hysteresis loop of a hydro-pneumatic spring (sinusoidal excitation)

The thermal losses are dependent on the excitation frequency, as well as the excitation amplitude. At low excitation frequencies, sufficient time is available for heat transfer to the surroundings, resulting in the isothermal characteristic. At higher excitation frequencies, the heat transfer process is too slow and the adiabatic characteristic is achieved. For excitation frequencies between isothermal and adiabatic, energy is transferred to the surroundings during the compression stage and not completely recovered during the expansion stage. This phenomenon results in the hysteresis loop, clearly visible in Figure 2-10. The area enclosed by the hysteresis loop indicates the amount of thermal damping at that specific excitation frequency.

Figure 2-11 indicates the amount of thermal damping for different excitation frequencies and amplitudes. From this figure, it can be seen that the thermal damping is frequency dependent and that for this specific case, the peak loss is below any frequency that is of interest in vehicle suspensions.

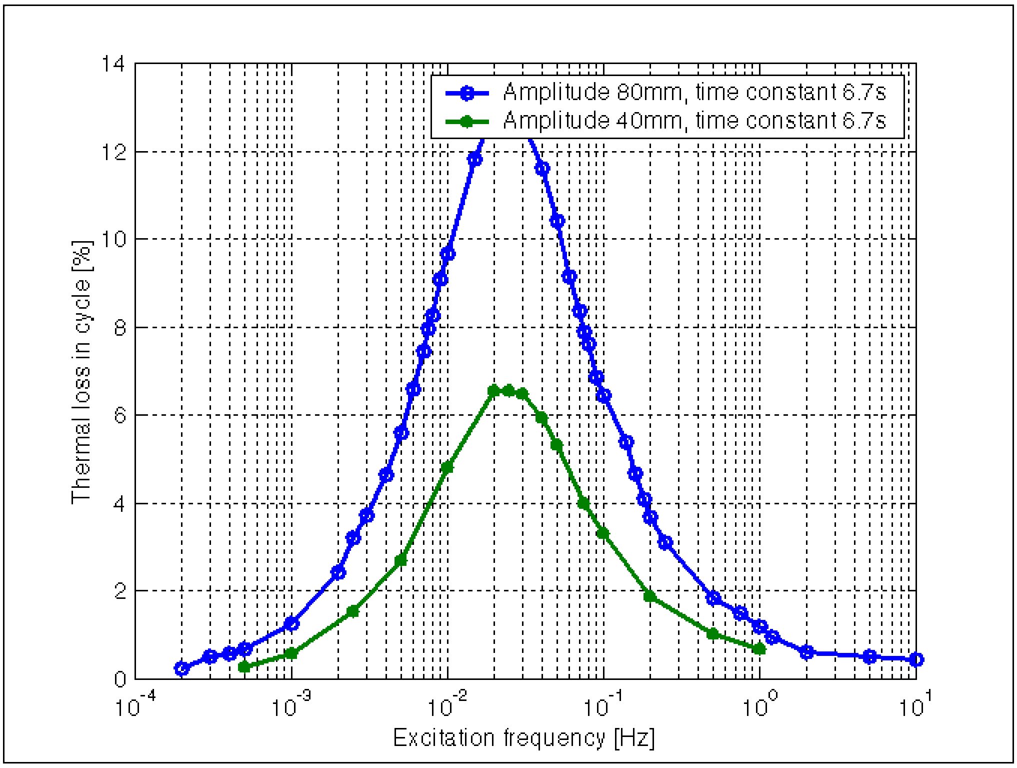

Figure 2-11: Thermal damping

In previous work of Pourmovahed and Otis (1984), a linear anelastic model was compared with the thermal time constant method. Figure 2-12 shows schematically the anelastic model, in which the spring ($k1$) and the damper (c) model the hysteresis loop in the spring characteristic.

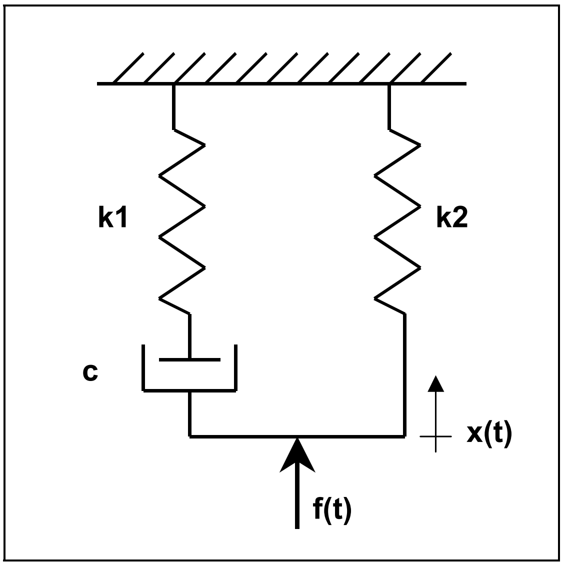

Figure 2-12: Anelastic model for modelling heat transfer in accumulators

The anelastic model was shown to be mathematically the same as the time constant model. The thermal time constant and real gas approach is used in the modelling of the semi-active spring/damper system investigated in this study (see Chapter 3 for more detail).

e.) Bond graphs

The bond graph method was developed for simplifying the process of deriving the equations for mathematical models. This method was used by Félez and Vera (1987) to model a hydro-pneumatic spring system. In their model the damper is treated as a resistive element, while the hydro-pneumatic spring is a capacitive element. A polytropic process was used to model the capacitance of the hydro-pneumatic spring.

The bond graph method can also be used to model the parallel accumulator system of this study, but the causality laws at the split results in derivative, as well as integral equations, which can be troublesome to solve.

2.4.3 Controllable hydro-pneumatic / pneumatic suspensions

Several examples of controllable hydro-pneumatic or pneumatic springs can be found in the literature. The idea of obtaining more than one spring rate by changing the gas volume is not new. Karnopp en Heess (1991) suggested connecting two accumulators in parallel, in order to obtain different spring rates. They remark that it is possible to vary the spring rate, but it is not possible to directly control the force, as can be done with dampers.

The Electronic Modulated Air Suspension System for the 1986 Soarer of Toyota is fitted with an adjustable pneumatic spring and damper (Hirose et al 1988). Valve response times of 70*ms* were attained through an electromagnetic drive system. Modulating a rotary valve between a main and smaller air chamber alter the spring rate. The same principle is used for adjusting the damper characteristics. The spring and damper characteristics are not adjusted individually and a combination of input driven (steering, clutch, throttle or brake input) and reaction driven (measured acceleration, velocity and displacement) control strategies are used. Although fast response times are achieved the control of this suspension can be classified as adaptive rather than semi-active, since the reaction driven strategies react to vehicle speed and ride height, not body motion.

A controllable parallel accumulator suspension system was proposed by TACOM (1975). Figure 2-13 shows the variable spring rate concept. This system was proposed as an operator controlled system.

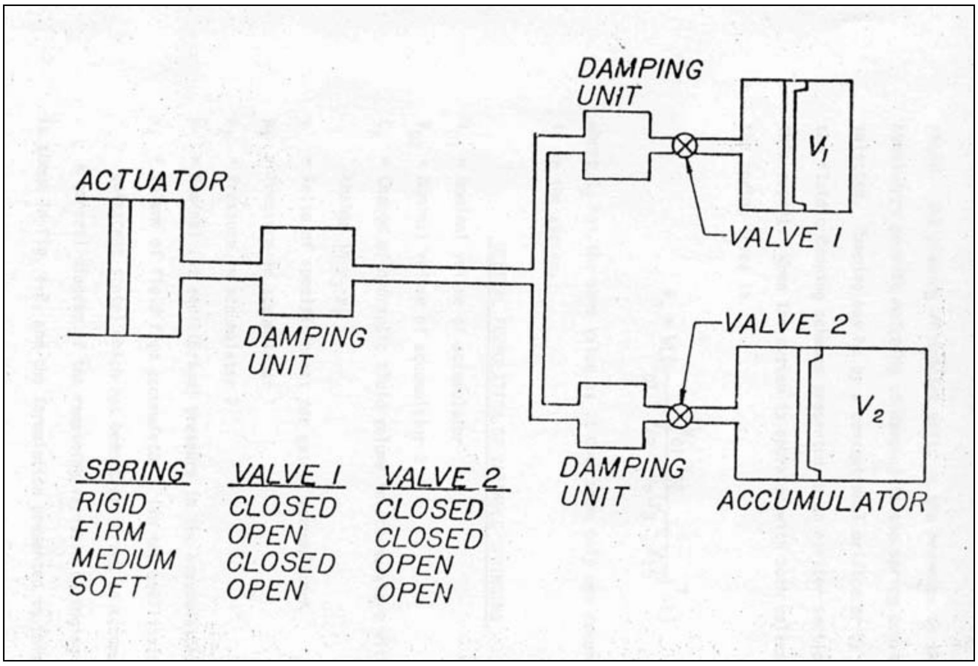

Figure 2-13: Variable spring rate suspension (parallel accumulators)

Another controllable spring system proposed by TACOM (1975) is an accumulator system connected in series. In principle, this concept works the same as accumulators connected in parallel. It is unknown if prototypes of these suspension concepts were ever built.

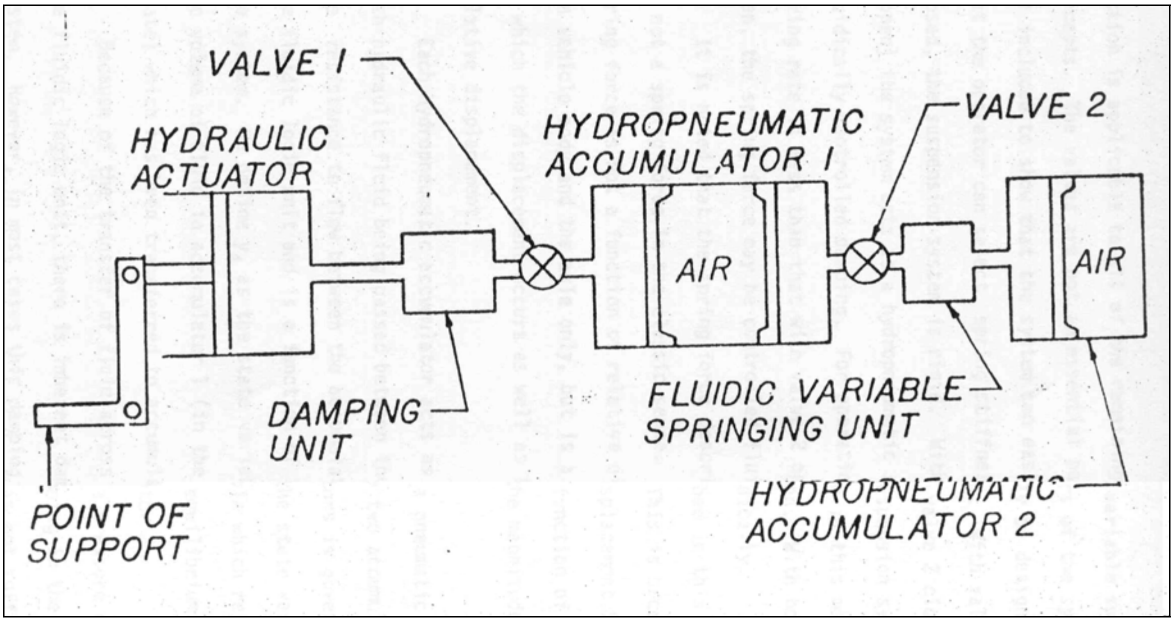

Figure 2-14: Variable spring rate suspension (series accumulators)

The twin-accumulator suspension described by Abd El-Tawwab (1997) was investigated. The twin accumulator suspension consists of two hydro-pneumatic springs in parallel and a control valve in series with each of the accumulators (see Figure 2-15).

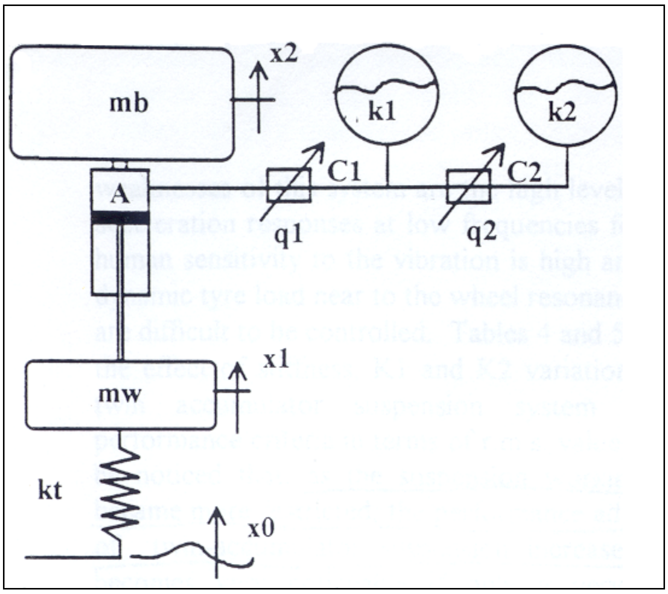

Figure 2-15: Twin accumulator suspension

The valves in this case are not switchable and have constant throttle properties. The accumulator parameters and throttle valve values can be chosen to result in an optimised, passive twin-accumulator suspension.

2.5 Closing

In closing the following conclusions can be made:

- Hydropneumatic springs are popular due the their non-linear characteristics.
- The semi-active hydro-pneumatic spring of this study was conceptually proposed by Karnopp & Heess (1991).
- No examples of semi-active hydro-pneumatic springs (exactly like to the one discussed in this study) could be found in the literature.
- An ideal gas approach is used by most researchers to model hydro-pneumatic springs.
- For pressures found in practical hydro-pneumatic springs, the ideal gas approach results in significant errors.
- The time constant approach for modelling heat transfer effects in hydro-pneumatic springs is in good agreement with experimental results.
- Parallel and serial accumulator suspension systems have been proposed, but not on the same scale (wheel loads) as in this study.
- Several systems employing controllable dampers have been developed and is currently used in production vehicles.
- The majority of controllable dampers employ fast acting valves, with adaptive control strategies.

****