Efficient Control Of Full-Bridge Oscillation Transformer

Kasper Juhl Mortensen

EMSD4

Master Thesis

AALBORG UNIVERSITET

Materials and Production
Electro-Mechanical System Design
Fibigerstræde 16
9220 Aalborg Øst

Title:
Efficient Control Of Full-Bridge Oscillation Transformer

Project:
Master Thesis

Project period:
1st of February to 30th of May 2025

Supervisor:
Per Johansen
Lasse Schmidt

Pages:
38

Appendices:
4

Participants:
Kasper Juhl Mortensen

Preface

In the creation of this project, generative AI has been used on occasion for the following purposes: spell checking, synonyms, sentence restructuring and grammar improvements; code suggestions, code review, and explanations; and initial brainstorming and research purposes.
Bold letters denote either vectors or matrices. All symbols are their definition can be found in the nomenclature.
All figures and tables have been produced by the author unless there is a source in the caption.

During the making of this project, the following programs have been used:

Overleaf LaTeX	For text processing and layout of this report.
MathWorks MATLAB and Simulink	For calculations, modelling, simulation, Real-time control and data processing.
Inkscape	For diagrams and illustrations.
SolidWorks	To view model of FBoT.

Kasper Juhl Mortensen

Aalborg University, 30th of May 2025

Abstract

This thesis explores efficient control strategies for the Full Bridge Oscillation Transformer (FBoT), a novel hydraulic transformer developed at Aalborg University. Hydraulic transformers aim to improve the efficiency of fluid power systems by replacing traditional proportional valves, which are prone to throttling losses.

The FBoT operates by oscillating a free-floating piston, enabling energy transfer via controlled actuation of On/Off valves assisted by pressure-controlled check valves. A high-fidelity simulation model of the system is developed to support analysis and control design. To minimise energy loss when the On/Off valves are opened or closed, a 12-step control sequence is proposed for pump-mode operation, reducing the control problem to determining six control timings which correspond to specific On/Off valves being open.

Due to the system's nonlinear dynamics and strong coupling, Reinforcement Learning (RL) is selected as the control approach. The symmetric structure of the FBoT is exploited to double the data collection rate by using both halves of each oscillation, which is critical since data acquisition was the bottleneck during training. The RL policy is trained iteratively using a reward function designed to minimise valve losses while achieving a centred oscillation at the desired amplitude.

Simulation results show a wide range of policy behaviours depending on the training stage. Notably, policies achieving the highest average rewards often fail to change the control timings in response to error signals. This behaviour stems from the reward function favouring policies that deliver near-optimal open-loop timings rather than learning to make corrections over time. Consequently, significant steady-state errors emerge in piston position and chamber pressures, leading to reduced efficiency due to large energy losses when the valves are opened.

Experimental validation demonstrates that while the trained policy can generate piston oscillations, it does so with large variations in the amplitude from oscillation to oscillation. Furthermore, a subpar efficiency was estimated.

Finally, a novel control framework is proposed to reduce system coupling through control signal and error transformations. This structure may enable the application of classical PI control for error correction, with the initial feedforward control signal set based on intuition or via reinforcement learning.

Nomenclature

Symbol	Definition
A_1, A_2	FBoT piston areas
A_{BP}	Boom piston area
Amp	Reference piston amplitude
B_p	Piston viscous friction coefficient
D_p	Pump displacement
e	Euler's Number
g	Gravitational acceleration
$k_{v1,CV}$	Orifice constant for check valve in transition
$k_{v2,CV}$	Orifice constant for check valve fully open
$k_{v,oo}$	Orifice constant for on/off valves
L	Overlap between piston and housing
L_k	Loss function for critic K
M	Mini batch size
m_p	FBoT piston mass
$\mathbf{N}_g$	Noise added to action during training
n_o	Poly-tropic index for bulk modulus
obs	Vector containing the observations
$\mathbf{obs}_N$	Vector containing the normalised observations
p_0	Initial pressure
$p_1, \dot{p}_1$	Pressure and gradient in chamber 1
$p_2, \dot{p}_2$	Pressure and gradient in chamber 2
$p_3, \dot{p}_3$	Pressure and gradient in chamber 3
$p_4, \dot{p}_4$	Pressure and gradient in chamber 4
$\Delta p_{1,L}$	Error in chamber 1's pressure at the left end
$\Delta p_{1,R}$	Error in chamber 1's pressure at the right end
$\Delta p_{2,L}$	Error in chamber 2's pressure at the left end
$\Delta p_{2,R}$	Error in chamber 2's pressure at the right end
Δp	General pressure difference
p_A	Generalised pressure
p_B	Generalised pressure
p_{BC}	Boom cylinder pressure and gradient
p_{CV}	Pressure difference to open check valve
p_{FO}	Pressure difference, check valve fully open
P_{in}	Power into the FBoT
P_{out}	Power out of the FBoT
p_s	Supply pressure
p_T	Tank pressure
Q_{1s}	Supply flow to FBoT chamber 1
Q_{1T}	Flow from chamber 1 to tank
Q_{2s}	Supply flow to FBoT chamber 2
Q_{2T}	Flow from chamber 2 to tank
Q_{3L}	Boom cylinder flow to chamber 3
Q_{3T}	Flow from chamber 3 to tank

Continued on next page

Symbol	Definition
Q_{4L}	Boom cylinder flow to chamber 4
Q_{4T}	Flow from chamber 4 to tank
Q_{BC}	Boom cylinder flow
Q_{CV}	Check valve flow
Q_{in}	Flow into FBoT
Q_k	Value function number k
$Q_{le,13}$	Leakage flow from chamber 1 to chamber 3
$Q_{le,1T}$	Leakage flow from chamber 1 to tank
$Q_{le,24}$	Leakage flow from chamber 2 to chamber 4
$Q_{le,2T}$	Leakage flow from chamber 2 to tank
$Q_{le,3T}$	Leakage flow from chamber 3 to tank
$Q_{le,4T}$	Leakage flow from chamber 4 to tank
Q_{out}	Flow out of FBoT
Q_p	Flow from pump
Q_{tk}	Target value function number k
Q_v	On/Off valve flow
R	Reward
S	State in operation sequence
States	Vector containing the measured states
t	Time
$t_{1..6}$	FBoT control timings
t_{end}	end time for each episode during training
t_l	Local time since last end value sample
$t_{p1,L}$	Local time where where p_1 crosses p_S while moving left
$t_{p1,R}$	Local time where where p_1 crosses p_T while moving right
$t_{p2,L}$	Local time where where p_2 crosses p_T while moving left
$t_{p2,R}$	Local time where where p_2 crosses p_S while moving right
$\Delta t_{p1,L}$	Excess time with pressure above p_S in chamber 1
$\Delta t_{p1,R}$	Excess time with pressure below p_T in chamber 1
$\Delta t_{p2,L}$	Excess time with pressure below p_T in chamber 2
$\Delta t_{p2,R}$	Excess time with pressure above p_S in chamber 2
$\mathbf{t}_a$	Control timings from agent
$\mathbf{t}_{aN}$	Normalised control timings from agent
$\mathbf{t}_{ctrl}$	Vector containing control timings
$\mathbf{t}_{sat}$	Saturated control timings
$\mathbf{t}_{satN}$	Normalised saturated control timings
T_d	Valve opening dead time
$T_{fully-open}$	Time until valve fully opened
T_s	Sample time
$Trig$	Variable used to trigger the agent subsystem and define measurement side
V_{01}	Dead volume in FBoT chamber 1,4
V_{02}	Dead volume in FBoT chamber 2,3
$x_{BP}, \dot{x}_{BP}$	Boom piston position and velocity
$x_p, \dot{x}_p, \ddot{x}_p$	FBoT piston position, velocity and acceleration

Continued on next page

Symbol	Definition
x_v	Orifice valve opening
$x_{v,ref}$	Valve position reference
$\mathbf{x}_{v,ref}$	Vector containing valve position reference for all valves
$\Delta x_{p,L}$	FBoT piston position error measured at the left end
$\Delta x_{p,R}$	FBoT piston position error measured at the right end
$\Delta x_{pref,L}$	FBoT piston position reference for the left end
$\Delta x_{pref,R}$	FBoT piston position reference for the left end
y_i	Value function target for i'th experience
α	Air volume fraction content of fluid
β_0	Constant bulk modulus
β_{eff}	Effective fluid bulk modulus
δ	Gap between piston and housing
η	Efficiency
η_{vp}	Volumetric efficiency of pump
γ	Discount factor for future rewards
μ	Viscosity
ω_p	Angular Speed of pump
π	Policy function
π_t	Target policy function
ϕ_k	Parameters for the k'th critic
ϕ_{tk}	Parameters for the k'th target critic
ρ	Density
τ	Periodic smoothing factor
τ_v	Time constant for valve opening dynamics
θ	Parameters for the actor
θ_t	Parameters for the target actor

Abbreviations

Symbol	Definition
DDPG	Deep Deterministic Policy Gradient
EMSD	Electro Mechanical Systems Design
FBoT	Full-Bridge Oscillation Transformer
ILC	Iterative Learning Control
MIMO	Multiple Input Multiple Output
MPC	Model Predictive Control
NN	Neural Network
RGA	Relative Gain Array
RL	Reinforcement Learning
SISO	Single Input Single Output
TD3	Twin-Delayed Deep Deterministic

Contents

1 Introduction

The low energy efficiency of off-road hydraulics is increasingly under scrutiny due to its significant contribution to environmental pollution. These machines typically have an average energy efficiency of 21%, with excavators sometimes as low as 12% (Hansen and Schmidt 2024). The primary source of energy loss in excavators lies in the hydraulic system, which operates at only 30% efficiency (Danfoss 2023, p. 17). The climate impact of these machines in the construction sector is comparable to that of international aviation, which receives much greater attention and scrutiny (Hansen and Schmidt 2024).

A key factor behind the inefficiency of traditional fluid power systems is the widespread use of conventional proportional valves. These valves restrict flow, resulting in significant throttling losses. Alternatives being explored include electro-hydraulic drives and hydraulic transformers.

1.1 The Principle Behind Hydraulic Transformers

The concept of an ideal hydraulic transformer is to maintain a constant power level $P = p \cdot Q$ by converting hydraulic power along its constant power curve, seen in Figure 1.1. This system allows for the adjustment of pressure while maintaining a steady power output. In contrast, throttle valves adjust the desired power output by reducing pressure without altering the flow, as depicted in Figure 1.1. Here, an actuator requires a flow Q and the pressure p_B with a supply pressure of p_A. With a throttle valve, a flow Q is required at the supply side, whereas with the ideal transformer, a flow lower than Q on the supply side can be transformed to the desired flow.

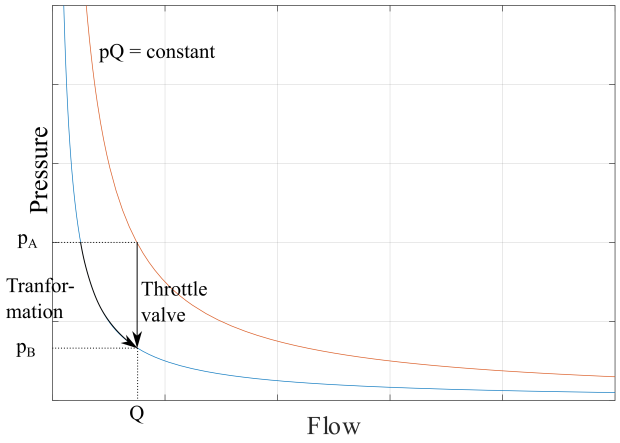


Figure 1.1. Plot showing the difference between a power-conserving transformation and a throttle valve. The power-conserving transformation follows the power-conserving line, where as the throttle valve goes from a higher energy state to a lower power state to supply the desired flow Q at the pressure p_B. (Mortensen et al. 2025)

This pressure drop through the throttle valve corresponds to an energy loss. The efficiency of a valve can be expressed using Equation (1.1):

$$\eta = \frac{p_B \cdot Q}{p_A \cdot Q} = \frac{p_B}{p_A} \tag{1.1}$$

Valves operate most efficiently when the pressure drop across them is minimal, which can be managed by regulating the supply pressure. However, in construction machinery, a common-pressure rail is typically used to supply all cylinders, as there is only one pump serving them. Since different cylinders require varying pressures to achieve the desired movement, it becomes impractical to maintain low-pressure drops across all valves.

A more efficient alternative is the use of hydraulic transformers, which supply the required flow to each cylinder with minimal energy loss. Additionally, hydraulic transformers enable energy recovery. For instance, when lowering a load, the fluid can be transformed from load pressure to supply pressure and stored in a high-pressure accumulator until needed, thereby further enhancing overall energy efficiency.

1.2 Initial Problem Statement

One of the potential transformer types is the Full-Bridge Oscillation Transformer (FBoT) designed and tested by Anders Hedegaard Hansen and Per Johansen. The FBoT has a mechanically simple design, but requires a complicated control system to work properly (Johansen and Hansen 2023).

This leads to the initial problem statement:

How does the Full-Bridge oscillation Transformer work?

System Description 2

The FBoT consist of a free-floating piston in a housing. The movement of the piston is controlled with 8 ON/OFF valves. For safety reasons and to simplify the control, 8 check valves are implemented. This means that 4 valves are connected to each of the chambers, with p_1 corresponding to chamber 1, and so on in the general schematic of the FBoT can be seen in Figure 2.1.

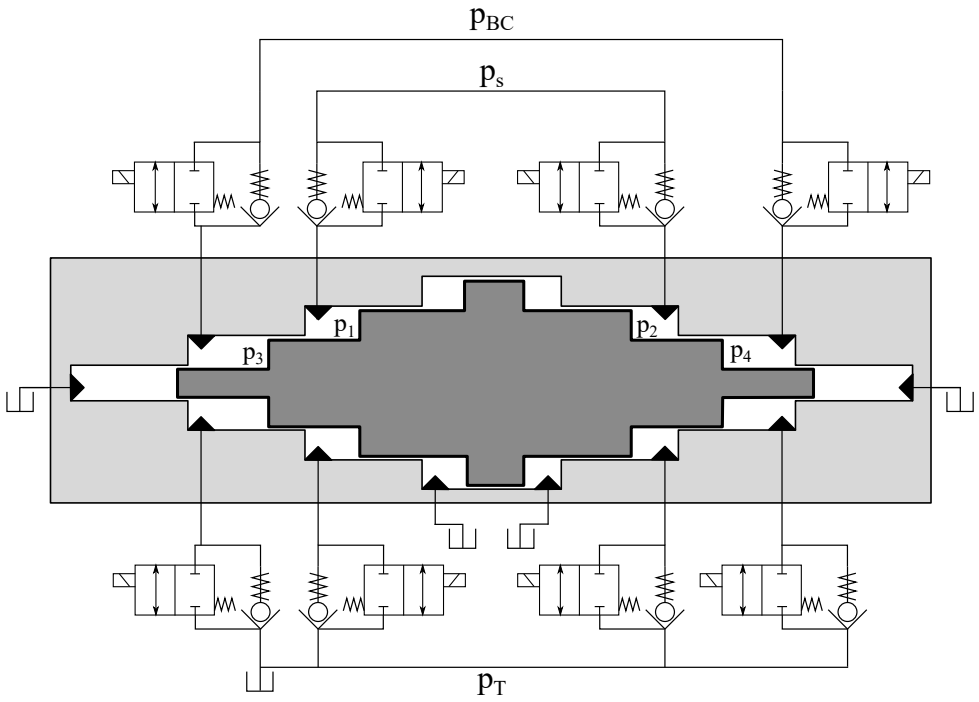


Figure 2.1. General schematic of the FBoT.

The FBoT can either transform from p_S to p_{BC}, called pump mode, or from p_{BC} to p_S called regenerative mode. In pump mode, p_s is used to drive the movement of the piston, whereas in regenerative mode p_{BC} generates the movement.
The transformation consists of 12 distinct states, which are described below for pump mode. The principle is the same for regenerative mode. An animation of the transformation can be found with the following link: `https://drive.google.com/file/d/15Ox8ZqvGdZykMcbXJ29rVbOOIHSdvWdE/view?usp=sharing`. The animation is at $1/500$ of the real speed, with the arrows indicating the flow of oil, t indicating the time and S indicating the state corresponding to the list below. Note that transient valve dynamics have been ignored in the animation.

1. The kinetic energy of the piston is increased by connecting chamber 1 to the supply and the rest to the tank.
2. Chamber 4 is isolated so that pressure in it starts increasing, by converting the kinetic energy to pressure energy.
3. The pressure in chamber 4 surpasses p_{BC} and oil starts to flow through the check valve.
4. Chamber 1 or 2 is isolated here, depending on which takes the longest time to reach the desired pressure to minimise valve losses when going from state 6 to 7. (Illustrated with chamber 1 being isolated first)
5. Both chambers 1 and 2 are isolated so that chamber 1 reaches p_t at the same time as chamber 2 reaches p_S.
6. The pressure in chamber 1 becomes lower than the tank pressure (p_T) and the check valve is opened. The pressure in chamber 2 becomes larger than p_s and the check valve is opened. The time in this state should be minimised since it is pure loss.
7. Chamber 2 is connected to supply and chambers 1 and 3 are connected to tank. This increases the kinetic energy of the piston.
8. Chamber 3 is isolated so that pressure in it starts increasing, by converting the kinetic energy to pressure energy.
9. The pressure in chamber 3 surpasses p_{BC} and oil starts to flow through the check valve.
10. Chamber 1 or 2 is isolated here, depending on which takes the longest time to reach the desired pressure to minimise valve losses when going from state 12 to 1. (Illustrated with chamber 2 being isolated first)
11. Both chambers 1 and 2 are isolated so that chamber 1 reaches p_S at the same time as chamber 2 reaches p_T
12. The pressure in chamber 2 becomes lower than p_T, and the check valve is opened. The pressure in chamber 1 becomes larger than p_s, and the check valve is opened. This ends with the piston at a standstill.

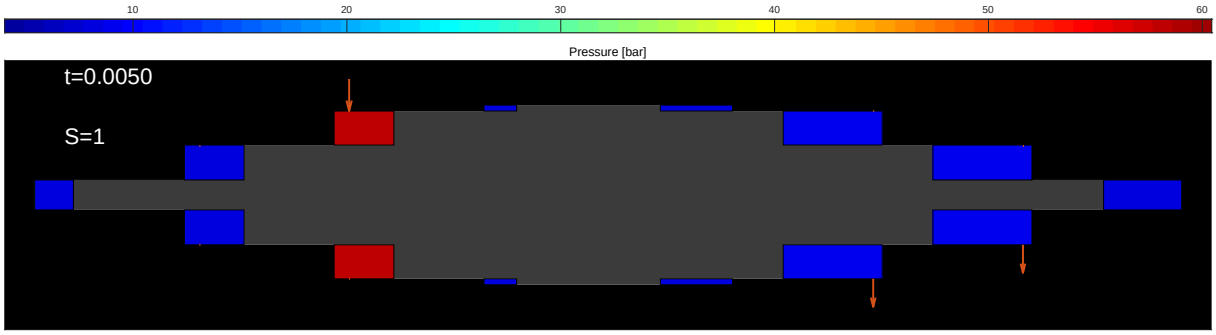
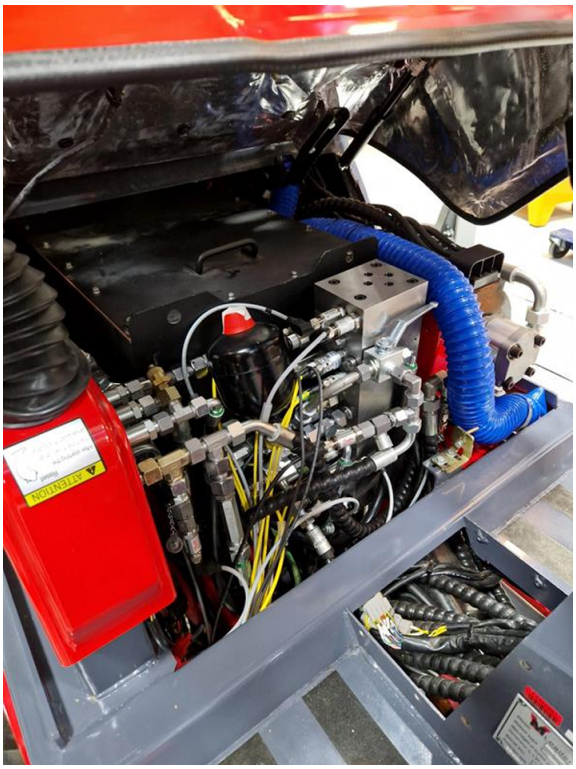


Figure 2.2. Still picture from the animation of the FBoT operation principle available at `https://drive.google.com/file/d/15Ox8ZqvGdZykMcbXJ29rVbO0IHSdvWdE/view?usp=sharing`. Arrows indicate flow, t indicates the time, with S indicating the state from the list above. For the animation $p_S = 60\,\text{bar}$ and $p_{BC} = 30\,\text{bar}$

2.1 Lab Setup

To evaluate the performance of the FBoT, it is integrated into a wheel loader within a laboratory setting. In this setup, chambers 3 and 4 of the FBoT are connected to the wheel loader's boom cylinder. The lifted mass can be adjusted by varying the number of weight plates on the pallet. Figure 2.3a illustrates the wheel loader, while Figure 2.3b shows the FBoT installed in the system. The setup is equipped with position sensors to measure x_p and x_{BP} along with pressure sensors to measure p_1 to p_4, p_S, p_T and p_{BC}. An encoder is used to measure the pump speed. All sensors are sampled at 10 kHz.

(a) Laboratory setup.



(b) Implementation of the FBoT in the wheel loader.

Figure 2.3. Laboratory setup (Johansen and Hansen 2023).

2.2 Problem Formulation

Based on the introduction and system description, the problem formulation is:

How is the FBoT controlled to minimise valve losses and accurately track a positional amplitude reference?

System Modelling 3

In this chapter, the system is modelled so that it can be utilised to design a controller. The modelling in this chapter is, unless otherwise stated, based on the book "Fluid Power Systems: A Lecture Note in Modelling, Analysis and Control" by Anders Hedegaard Hansen (Hansen 2023). The modelling is based on the hydraulic diagram and notation shown in Figure 3.1.

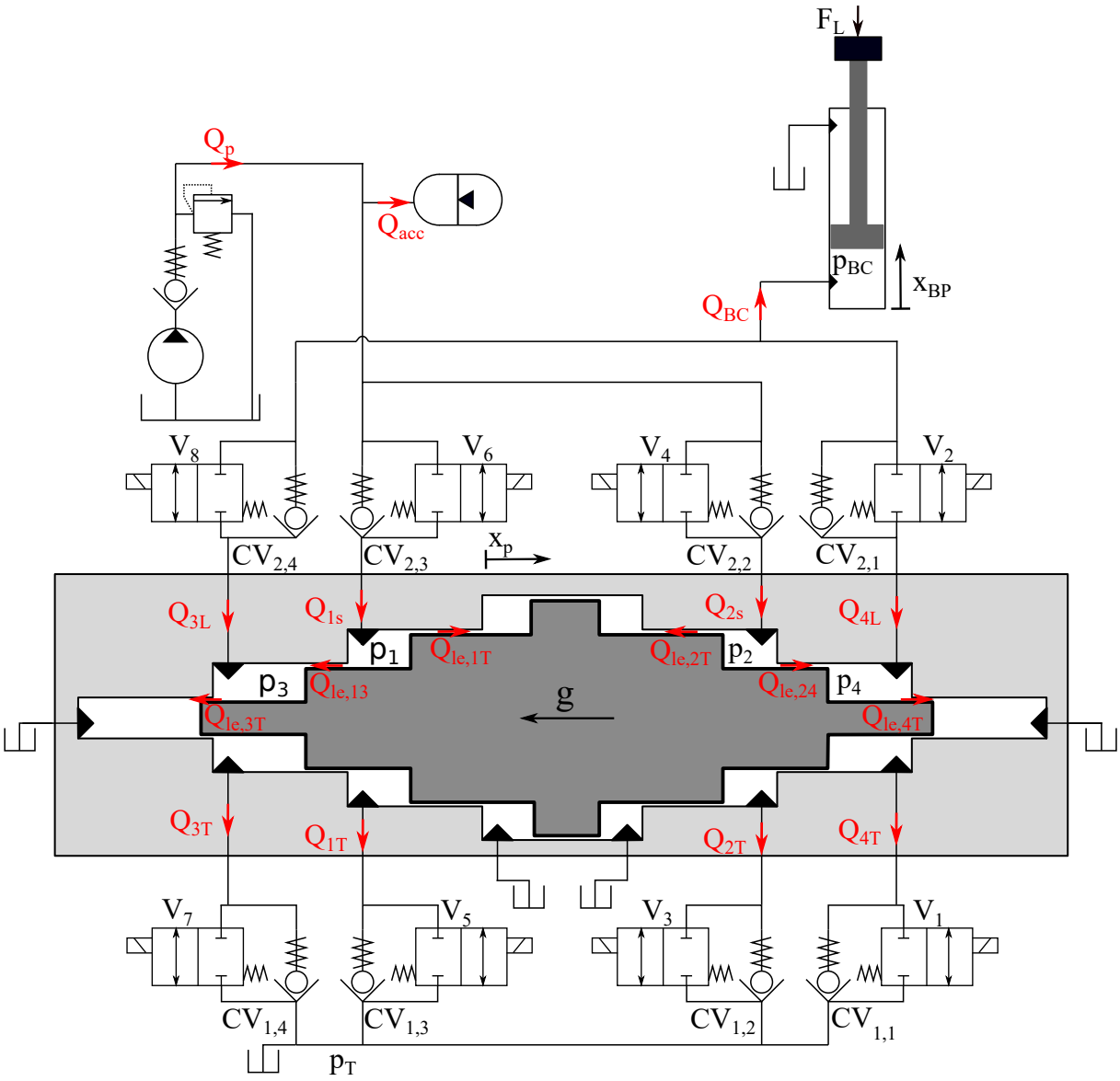


Figure 3.1. Hydraulic diagram of FBoT in the laboratory setup

3.1 Bulk Modulus

The stiffness of the fluid is described with the bulk modulus. It varies considerably with pressure due to the air content in the oil, particularly at low pressures. This is especially important because the fluid pressure in the FBoT's chambers ranges between tank pressure and supply pressure. This can be modelled with Equation (3.1).

$$\beta_{eff} = \frac{(1-\alpha)\cdot e^{\left(\frac{p_0-p}{\beta_0}\right)} + \alpha\cdot\left(\frac{p_0}{p}\right)^{\frac{1}{n_o}}}{\frac{1-\alpha}{\beta_0}\cdot e^{\left(\frac{p_0-p}{\beta_0}\right)} + \frac{\alpha}{n_o\cdot p_0}\cdot\left(\frac{p_0}{p}\right)^{\left(\frac{n_o+1}{n_o}\right)}} \tag{3.1}$$

3.2 Supply Manifold Model

The pump itself is a flow machine driven by a variable-speed electric motor. The flow of the ideal pump is defined by Equation (3.2).

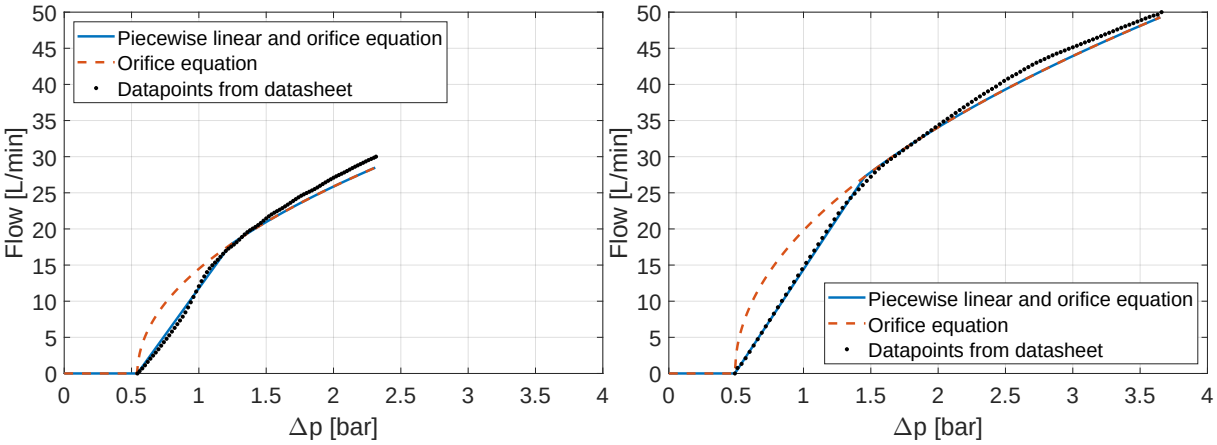
$$Q_p = D_p \cdot \omega_p \tag{3.2}$$

The speed of the motor is controlled with a frequency converter, which is controlled by the computer in the laboratory. This is utilised with the feedback pressure controller designed by (Mortensen et al. 2025) to achieve an approximately constant supply pressure, which leads to the supply pressure being modelled as a constant.

3.3 Check Valves

The check valve's function is to only allow fluid flow in one direction when a specific pressure difference, which varies depending on the valve, is reached. Assuming the valve's spring behaves linearly, the opening will increase proportionally with the pressure difference until it reaches its fully open state.

In the FBoT, two different check valves are utilised: the Bucher RKVG-08, which is installed between the tank and the FBoT, and the Bucher RKVE-G-08, used for the load and supply connections. The datasheets for these valves are provided in Appendix D. The black dots in Figure 3.2 represent the flow characteristics of both check valves. Rather than following the standard orifice equation from the initial opening at the check valve pressure difference threshold (p_{CV}), the data indicates a linear transition phase leading to what is assumed to be the fully open pressure (p_{FO}). At the lower pressure differences, this linear region may result from a laminar flow regime. As the pressure drop increases so does the fluid flow, which leads to a flow regime dominated by turbulence, causing the nonlinear behaviour described by the orifice equation.



(a) Modelling of flow characteristic of RKVG-08 ***(b)*** Modelling of flow characteristic of RKVE-08

Figure 3.2. Modelling of flow characteristic of the check valves. (Mortensen et al. 2025)

The flow through a check valve will therefore be modelled as in Equation (3.3). Here p_A and p_B are pressures on each side of the check valve, as seen in Figure 3.3, p_{CV} is the minimum pressure to overcome the spring tension and p_{FO} is the pressure where the check valve is fully opened.

$$Q_{CV} = \begin{cases} 0, & (p_A - p_B) < p_{CV} \\ k_{v1,CV} \cdot (p_A - p_B - p_{CV}), & (p_A - p_B) \geq p_{CV} \;\&\; < p_{FO} \\ k_{v2,CV} \cdot \sqrt{(p_A - p_B) - p_{CV}}, & (p_A - p_B) \geq p_{FO} \end{cases} \tag{3.3}$$

Figure 3.3. Illustration of a check valve. (Mortensen et al. 2025)

3.4 On/Off Valves

The On/Off valves control the movement of the FBoT piston. To model this, the flow through the On/Off valves has to be modelled, which can be done with the orifice equation.

$$Q_v = x_v \cdot k_{v,oo} \cdot \sqrt{\frac{2}{\rho} |p_A - p_B|} \cdot sign\,(p_A - p_B) \tag{3.4}$$

Here, x_v is the normalised valve position between 0 and 1. From Figure 3.4, which comes from the data sheet in Appendix C, the term $k_{v,oo} \cdot \sqrt{\frac{2}{\rho}}$ can be derived. Once the fluid density is known, $k_{v,oo}$ can be determined yielding $k_{v,oo} = 4.88 \cdot 10^{-6}$.

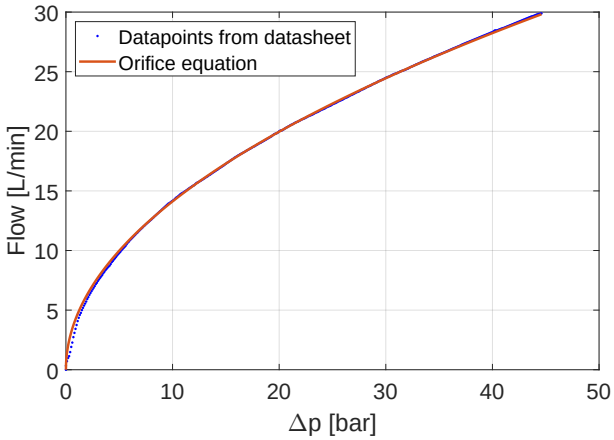


Figure 3.4. Relation between flow and pressure drop for the on/off valves. (Mortensen et al. 2025)

The valve opening exhibits a transient response that consists of a time delay and a transient response resembling a first-order system (Andersen et al. 2023, p. 17). The time delay and first-order response can be modelled using Equation (3.5) (MathWorks 2025b).

$$\frac{x_v}{x_{v,ref}} = e^{-T_d \cdot s} \cdot \frac{1}{1 + \tau_v \cdot s} \tag{3.5}$$

The opening times given in the datasheet are very dependent on fluid pressure, viscosity, and flow. The approximate time constant and dead time of $T_d = 0.0035s$ & $\tau_v = 0.0030$ will be used, which is at the fast end of the average time for energising and deenergising according to the datasheet in Appendix C.

3.5 FBoT Chamber Models

The continuity equation is used to model the 4 chambers of the FBoT, which are not always connected to the tank. This gives Equations (3.6) to (3.9).

$$Q_{1s} - Q_{le,1T} - Q_{1T} - Q_{le,13} = \dot{x}_p \cdot A_2 + \frac{V_{01} + x_p \cdot A_2}{\beta eff(p_1)} \cdot \dot{p}_1 \tag{3.6}$$

$$Q_{2s} - Q_{le,24} - Q_{2T} - Q_{le,2T} = -\dot{x}_p \cdot A_2 + \frac{V_{02} - x_p \cdot A_2}{\beta eff(p_2)} \cdot \dot{p}_2 \tag{3.7}$$

$$Q_{3L} + Q_{le,13} - Q_{3T} - Q_{le,3T} = \dot{x}_p \cdot A_1 + \frac{V_{02} + x_p \cdot A_1}{\beta eff(p_3)} \cdot \dot{p}_3 \tag{3.8}$$

$$Q_{4L} + Q_{le,24} - Q_{4T} - Q_{le,4T} = -\dot{x}_p \cdot A_1 + \frac{V_{01} - x_p \cdot A_1}{\beta_{eff}(p_4)} \cdot \dot{p}_4 \tag{3.9}$$

The leakage flow depends on the piston's position and the pressure difference between the two chambers. The piston's position determines the overlap length between the piston and the housing, which affects the leakage. When the diameter is sufficiently large, the leakage flow resembles the flow between two parallel plates, as described by Equation (3.10). This equation applies to a steady-state scenario, neglecting the effects of fluid inertia.

$$Q = \frac{\delta^3 \cdot \pi \cdot D}{12 \cdot \mu \cdot L} \cdot \Delta p \tag{3.10}$$

Equation (3.10) highlights the significant influence of the gap δ between the piston and housing on leakage. This gap is determined by the tolerances on the piston and the housing where the mean of the two tolerances is used here. However, the steady-state assumption is invalid for the FBoT as the piston direction changes rapidly, preventing the fluid flow from fully developing a steady-state profile. Upon studying this in simulation, the actual leakage used was scaled by a factor 10^{-4} giving more resemblance to the actual behaviour.

The movement of the FBoT piston can be described using Newton's second law. This is done in Equation (3.11) where a viscous friction model is utilized. Furthermore, gravity is included since the piston moves vertically in the FBoT.

$$\ddot{x}_p \cdot m_p = (p_1 - p_2) \cdot A_2 + (p_3 - p_4) \cdot A_1 - \dot{x}_p \cdot B_p - m_p \cdot g \tag{3.11}$$

3.6 Load Model

The FBoT's load side is connected to the boom cylinder in the laboratory setup. The pressure in the boom cylinder varies depending on the load and the position of the boom. To simplify the training process in Chapter 4, the boom cylinder pressure, p_{BC}, is modelled as a constant which can be varied during training.

Control 4

Throughout this chapter, a control approach is developed for the FBoT operating in pump mode. The same principles can be used for regenerative mode by substituting p_1 and p_2 with p_3 and p_4. The relevant state measurements are determined before different control approaches are investigated. This leads to the implementation of a Reinforcement Learning (RL) based control strategy.

4.1 Control Considerations

The frequency of the FBoT piston's movement is fairly high compared to the response times of the valves. This makes it difficult to control it with classic feedback control since the valves will close too late here due to their transient response. Furthermore, to reduce losses when the valves are switched, pressure transition periods are needed where the chambers go from either high to low pressure or low to high. This is illustrated with the 12 steps from Chapter 2 in Figure 4.1. Due to this, it is determined that some sort of feed-forward control is required. To do this, the timings for states 1-4 and 7-10 are used as control signals. States 5-6 and 11-12 have the same ON/OFF valve positions, which are held until the piston is at standstill.
Furthermore, it should be noted that for states 4 and 10, there are two different options depending on which chamber takes the longest to achieve the desired pressure at standstill. In states 4 and 10, the specific substate is chosen based on the sign of the corresponding time.

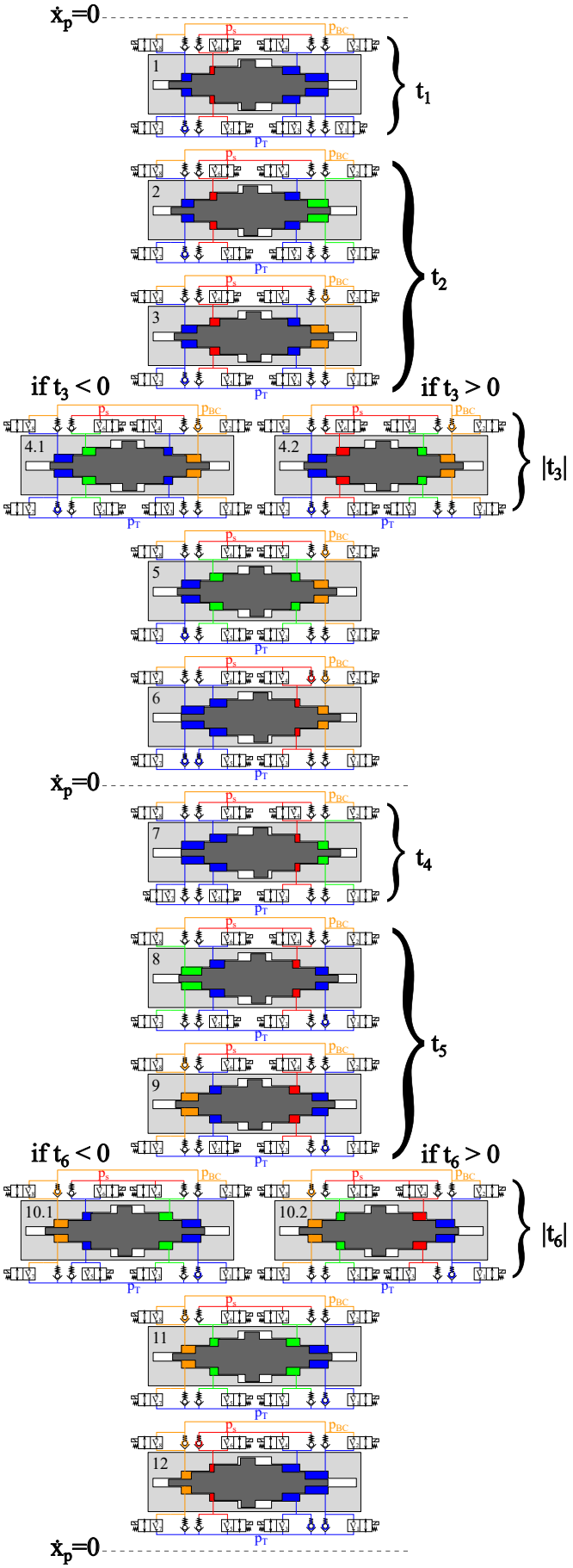


Figure 4.1. FBoT states when operating in pump mode. A larger version can be seen in Appendix A.

4.1.1 State Measurement

Due to the valves being fairly slow compared to the oscillation of the FBoT piston, the specific movement cannot be controlled to, for example, a Sine wave.
Due to this, the states are only measured at each end of the piston oscillations, which here is defined as when the velocity is 0. It is desired to control the amplitude of the piston oscillation while keeping it centred. The position error at the endpoints is therefore measured, as illustrated in Figure 4.2 with separate reference values and measurements for the two ends. Here, the subscript "$_{,R}$" indicates that it is a measurement at the right end of the oscillation relative to Figure 4.1, whereas the subscript "$_{,L}$" indicates that it is a measurement at the left end of the oscillation.

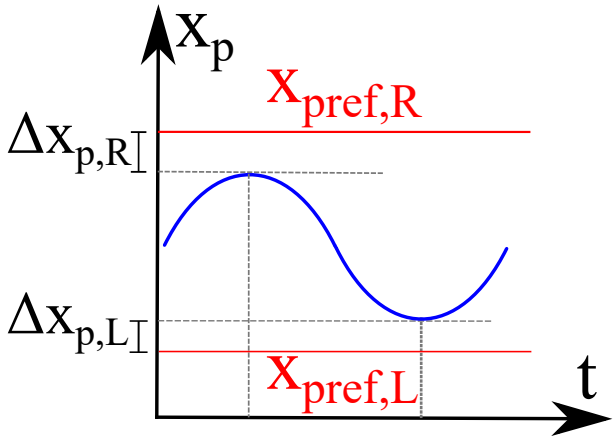
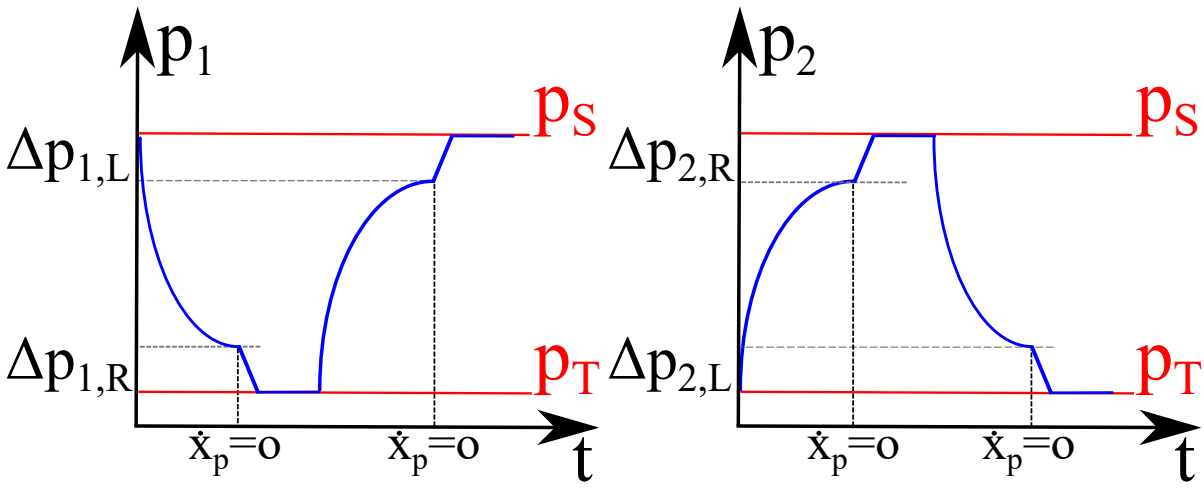


Figure 4.2. Measurement of position error at the two ends poits where $\dot{x}_p = 0$.

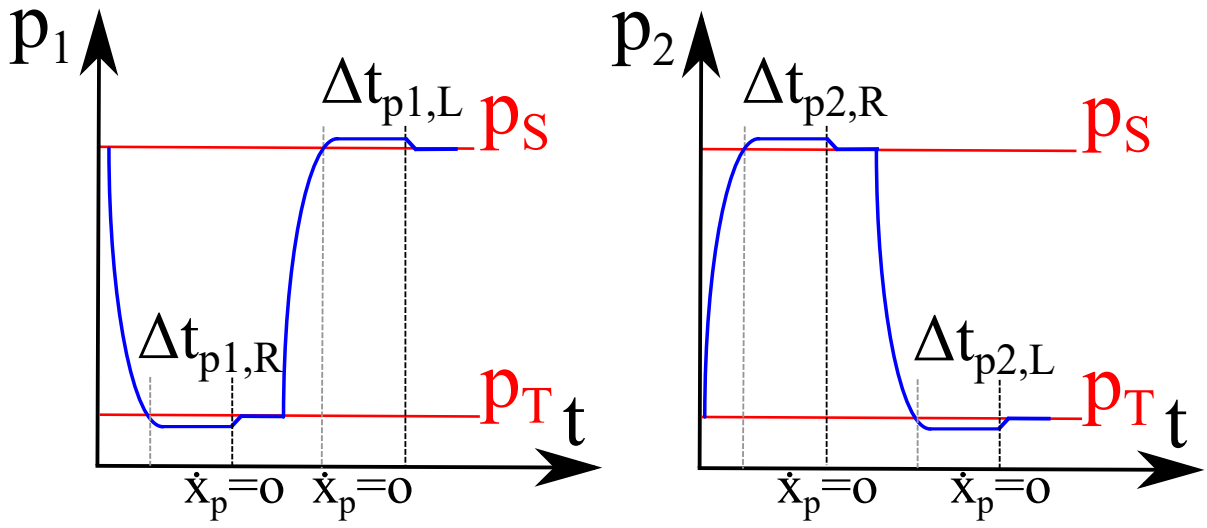
To reduce losses, it is desired that states 6 and 12, from figure 4.1, are reached so that the pressure drops over the valves are minimised when they are opened. This is measured by the pressure difference between the desired pressure, p_S or p_T, and the chamber pressure for each side when $\dot{x}_p = 0$. This is illustrated in Figure 4.3.



(a) Chamber 1 ***(b)*** Chamber 2

Figure 4.3. Measurement of the pressure difference between desired pressure,p_S or p_T, and the chamber pressure when at end points where $\dot{x}_p = 0$

However, spending excess time in states 6 and 12, where the desired pressure is reached, leads to unnecessary losses. Due to the check valves, the pressure does not increase or decrease significantly beyond the desired pressure. The time since the desired pressure has been reached is therefore added as a state for each side and chamber. This is illustrated in Figure 4.4 where the desired pressure is reached at the grey dashed line while the piston is at a standstill at the black dashed line. It is desired that this time becomes as small as possible.



(a) Chamber 1 *(b)* Chamber 2

Figure 4.4. Excess time spent at the desired pressure state, corresponding to states 6 and 12, before $\dot{x}_p = 0$ is measured from the grey dashed line to the black dashed line.

4.1.2 MIMO System

The system has multiple inputs and multiple outputs and is therefore a MIMO system. For MIMO systems, it is relevant to investigate the coupling between the inputs and outputs. This is done to determine if SISO control can be utilised. This is usually done with an RGA analysis but that is not possible here due to the choice of control signals.

The coupling has instead been investigated by increasing t_1, t_2 and t_3 independently by 2 ms at three different reference points while measuring the change in error measurements. It is only done for t_1 to t_3 and the corresponding measurements at the right end, since the FBoT is symmetrical except for gravity. Table 4.1 shows that t_1 and t_2 effects on the position error are similar at the three reference points, whereas their effect on the pressure-related states varies significantly. While at some points being negligible, that is not the case for all situations. t_3 generally has the largest effect on the pressure-related states, but the effect varies significantly between the three reference points. For the position-related errors, the effect of t_3 changes a lot but is generally lower than for t_1 and t_2.

Based on this, it is concluded that the system shows significant coupling and non-linearities at times.

Reference	Changed time	$\Delta(\Delta x_{p,R})$ [mm]	$\Delta(\Delta p_{1,R})$ [bar]	$\Delta(\Delta p_{2,R})$ [bar]	$\Delta(\Delta t_{p1,R})$ [ms]	$\Delta(\Delta t_{p2,R})$ [ms]
$t_1 = 2$, $t_2 = 2$, $t_3 = 1$	t_1	-3.59	1.00	0.160	0	1.10
	t_2	-2.83	-0.515	-0.0190	0	0.200
	t_3	-1.29	-11.9	-0.868	0	2.20
$t_1 = 0$, $t_2 = 4$, $t_3 = -1$	t_1	-4.11	1.77	-7.75	0	0
	t_2	-2.50	0.331	-2.12	0	0
	t_3	$8.22 \cdot 10^{-3}$	-6.31	-15.8	0	0
$t_1 = 10$, $t_2 = 0$, $t_3 = -1$	t_1	-4.61	$6.58 \cdot 10^{-4}$	-2.71	0.500	0
	t_2	-3.78	$4.48 \cdot 10^{-4}$	0.944	-0.300	0
	t_3	2.12	-2.13	-11.8	-3.30	3.10

Table 4.1. The effect of changing the control timings, t_1, t_2 or t_3, by 2 ms at three different reference points. Note that t_3 being negative means that state 4.1 is used instead of state 4.2 in Figure 4.1.

4.1.3 Control Approaches

There are multiple possible control approaches for the FBoT, each with its advantages and disadvantages. Here, Model Predictive Control (MPC), Reinforcement Learning (RL) and Iterative Learning Control (ILC) are investigated.

MPC is an optimal control technique where the goal is to find the control actions which minimise the defined cost function over a finite time horizon. To solve this optimisation problem, a fast and accurate model of the system is needed to get the proper control signals. (MathWorks 2025e) Due to this, MPC has not been investigated further since the current model can only be executed at approximately real time, which is too slow.

RL is a branch of machine learning which interacts with a dynamic environment and learns from these interactions based on trial and error. A reward function is used to quantify if the applied action was good or bad. This is then used to update the policy. Some advantages of RL are that it can learn directly from the non-linear model and that once it is trained, the final policy is generally fast to execute. A disadvantage is that it is sample-inefficient, which means that it can take a long time to train the algorithm.(MathWorks 2025f)

ILC updates the control parameters from iteration to iteration which is useful for systems that repeat the same motion again and again as the FBoT. The initial conditions, however, have to be the same, which is likely not the case for the FBoT since the movement in one direction can start from different positions. To use ILC for MIMO systems, a model of the relation between the control signals and errors is needed which as shown by (Mortensen et al. 2025) was difficult to achieve.

Based on this initial investigation into different control approaches, it has been decided to move forward with an RL-based control approach.

4.2 Reinforcement Learning

As stated earlier, reinforcement learning (RL) could potentially be an approach to control the FBoT. Firstly, general terminology is explained, after which important choices are considered and the general implementation is explained.

4.2.1 RL Structure

RL is a subsection of machine learning, separate from unsupervised and supervised learning. Resembling supervised learning, where an output is predicted, gets a correct or incorrect result and updates the policy to learn the right patterns. RL uses a reward function to determine if the applied control signals were good or bad. RL is therefore regarded as a data-driven control using machine learning techniques. RL iterates towards an optimum through enough data exploration. An illustration of an RL network is seen in Figure 4.5, and the RL terminology and the control analogy then becomes (MathWorks 2025a):

Agent - The agent is a combination of a policy (control law) and a reinforcement learning algorithm. The goal of the agent is to create a policy which can be deployed independently of the agent.

Policy - Control law The policy takes the observations as input and gives actions as output. The goal of the training process is to determine the optimal policy.

Environment - Plant + more The environment is everything which is not the agent. This includes system dynamics (plant), measurements, disturbances and reward calculations. The environment can also include reference values if the desired behaviour changes.

Observation - measured output Not to be confused with observers, observation in RL is everything that is given as input to the policy. This can be measured states, reference values, integrals or derivatives of states, previous actions or whatever is needed to determine the next appropriate action.

Action - Control signal The actions are similar to the control signals in control problems in that it is signals that are applied to the environment. Depending on the type of agent, both Discrete and Continuous actions can be applied.

Reward - Optimality criteria A custom function that RL agents try to maximise, i.e. take actions that provide the highest reward.

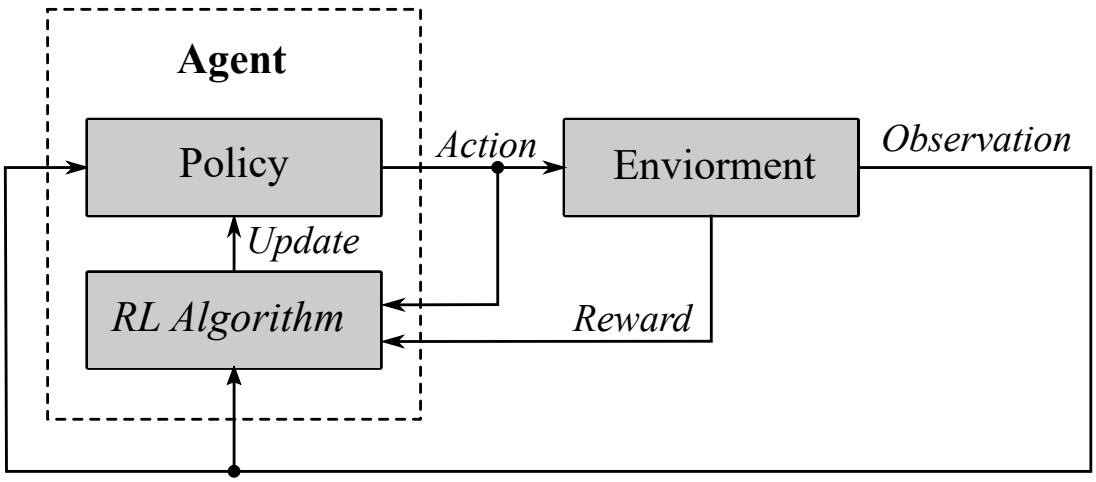


Figure 4.5. Working principle of reinforcement learning.

Beyond the reward function, there are multiple hyperparameters which affect the learning process. This has todo with the optimisation process, exploration and utilisation of the sampled data.

Optimizer Options

There are multiple optimization options that guide the learning process for RL. Some of them are:

Learn Rate Affects how fast the actor and critic learns during training. A too-large learn rate will lead to instability, whereas a too-low learn rate will make the training take a long time.

Gradient Threshold - This is used to clip the gradient to the specified value if it exceeds it. This is done to limit how much the model parameters can change each iteration to improve stability

Exploration vs. Exploitation

This is the tradeoff between using what the agent has already learned (exploitation) and exploring new actions to see if they are better. Initially, large exploration is desired to get the general area of the optimal policy, whereas lowering the exploration will help fine-tune the parameters. The exploration is done by adding a Gaussian-distributed noise signal to the action signal from the Actor.

Experience Buffer

The agent stores the experiences, which consist of the Action, reward, observation and next observation in the experience buffer. The length of the experience buffer can be varied. From the experience buffer, a random batch of experiences is sampled. This batch is used to update the Critic. A large experience buffer reduces variance but increases the computational expense. The number of times the agent learns from the data in the experience buffer can also be changed, which is known as the number of epochs.
The parameters related to the experience buffer are generally tuned based on whether the data gathering or learning process takes the longest. For a slow data-gathering process, the number of epochs is increased along with the mini-batch size.

Parallelisation

Parallelisation can be utilised to increase the creation of data from the environment, as multiple combinations of environment and agents are simulated at the same time. This can greatly increase the data available for the learning process.
The training can be run synchronously, where the workers wait until all workers have finished or in asynchronous mode, where the worker gets a new set of agents and environment when it has completed its task. Asynchronous is faster but less reproducible due to slight variation in worker execution time, which can give significantly different results in the end (MathWorks 2025c).

4.3 Reinforcement Learning For FBoT Control

With general information about RL given above, some specific choices can be made with regard to implementing it to control the FBoT. To train the RL Agent, the data from the simulation has to be read, the agent has to be executed at the right time, and the determined actions have to be applied. To do this, the system shown in Figure 4.6 is used. The Agent and all the functions within the grey dashed box in Figure 4.6 run at a variable sample time, double the piston oscillation frequency, since it is only executed at each end. The "State Measurement" and "Timings To Valve Signals" Algorithms are executed with a constant sample rate of 10 kHz to mimic the laboratory setup. The function of the different blocks is described in further detail in the following paragraphs.

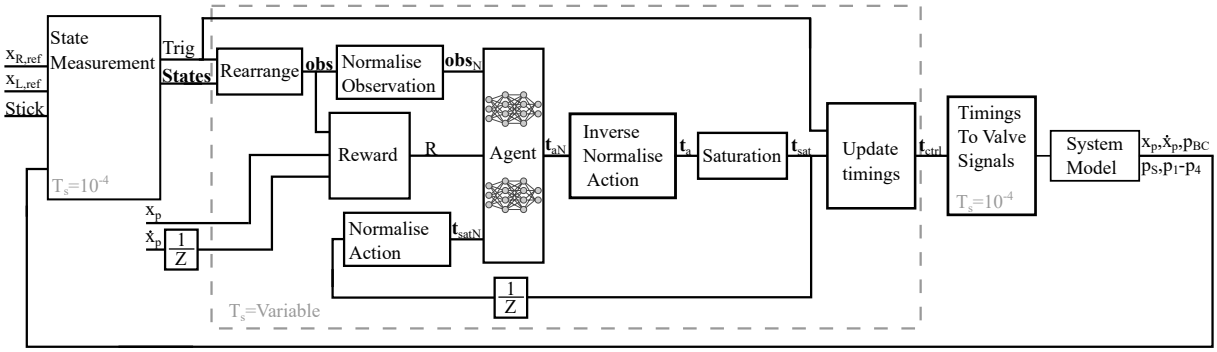


Figure 4.6. Block diagram of the setup for the training process.

State Measurement The states are measured with Algorithm 1. An initial state measurement is done if the piston is at standstill and the joystick value is non-zero. Otherwise, it measures the states at the endpoint of the piston oscillations, based on the change in sign of the piston velocity, if the joystick signal is non-zero. *Trig* is a variable which is set to -0.5 at the left end and 0.5 at the right end while keeping the value until the next end. This is illustrated in Figure 4.7.

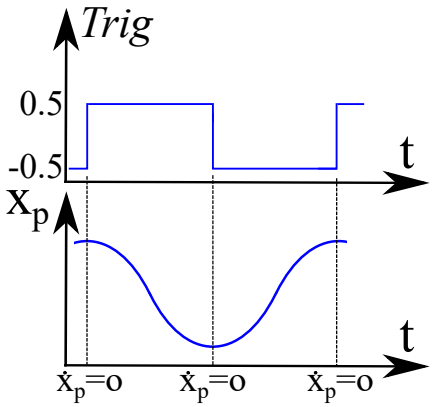


Figure 4.7. Illustration showing how the *Trig* variable varies with the position oscillation.

Algorithm 1: State Measurement

Determine the initial measurement side and movement direction at the first sample after stick is changed from 0

if $Stick \neq 0$ **then**
  if $Trig == 0.5$ **then**
    $t_{p1,L} = 0; \quad t_{p2,L} = 0$
    if $p_1 < p_T \ \& \ t_{p1,R} == 0 \ \& \ t_l > 0.005$ **then**
      $t_{p1,R} = t_l$
    end
    if $p_2 > p_S \ \& \ t_{p2,R} == 0 \ \& \ t_l > 0.005$ **then**
      $t_{p2,R} = t_l$
    end
  end
  else if $Trig == -0.5$ **then**
    $t_{p1,R} = 0; \quad t_{p2,R} = 0$
    if $p_1 > p_S \ \& \ t_{p1,L} == 0 \ \& \ t_l > 0.005$ **then**
      $t_{p1,L} = t_l$
    end
    if $p_2 < p_T \ \& \ t_{p2,L} == 0 \ \& \ t_l > 0.005$ **then**
      $t_{p2,L} = t_l$
    end
  end
  if $sign(\dot{x}_p(t)) > sign(\dot{x}_p(t-1)) \ || \ calc_{i,L} == 1$ **then**
    $\Delta x_{p,L} = x_{p,refL} - x_p; \quad \Delta p_{1,L} = p_S - p_1; \quad \Delta p_{1,L} = p_T - p_2$
    if $t_{p1,L} == 0$ **then**
      $\Delta t_{p1,L} = 0$
    else
      $\Delta t_{p1,L} = t_l - t_{p1,L}$
    end
    if $t_{p2,L} == 0$ **then**
      $\Delta t_{p2,L} = 0$
    else
      $\Delta t_{p2,L} = t_l - t_{p2,L}$
    end
    $Trig = -0.5$
  end
  else if $sign(\dot{x}_p(t)) < sign(\dot{x}_p(t-1)) \ || \ calc_{i,R} == 1$ **then**
    $\Delta x_{p,R} = x_{p,refR} - x_p; \quad \Delta p_{1,R} = p_T - p_1; \quad \Delta p_{1,R} = p_S - p_2$
    if $t_{p1,R} == 0$ **then**
      $\Delta t_{p1,R} = 0$
    else
      $\Delta t_{p1,R} = t_l - t_{p1,R}$
    end
    if $t_{p2,R} == 0$ **then**
      $\Delta t_{p2,R} = 0$
    else
      $\Delta t_{p2,R} = t_l - t_{p2,R}$
    end
    $Trig = 0.5$
  end
end

Rearrange During training, gravity is removed from the simulation model to achieve a symmetric response since this doubles the number of samples. To do this, the order of the samples in the observation vector has to be rearranged, which is done in Algorithm 2.

Algorithm 2: Rearrange

if *Trig == -0.5* **then**
  obs =
  $\begin{bmatrix}\Delta x_{p,R} & \Delta p_{1,R} & \Delta p_{2,R} & \Delta t_{p1,R} & \Delta t_{p2,R} & \Delta x_{p,L} & \Delta p_{1,L} & \Delta p_{2,L} & \Delta t_{p1,L} & \Delta t_{p2,L} & p_{BC} & p_S & Amp\end{bmatrix}$
else
  obs =
  $\begin{bmatrix}\Delta x_{p,L} & \Delta p_{2,L} & \Delta p_{1,L} & \Delta t_{p2,L} & \Delta t_{p1,L} & \Delta x_{p,R} & \Delta p_{2,R} & \Delta p_{1,R} & \Delta t_{p2,R} & \Delta t_{p1,R} & p_{BC} & p_S & Amp\end{bmatrix}$
end

Reward The reward function is the basis for the policy updates. This means that a good reward function is essential for the RL algorithm to converge to a desirable result. It is in the reward function that the different errors can be weighted. For this project, the reward function in Equation (4.1) is used. Here, a positive reward of 100 is given every time step with a negative reward for different errors. Here, a position error of 1 mm gives the same negative as a pressure error of 10 bar and a time error of 1 ms. Furthermore, a negative reward is given if the piston hits the end position of 0 and 0.035, respectively. The penalty for hitting the end is dependent on the speed just before impact to guide the learning process.

$$R = \begin{cases} 100 - \left(\frac{\Delta x_{p,L}}{0.001}\right)^2 - \left(\frac{\Delta p_{1,L}}{10\cdot 10^5}\right)^2 - \left(\frac{\Delta p_{2,L}}{10\cdot 10^5}\right)^2 - \left(\frac{\Delta t_{p1,L}}{0.001}\right)^2 - \left(\frac{\Delta t_{p2,L}}{0.001}\right)^2 & \\ -(x_p \leq 0)\cdot(15 + 100\cdot|\dot{x}_p(t-1)|)\,, & \text{if } Trig = -0.5 \\ & \\ 100 - \left(\frac{\Delta x_{p,R}}{0.001}\right)^2 - \left(\frac{\Delta p_{1,R}}{10\cdot 10^5}\right)^2 - \left(\frac{\Delta p_{2,R}}{10\cdot 10^5}\right)^2 - \left(\frac{\Delta t_{p1,R}}{0.001}\right)^2 - \left(\frac{\Delta t_{p2,R}}{0.001}\right)^2 & \\ -(x_p \geq 0.035)\cdot(15 + 100\cdot|\dot{x}_p(t-1)|)\,, & \text{else} \end{cases} \tag{4.1}$$

Saturation The Actions are saturated with Algorithm 3 which saturates $\mathbf{t}_a(1:2)$ between 0 ms and 20 ms whereas $\mathbf{t}_a(3)$ is saturated between −10 ms and 10 ms. To ensure that the piston always oscillates during training, a combined minimum is set for $\mathbf{t}_{sat}(1)$ and $\mathbf{t}_{sat}(2)$, which accelerates the piston, to 0.1 ms.

Algorithm 3: Saturation

$\mathbf{t}_{sat} = max\left(\, min\left(\mathbf{t}_a,\ \begin{bmatrix}0.02 & 0.02 & 0.01\end{bmatrix}\right),\ \begin{bmatrix}0.0 & 0.0 & -0.01\end{bmatrix}\right)$
if $\boldsymbol{t}_{sat}(1) + \boldsymbol{t}_{sat}(2) < 0.0001$ **then**
  $\mathbf{t}_{sat}(1) = 0.0001$
end

Update Timings The saturated timings are given to Algorithm 4, which updates the vector t_{ctrl}, which contains t_1 to t_6. At the left end, entrances 1 to 3 are updated, and at the right end, entrances 4 to 6 are updated. This can be seen in Algorithm 4

Algorithm 4: Update Timings

if *First Execution* **then**
  Define $\mathbf{t}_{ctrl}$ as a persistent variable
  $\mathbf{t}_{ctrl} = zeros(6, 1)$
end
if $Trig == -0.5$ **then**
  $\mathbf{t}_{ctrl}(1:3) = \mathbf{t}_{sat}$
else if $Trig == 0.5$ **then**
  $\mathbf{t}_{ctrl}(4:6) = \mathbf{t}_{sat}$
end

Timings To Valve Signals The timings are converted to valve signals with Algorithm 5 according to Figure 4.1. $x_{v,ref}$ is a vector which contains the reference value for the 8 On/Off valves, where 1 is open and 0 is closed. The reference values are sent to the system model described in Chapter 3.

Algorithm 5: Determination of valve signals based on defined timings and measured time.

if $Trig = -0.5$ **then**
  if $t_l \leq t_1$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$
  else if $t_l > t_1 \ \& \ t_l \leq (t_1 + t_2)$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$
  else if $t_l > (t_1 + t_2) \ \& \ t_l \leq (t_1 + t_2 + |t_3|) \ \& \ t_3 < 0$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
  else if $t_l > (t_1 + t_2) \ \& \ t_l \leq (t_1 + t_2 + |t_3|) \ \& \ t_3 > 0$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
  else
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
  end
else if $Trig = 0.5$ **then**
  if $t_l \leq t_4$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \end{bmatrix}$
  else if $t_l > t_4 \ \& \ t_l \leq (t_4 + t_5)$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$
  else if $t_l > (t_4 + t_5) \ \& \ t_l \leq (t_4 + t_5 + |t_6|) \ \& \ t_6 < 0$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$
  else if $t_l > (t_4 + t_5) \ \& \ t_l \leq (t_4 + t_5 + |t_6|) \ \& \ t_6 > 0$ **then**
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$
  else
    $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
  end
else
  $\mathbf{x}_{v,ref} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$
end

4.3.1 Agent

For this project, a Twin-Delayed Deep Deterministic (TD3) policy gradient agent has been chosen, which is an improvement of the classic deep deterministic policy gradient (DDPG) agent. (MathWorks 2025d) TD3 has a continuous deterministic policy, which means that the same observations will result in the same actions as for a classical controller. This is appropriate for physical systems since they are deterministic in nature. The observations can be both discrete and continuous, but in this case, all the observations are continuous.
TD3 is an Actor-Critic method where the critic is used to guide the actors' learning by giving an estimate of the expected long-term reward called the Q-value. To estimate the Q-value, the critic gets the current observation and action as input and outputs a Q-value, which is compared to the reward received from the environment after the action has been applied, which is illustrated in Figure 4.8. The Critic is updated to minimise this difference with a least squares method. The Actor is updated by gradient ascent, where the gradient of the critic output with regards to the actions and the gradient of the actor with regards to the actor parameters are used. (MathWorks 2025d) The specific details with regard to when the learning takes place are described further down.

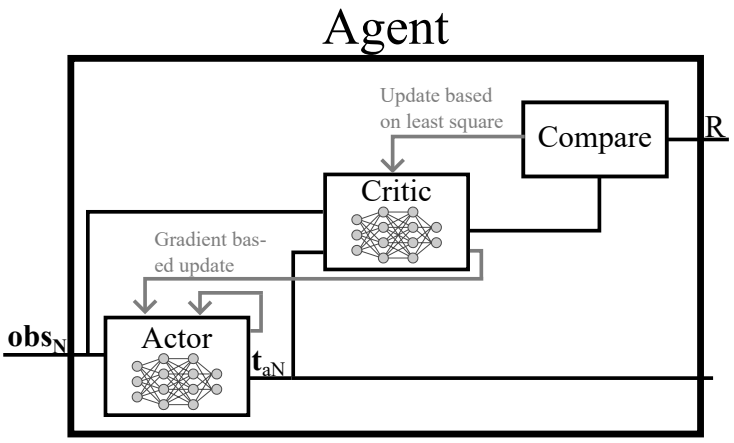


Figure 4.8. Block diagram illustrating the learning process for the actor and critic within the agent.

For this agent, there are multiple choices to take with regard to Actor, Critic, Optimiser Options and Exploration vs Exploitation.

Actor

A deep Neural Network is used to represent the policy since they are great at representing the complicated non-linear relations between the Observations and the desired Actions. More information on Neural Networks can be found in Appendix B. It is desired to have as small a Neural Network as possible while still being able to create desirable control signals. This leads to a network with 3 hidden layers, with each hidden layer having a width of 64. This was found through trial and error.
For the actor, a sigmoid action function has been used, which gives an output between 0 and 1. This is desirable since it can easily be scaled to the intervals of the control timings.

Critic

The TD3 agent uses 2 Q-value functions, which in this case are both represented with deep neural networks since they are great at representing the complicated non-linear relations between action, state and cumulative rewards. Each Q-value function is represented with a Neural Network with 4 hidden

layers, each with a width of 64. The layers are fully connected with the RELU activation function being utilised.

Training Process

With the overall update of the actor and critic described previously, the full training process can be described. All the agents in this project have been trained using experience-based parallelisation. This means that the different workers simulate the environment with an agent to create experiences, which are stored in a combined experience buffer. The experience buffer is used to update the parameters of the actor and the critic.
The training process is described in Algorithm 6 (MathWorks 2025d). To improve reproducibility, the same seed is used to create the "random" numbers that initialise the actor and critic every time a new training process is started. The TD3 agent consists of the standard actor and critic, along with the target actor and target critic. The target actor and critic are used to improve the stability of the optimisation process by updating at a slower rate.

Algorithm 6: Training algorithm

Initialise the two critics, $Q_k(\mathbf{obs}_N,\ \mathbf{t}_{aN};\ \phi_k)$, with random parameters, ϕ_k and the corresponding target critic $Q_{tk}(\mathbf{obs}_N,\ \mathbf{t}_{aN};\ \phi_{tk}$ with random parameters ϕ_{tk}

Initialise actor,$\pi(\mathbf{obs}_N;\ \theta)$, and target actor $\pi_t(\mathbf{obs}_N;\ \theta_t)$ with the same "random" parameters θ and θ_t

while *Training == 1* **do**

  while *StopEpisode* == 0 **do**

    Calculate the Action based on the current observation and policy. $\mathbf{t}_{aN,i} = \pi(\mathbf{obs}_{N,i};\theta) + \mathbf{N}_g$ where $\mathbf{N}_g$ is the noise from the exploration model.

    Execute the Action in the environment and measure the next observation, $obs_i^{'}$, and the corresponding reward.

    Add the experience $\left(\mathbf{obs}_{N,i}, \mathbf{t}_{aN,i}(\ \mathbf{obs}_{N,i}\ ), R_i, \mathbf{obs}_{N,i}^{'}\right)$ to the experience buffer.

    if $x_p \geq 0.035\ \ ||\ \ x_p \leq 0.0\ \ ||\ \ t \geq t_{end}$ **then**

      $StopEpisode = 1$

    end

  end

  if $n_{experience} \geq M$ **then**

    Sample a random mini batch of M experiences from the experience buffer

    Calculate the value function target, y_i, from the current reward and the discounted reward from the next step based on the critic.

$$y_i = R_i + \gamma \cdot \min_k \left(Q_{tk}\left(\mathbf{obs}_{N,i}^{'},\ \pi_t\left(\mathbf{obs}_{N,i}^{'};\theta_t\right)+\varepsilon;\phi_{tk}\right)\right)$$

    Update the critic by minimising the loss function

$$L_k = \frac{1}{2M}\sum_{i=1}^{M}\left(y_i - Q_k\left(\mathbf{obs}_{N,i},\ \mathbf{t}_{aN,i};\ \phi_k\right)\right)^2$$

    Update the policy every other episode using gradient ascent to maximise the discounted cumulative long-term reward.

    Update the target actor and critic every other time the critic is updated with periodic smoothing

    $\phi_{tk} = \tau\phi_k + (1-\tau)\phi_{tk}$

    $\theta_t = \tau\theta + (1-\tau)\theta_t$

  end

  if $n_{episodes} \geq max_{episodes}$ **then**

    Training = 0

  end

end

Training and Test of RL-Agent 5

With the setup completed, a policy can be trained with the RL agent before it is implemented in it simulation model and laboratory for testing.

5.1 Training of Policy

To test the possibility of RL being able to control the FBoT, a policy has to be trained. Before the training process is started, multiple hyperparameters have to be set. During this project, many different combinations of hyperparameters have been investigated. The combination which yielded the best results can be seen in Table 5.1. The large mini-batch relative to the size of the experience buffer has been chosen since the generation of data was slow relative to the time it takes to update the neural networks.

Parameter	Learn rate	Gradient threshold	Mini-batch	Experience buffer	Number of epochs	Standard Deviation (STD)	Min STD	STD decay	t_{end}
Value	10^{-5}	1	10000	10000	2	0.1	0	0	0.5

Table 5.1. Hyperparameters which gave the best performing policies.

For the policy to work at different manifold pressures, it has to be trained with this in mind. To do this, p_S was set as a uniformly distributed random value between 40 bar and 80 bar at the start of each episode whereas p_{BC} was set between 20 bar and 40 bar

The episode reward and average episode reward, over 200 episodes, from the training with these hyperparameters can be seen in Figure 5.1. Here, the mean reward is fairly constant until after 10000 episodes. This is because the experience buffer needs to be filled with 10000 experiences before the training begins, since that is the size of the mini-batches used during training.

After the initial 10000 episodes, the training starts, where multiple spikes are observed in the mean reward, but all below the threshold of 200, which was found to deliver agents that could operate continuously in a simulation without gravity based on the current reward function. This leads to a maximum mean reward just below 400 after 34400 episodes, whereafter the average reward is reduced before it starts rising again.

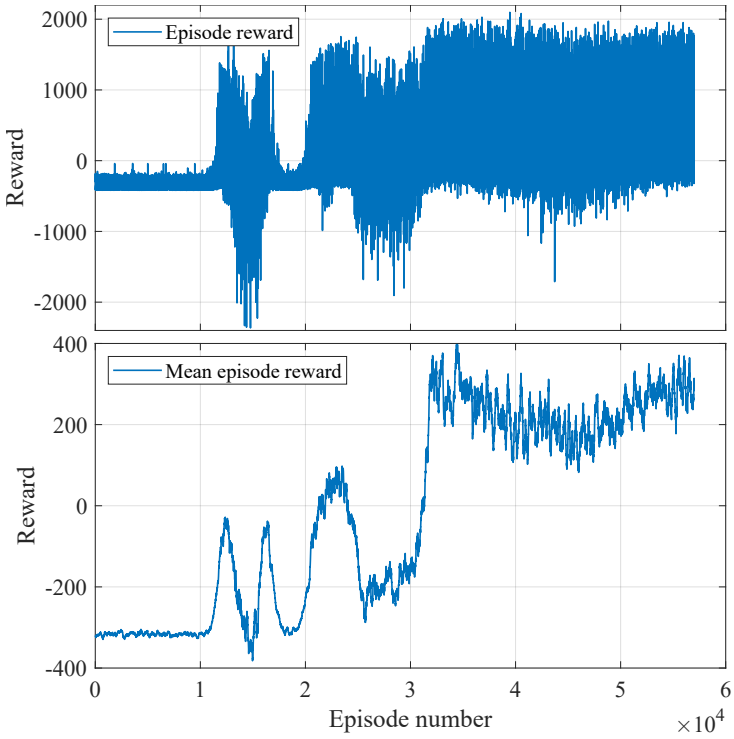


Figure 5.1. The episode reward and mean reward from the training of the agent.

5.2 Test of RL-Controller

With the training completed, the trained policy can be tested, which is done by replacing the agent with the policy. Since the policy only needs the normalised observation as an input, the block diagram can be simplified, which can be seen in Figure 5.2.

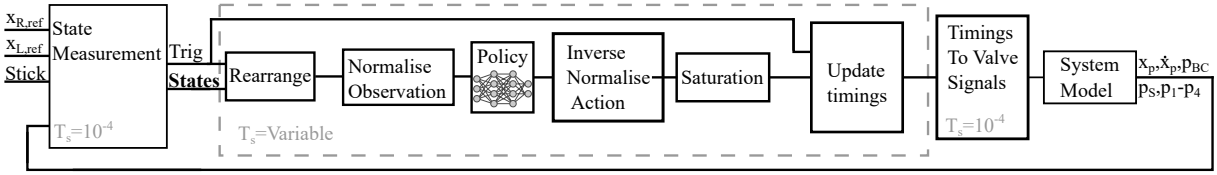


Figure 5.2. Block diagram for testing/implementing the policy.

The policy from the agents after 34400, 50000 and 57000 episodes is compared to each other to see how the policy changes through the training and if the mean reward represents the best performing agent. Firstly they are compared at $p_S = 50\,\text{bar}$ and $p_{BC} = 30\,\text{bar}$ which is shown in Figure 5.3. Here, it can be seen that all the policies can oscillate the piston nicely, but with some steady-state error. The only policy that somewhat corrects this over time is the one after 50000 episodes, which increases t_2 and t_5 through the test. It however has the largest error for $\Delta p_{2,R}$ and $\Delta p_{1,L}$ which might have to do with the significantly lower t_3 and t_6 which means that state 4.1 and 10.1 from Figure 4.1 is held significantly longer.

Interestingly the policies keep t_1 and t_4 at 0 all the time even though an increase in t_1 and corresponding decrease in t_2 will likely lead to lower pressure errors due to an increased amount of kinetic energy when the chambers are isolated in state 5.

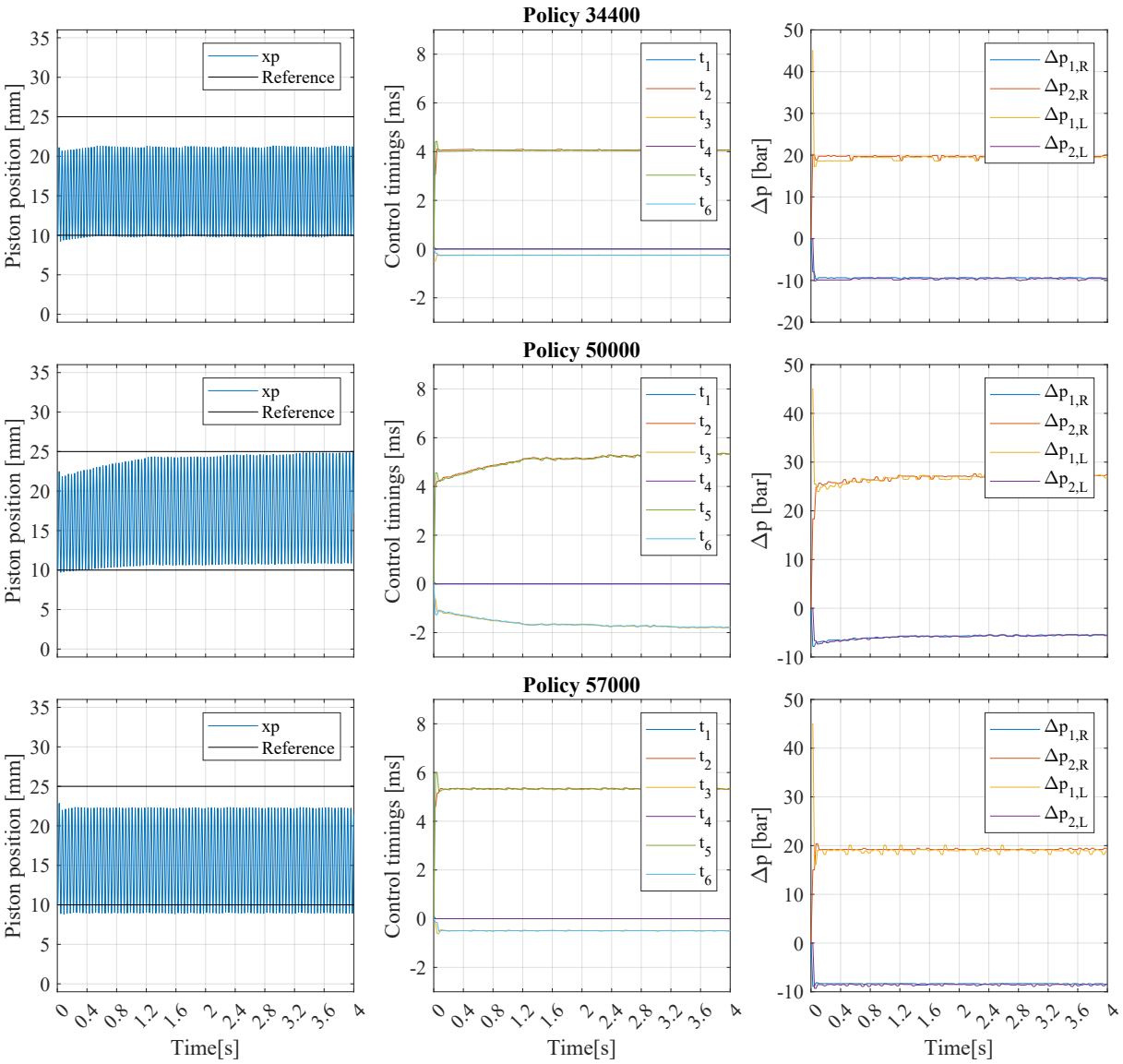


Figure 5.3. Comparison of the trained policies after 34400, 50000 and 57000 episodes with $p_S = 50\,\text{bar}$ and $p_{BC} = 30\,\text{bar}$.

Throughout the training, p_s and p_{BC} have been varied so that the policies should work at different manifold pressures. This is investigated in Figure 5.4 where $p_S = 70\,\text{bar}$ and $p_{BC} = 30\,\text{bar}$. Interestingly, it can be seen that the policy after 34400 episodes gives approximately the same control timings here, which suggests that the control timings are minimally affected by the manifold pressures, which are some of the inputs. The policy after 50000 episodes again slowly corrects its amplitude, which suggests that it might have learned how to correct its control timings based on the position error $\Delta x_{p,R}$ and $\Delta x_{p,L}$. Looking at the pressure error $\Delta p_{1,L}$ and $\Delta p_{2,R}$ is significantly reduced here compared to the simulation with $p_S = 50\,\text{bar}$ this is likely due to the increased ratio between p_S and p_{BC} which means that the kinetic energy should be higher when going into states 5 and 11 which leads to an increased chamber compression/decompression.

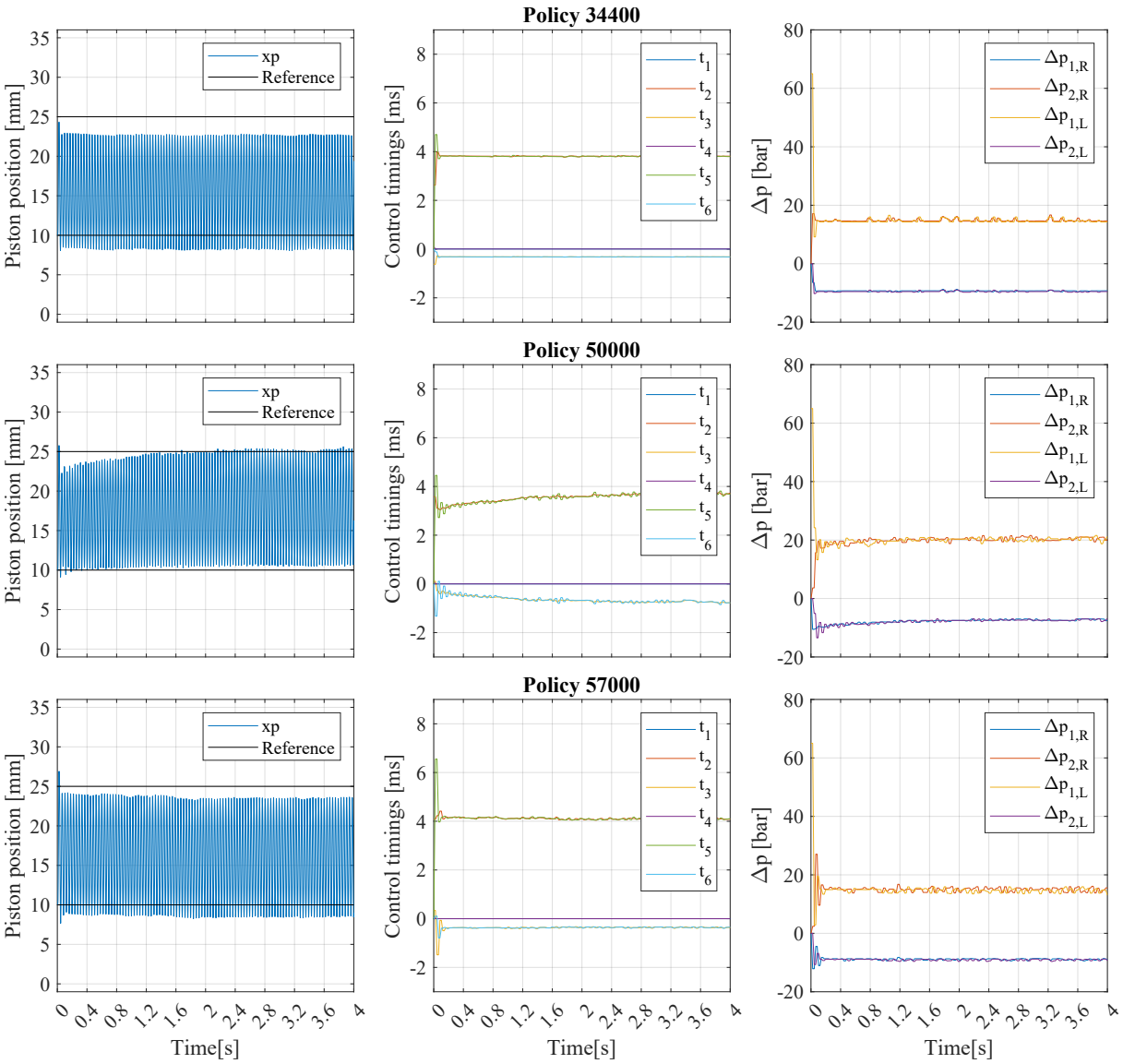


Figure 5.4. Comparison of the trained policies after 34400, 50000 and 57000 episodes with $p_S = 70\,\text{bar}$ and $p_{BC} = 30\,\text{bar}$.

To see how the policies handle a variation in oscillation amplitude, the amplitude is increased throughout the simulation. Here, it can be seen that none of the policies change the control timings to keep the oscillation centred based on the position error $\Delta x_{p,R}$ and $\Delta x_{p,L}$. Both policies 50000 and 57000 seem to have learned to increase t_2 and t_5 as the amplitude increases, which is unexpected since the amplitude was kept constant during training. This does however not lead to an oscillation that could be continued since they are approaching the maximum position of 35 mm quickly.

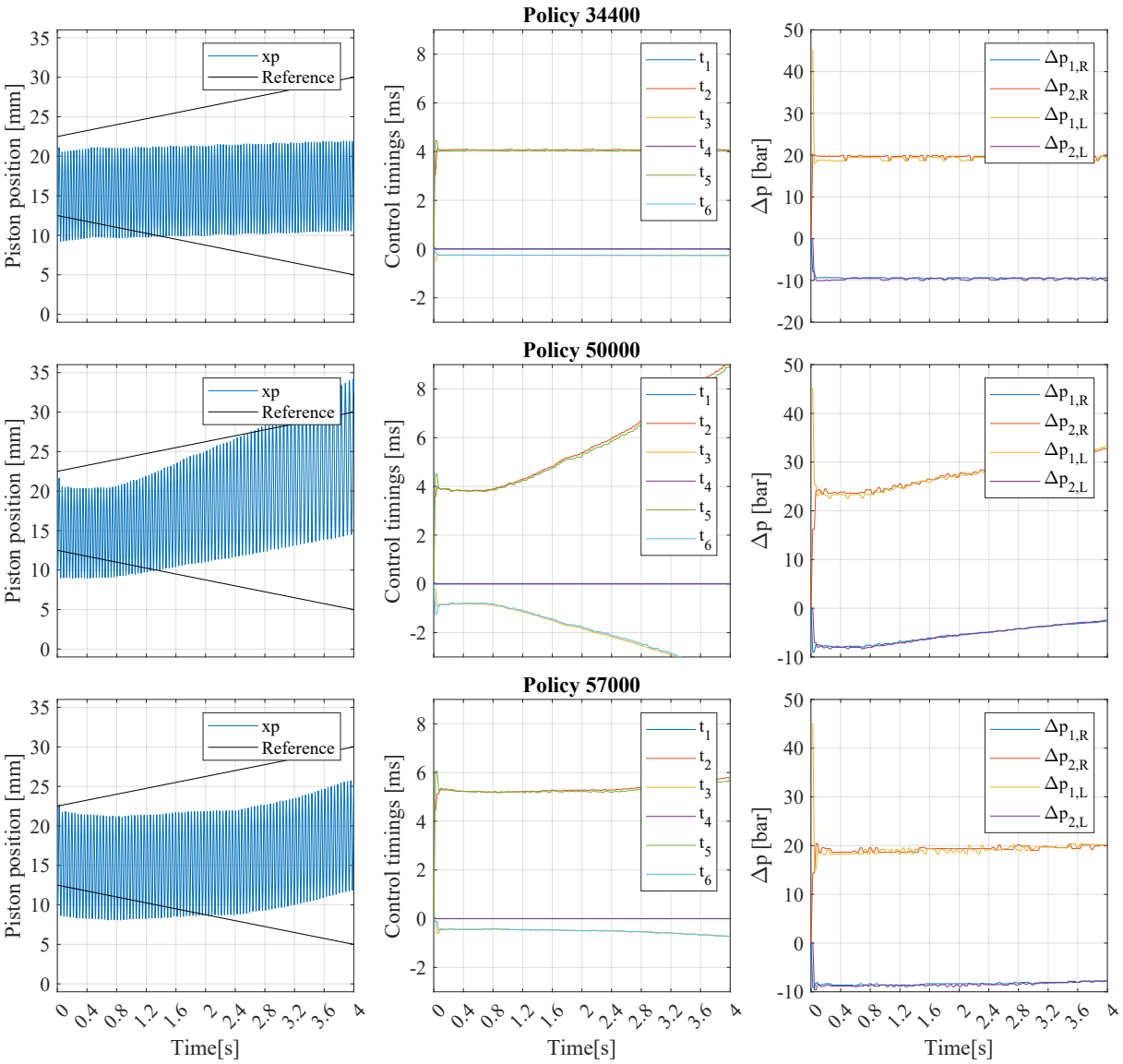


Figure 5.5. Comparison of the trained policies after 34400, 50000 and 57000 episodes with varying amplitude. p_S =50 bar and p_{BC} =30 bar

5.2.1 Efficiency Test

The goal of the FBoT is to increase the efficiency of hydraulic systems. The efficiency has been calculated as a moving average of the ratio between the power in and out of the FBoT:

$$\eta = \frac{P_{out}}{P_{in}} = \frac{-(Q_{3L} + Q_{4L}) \cdot p_{BC}}{(Q_{1S} + Q_{2S}) \cdot p_S} \tag{5.1}$$

The efficiency is seen to increase as the oscillation amplitude increases in Figure 5.6 where the efficiency ends at around 50%. Furthermore, it can be seen that the losses primarily come from the On/Off valves. Beyond the two losses shown, there is a loss in the check valves and through leakage within the FBoT.

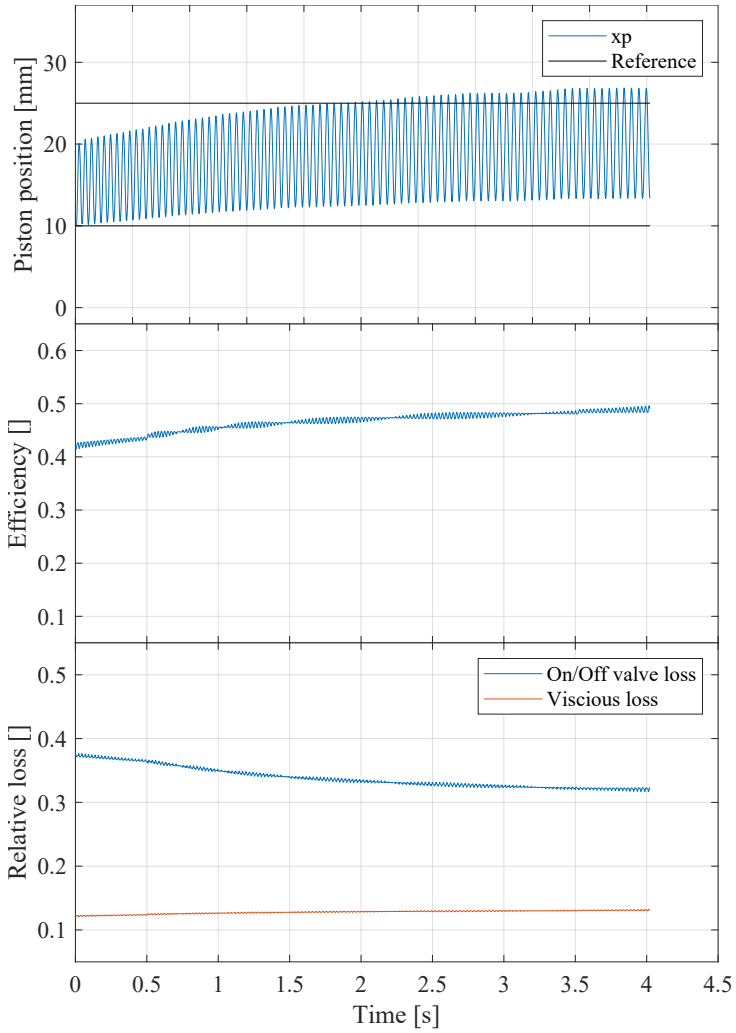


Figure 5.6. Plot showing that the efficiency increases as the amplitude increases. The plot is created with the policy from episode 50000 while p_S =40 bar and p_{BC} =30 bar.

To investigate where in the transformation the losses happen, a single oscillation is investigated. This is done by an animation which can be found at `https://drive.google.com/file/d/1gv-nNIcI5lppQFCtHojNH6ldgsdXyJ7/view?usp=sharing` with one of its frames shown in Figure 5.7.
Here, it can be seen that there is a delay from the endpoint until the valves start opening, due to the delay in the valve opening response. This results in a large spike in power loss since there is a large pressure difference when the valves are opened, combined with a large flow. The other significant point of energy loss is when the valves are closed, where the power loss happens over a longer period since the flow continues as the valve closes with an increasing pressure difference.
The energy loss at the start of the oscillation could be reduced significantly if there was minimal pressure difference as the valve was opened. To do this, the FBoT has to be controlled so that state 6/12 from Figure 4.1 is reached. Furthermore, it has to be anticipated when the endpoint is reached so that the valves are opened at that exact moment. To do this properly, you have to accurately determine the time until the end point so that the valve signals can be sent ≈ 0.35 ms before. The energy loss in the middle of the oscillation, when the valves are closing, cannot be reduced by smarter controls without neglecting the pressure buildup/reduction phase. A way to reduce these losses is to get faster operating valves.

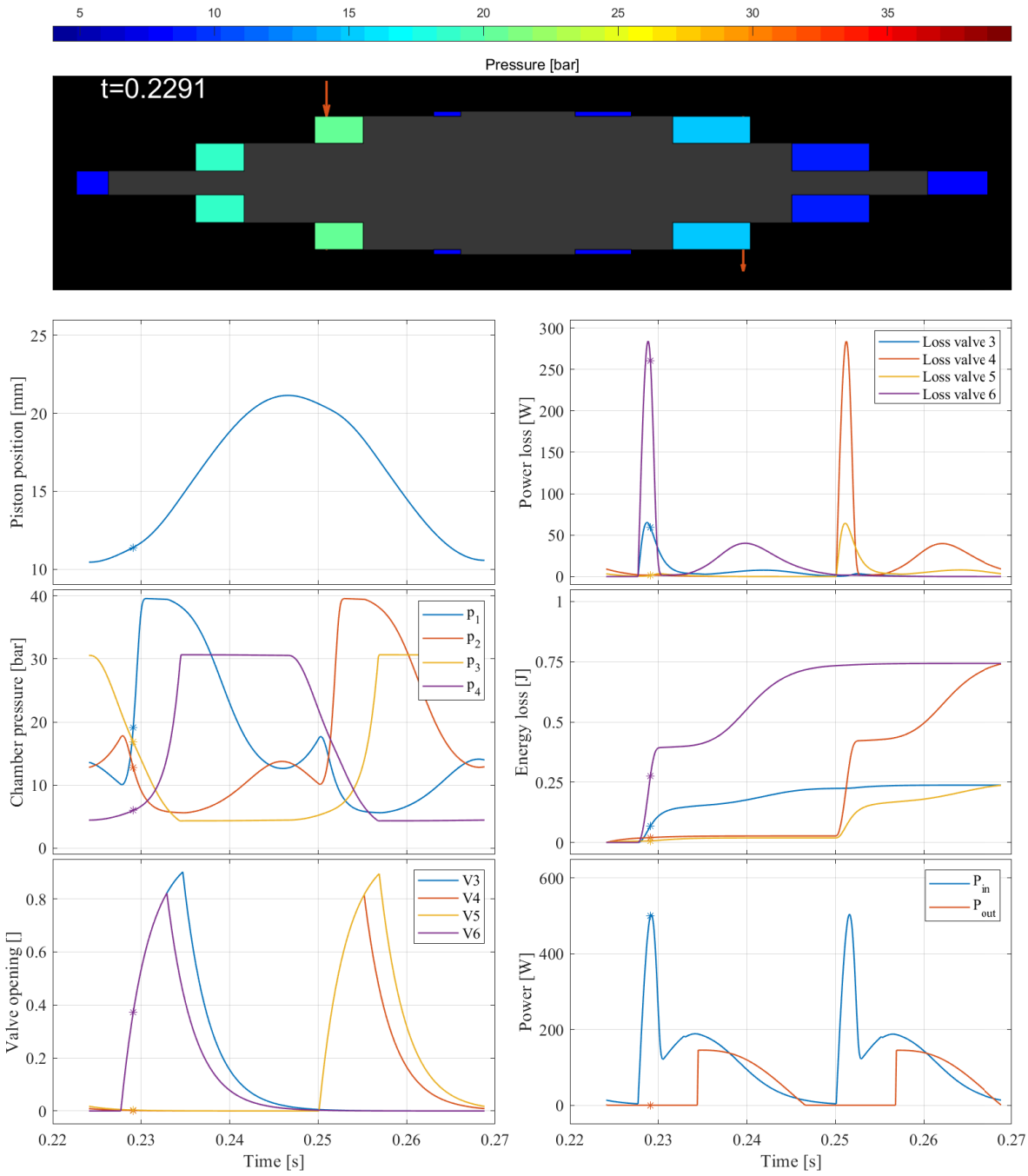


Figure 5.7. Investigation of energy loss during one oscillation. An animation of this plot is available at `https://drive.google.com/file/d/1gv-nNIcI5lppQFCtHojNH6ldgsdXyJ7y/view?usp=sharing`. Note that the definition of valve numbers can be found on Figure 3.1.

5.3 Laboratory Test

With the somewhat successful test in the simulation model, the policy from episode 50000 is deployed in the laboratory setup. This gives the results shown in Figure 5.8. Here, it can be seen that the oscillation amplitude varies significantly from oscillation to oscillation. This corresponds with significant variations, especially in t_5 and t_6. It is suspected that the change in control timings is due to the variations in p_{BC} or $\Delta p_{1,L}$ since the position errors, $\Delta x_{p,R}$ and $\Delta x_{p,L}$, have shown no effect on the control timings in the simulations. Note that it is $p_{BC,a}$ that is given to the policy.

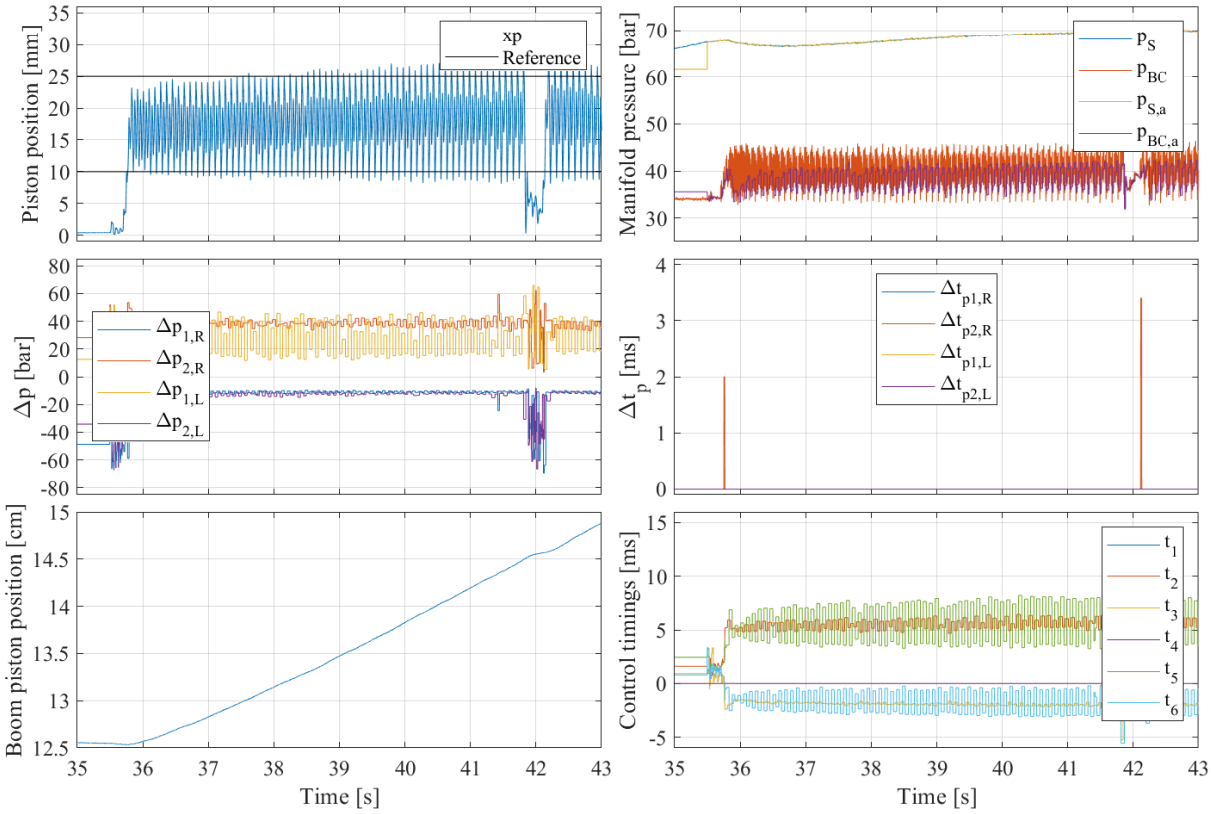


Figure 5.8. Long timeseries for policy 50000 executed with the laboratory setup.

During the test around 42 s, the oscillation suddenly varies to very small oscillations away from the centre. To investigate this, a shorter timeseries is shown in Figure 5.9. Here it is observed that both t_2 and t_5 are updated very close together around 41.81 s, which suggests that the policy has been called in quick succession due to either noise in the position measurement or an actual physical oscillation which affects the speed signal that is determined based on the position.

The bump in the position measurement shortly after the end points happens when the chamber pressures are about to equalise on both sides of the piston before the valves are opened due to their transient response. This is also evident in the p_1 and p_2 measurements, which have a significant bump at the ends of the oscillations where they are supposed to stay at tank pressure. This is due to the mistiming of the opening of the On/Off valves for the next state. This is not the case for p_3 and p_4 since they are only affected by the check valves, which are pressure-actuated.

This can also be done for chambers 1 and 2 with a controllable check valve similar to what Danfoss does with their digital displacement pumps.

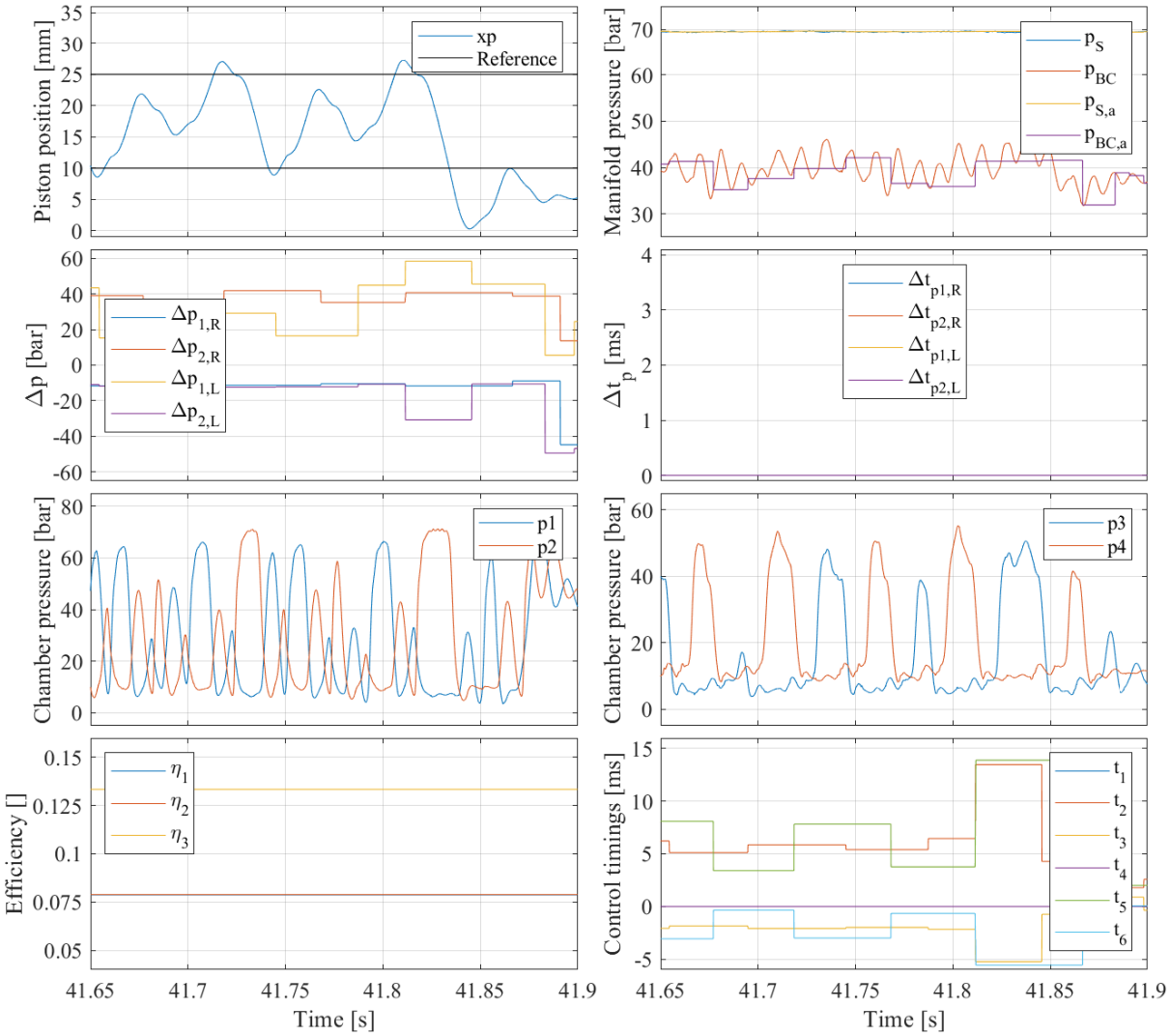


Figure 5.9. Short timeseries for policy 50000 executed with the laboratory setup.

5.3.1 Efficiency Test

To determine the efficiency of the FBoT, the power in and out of the FBoT is needed. The pressure is measured at both the source and the boom cylinder, but no flow measurements exist. The flow out of the FBoT is determined from the volume expansion of the boom cylinder chamber, assuming constant pressure.

$$Q_{out} = \dot{x}_{BP} \cdot A_{BP} \tag{5.2}$$

The flow into the FBoT can be determined in multiple ways which each comes with its own assumptions. Firstly, it can be based on the flow from the pump. This includes the leakage outside the FBoT and the flow needed to produce potential pressure changes in the supply chamber and accumulator. Here, a volumetric efficiency, η_{vP}, of 90% is assumed, which was found from a similar pump.

$$Q_{in1} = \omega_p \cdot D_p \cdot \eta_{vP} \tag{5.3}$$

Another approach is to base the flow into the FBoT on the On/Off valve models created in Chapter 3 where the flow is dependent on the opening of the valve, which is modelled as a first-order system, and the pressure difference. This model is a simplification of the actual opening, which is significantly

dependent on pressure difference, flow, fluid viscosity, and more.

$$Q_{in2} = x_{v,ref}(6) \cdot e^{-T_d \cdot s} \cdot \frac{1}{1 + \tau_v \cdot s} \cdot k_{v,oo} \cdot \sqrt{\frac{2}{\rho} |p_S - p_1|} \cdot sign(p_S - p_1) + \tag{5.4}$$

$$x_{v,ref}(4) \cdot e^{-T_d \cdot s} \cdot \frac{1}{1 + \tau_v \cdot s} \cdot k_{v,oo} \cdot \sqrt{\frac{2}{\rho} |p_S - p_2|} \cdot sign(p_S - p_2) \tag{5.5}$$

The third approach is to base the flow into the FBoT on the continuity equation for chambers 1 and 2. Here, only the positive flow is counted since the flow is out of the FBoT when it is negative. This ignores the leakage flow inside the FBoT and assumes that the pressure gradient can be accurately estimated based on the pressure measurements.

$$Q_{in3} = max\left(\dot{x}_p \cdot A_2 + \frac{V_{01} + x_p \cdot A_2}{\beta eff(p_1)} \cdot \frac{\Delta p_1}{\Delta t} - \dot{x}_p \cdot A_2 + \frac{V_{02} - x_p \cdot A_2}{\beta eff(p_2)} \cdot \frac{\Delta p_2}{\Delta t}, 0\right) \tag{5.6}$$

The efficiency can be calculated for the 3 different approaches with Equation (5.7).

$$\eta = \frac{Q_{out} \cdot p_{BC}}{Q_{in} \cdot p_S} \tag{5.7}$$

The efficiency from the three different methods is shown in Figure 5.10, where a moving average over 1 second is used. Here, it can be seen that the efficiency is between 7% and 12.5% for the given transformation with η_1 and η_2 showing the lowest efficiency while η_3 shows the highest. η_1 is expected to be the lowest since it gets the input power from the pump, thereby ignoring losses before the FBoT, whereas η_3 ignores the leakage losses within the FBoT, which means it should have the highest efficiency.

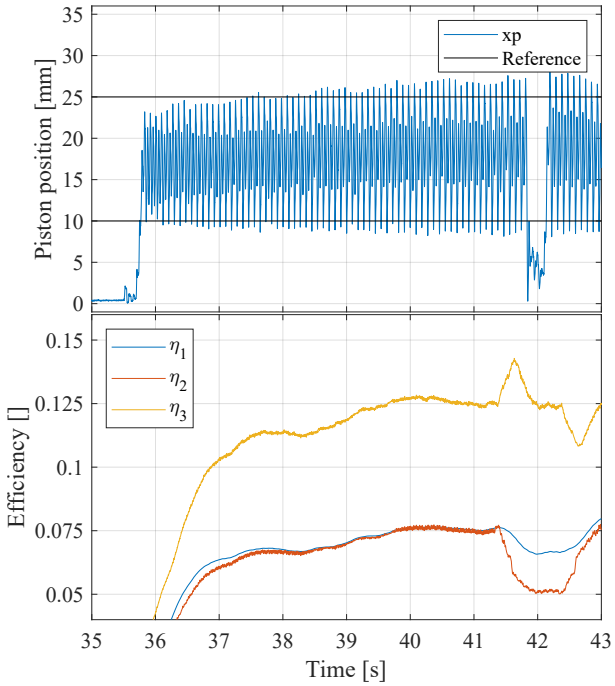


Figure 5.10. Comparison of the different efficiency estimates for this specific oscillation in the laboratory.

Discussion 6

Through this project, some things have been learned which warrant further discussion for the development of future control strategies for the FBoT, whether they are based on RL or some other method

6.1 Learnings From The Training of Agents For FBoT Control

Throughout the course of the project, numerous agents were trained, resulting in valuable insights and a deeper understanding of the learning dynamics. Some of these insights are described below.

Reproducibility One of the primary challenges has been reproducibility since it is difficult to iterate and improve something that cannot be reproduced, since there is no way to tell if an improvement is from better hyperparameters or a "random" coincidence. It was found that it was easier to reproduce the results when the agents were trained with an exploration policy with a constant standard deviation. This is since small variations in the training could give some good episodes, which lowers the standard deviation of the exploration policy away from the other training results. Furthermore, an increased mini-batch size was observed to improve the reproducibility.

Convergence Challenges When training agents, one has to determine when it has converged based on the episode reward. However, through many of the trainings, significant oscillations were seen in the reward. Increasing the mini-batch size was seen to reduce this since more experiences were used, which gave smoother convergence. This could also be achieved by reducing the standard deviation in the exploration model, but that comes with its own challenges.

Episode Length The length of the episodes has been limited by the variable t_{end}, which is the maximum time for each episode. Long episodes often lead to policies that only worked at specific manifold pressure combinations, since a long episode could give over 50 experiences, while the other manifold pressure combinations gave 1-5 experiences. This led to this specific pressure combination dominating the experience buffer. Too short episodes meant that it did not observe enough of the observation space, the actions could lead to over time.

Error Correction At the start of the project, it was expected that the agents could learn to correct the timings based on the error measurement from the previous oscillation to that side. However, through the project, it became clear that the agent learned the timings based on the manifold pressures, which meant that it behaved similarly to open-loop control. With a deeper understanding of the training process, this observation becomes intuitive: in a disturbance-free environment, the most effective strategy for maximising the reward, based on errors observed on both sides, is to provide optimal parameters from the outset in an open-loop manner.

6.2 Control Signal Transformation

Throughout this project, the system has been treated as coupled and non-linear as determined in Section 4.1.2. They could potentially be reduced by changing the control and error signals.
Instead of controlling t_1 and t_2, the ratio $r = \frac{t_1}{t_1+t_2}$ could be controlled based on a normalised average pressure vs. normalised time error for chambers 1 and 2. A pressure error should increase the ratio since more kinetic energy is needed to compress/decompress the chambers before standstill, while still outputting power to the load.
To get the actual values for t_1 and t_2 a scalar, M_t, is used so that $t_1 = M_t \cdot r$ and $t_2 = M_t \cdot (1 - r)$. M_t should be determined based on the position error, where a positive error should lead to an increased M_t since there was not enough input power.
t_3 should be based solely on which chamber is closest to reaching the desired pressure at the correct time. This could be determined by the difference in pressure error and time error for the two chambers.
This setup is sketched with the block diagram in Figure 6.1 for calculating t_1 to t_3 while at the left end. Here, it is illustrated with an agent trained to give the initial parameters, but these could potentially be constant or determined by simple linear relations.

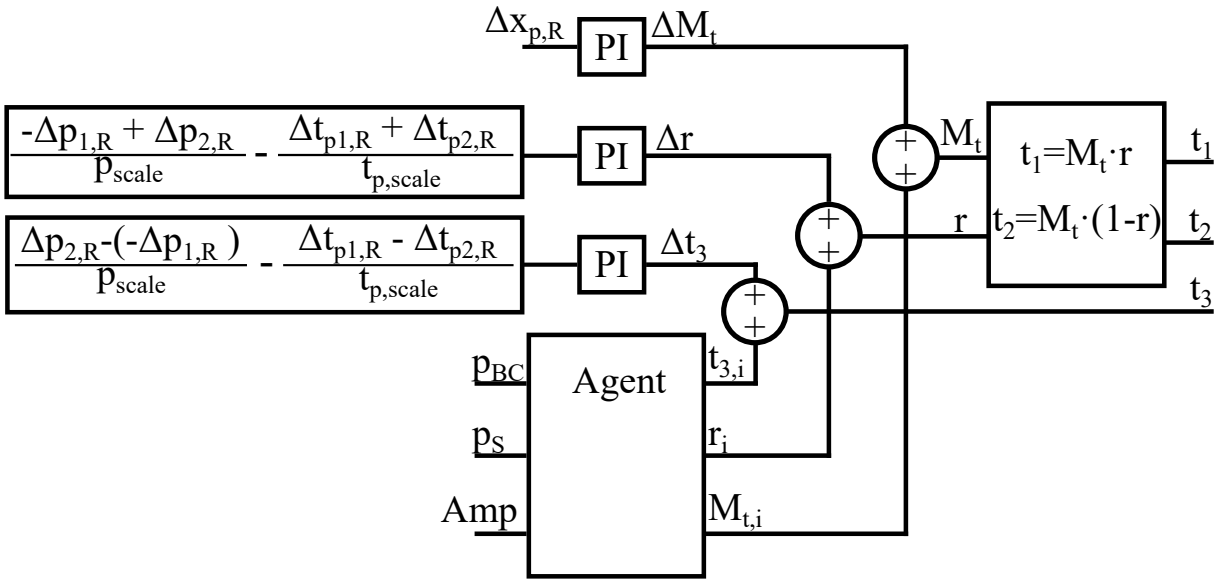


Figure 6.1. A block diagram for a potential new control structure based on the learnings from this project.

In the block diagram, PI-controllers are proposed to correct the control signals based on the errors in the previous oscillation. One potential problem with this is that the initial condition for each oscillation varies. This means that the previous control signal might give a different result this oscillation. This could potentially lead to an overall oscillatory behaviour as was observed by (Mortensen et al. 2025). The varying initial condition is sketched for the position in Figure 6.2. One way to overcome this could be to have different integrator gains for each side so that one side settles first, leading to a less oscillatory response.

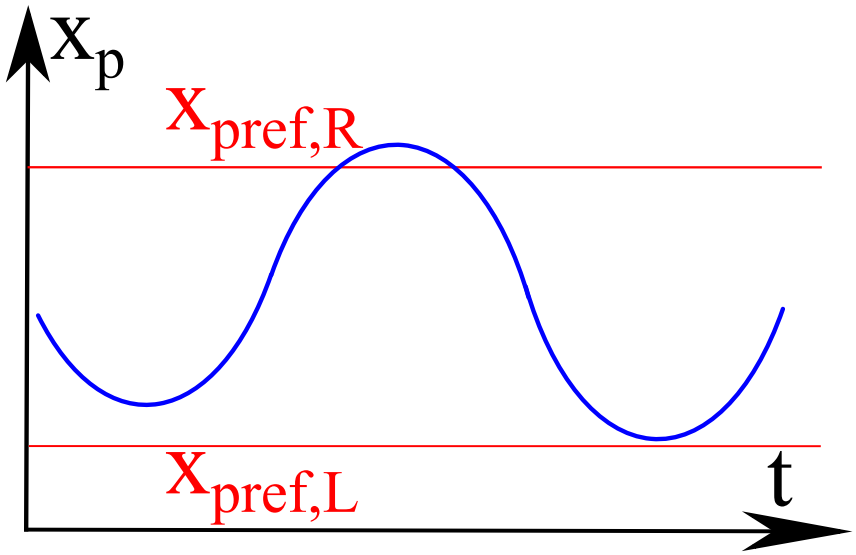


Figure 6.2. Case of when the variation in left position might fix the right end position without any changes in control timings for the movement towards the right.

Conclusion 7

This thesis investigated efficient control strategies for the Full Bridge Oscillation Transformer (FBoT), a novel hydraulic transformer developed at Aalborg University. To minimise energy losses associated with valve switching, a 12-step control sequence was proposed for pump-mode operation to simplify the control task. Given the nonlinear and coupled nature of the FBoT dynamics, reinforcement learning (RL) was employed to develop a control policy.

The RL-based controller was trained in simulation using a reward function designed to minimise valve losses while maintaining a desired oscillation amplitude. However, simulation results revealed that high-reward policies often failed to adjust control timings based on feedback errors. Instead, these policies favoured delivering near-optimal open-loop timings from the outset. While this approach maximised the reward function, it led to significant steady-state errors and inefficient valve actuation, ultimately reducing the system's overall efficiency.

Experimental validation confirmed that the trained policy could induce piston oscillation, but with considerable variability in amplitude and pressure errors. The efficiency, estimated using three different flow approximation methods, remained low, likely due to substantial valve loss due to the controller's inability to minimise the pressure differences as the valves were opened or closed.

Bibliography

Andersen, Benjamin Holm et al. (2023). *Bi-Directional Check Valve Design for Full-Bridge Oscillation Transformer*. Bachelor Project. Aalborg University.

Danfoss (2023). "Roadmap for decarbonizing cities". In: URL: https://cdn.sanity.io/files/5zabm86v/production/0397994639d425772172c5dec4f83180a9095ad4.pdf.

Hansen, Anders Hedegaard (2023). *Fluid power systems*. Springer. ISBN: 978-3-031-15088-3.

Hansen, Anders Hedegaard and Lasse Schmidt (2024). *Lektorer i energi: De store entreprenørmaskiner er kæmpe, oversete klimasyndere*. Altinget. URL: https://www.altinget.dk/transport/artikel/forskere-de-store-entreprenoermaskiner-er-kaempe-oversete-klimasyndere.

Johansen, Per and Anders Hedegaard Hansen (Oct. 2023). "A Digital Hydraulic Full-Bridge Oscillation Transformer". English. In: *Proceedings of ASME/BATH 2023 Symposium on Fluid Power and Motion Control, FPMC 2023*. 2023 ASME/BATH Symposium on Fluid Power and Motion Control, FPMC 2023 ; Conference date: 16-10-2023 Through 18-10-2023. United States: The American Society of Mechanical Engineers (ASME). DOI: 10.1115/FPMC2023-111865. URL: https://event.asme.org/FPMC.

MathWorks (2025a). *Reinforcement Learning for Control Systems Applications*. URL: https://www.mathworks.com/help/reinforcement-learning/ug/reinforcement-learning-for-control-systems-applications.html.

– (2025b). *Time Delays in Linear Systems*. MathWorks. URL: https://se.mathworks.com/help/control/ug/time-delays-in-linear-systems.html.

– (2025c). *Train Agents Using Parallel Computing and GPUs*. URL: https://www.mathworks.com/help/reinforcement-learning/ug/train-agents-using-parallel-computing-and-gpu.html.

– (2025d). *Train Agents Using Parallel Computing and GPUs*. URL: https://www.mathworks.com/help/reinforcement-learning/ug/td3-agents.html.

– (2025e). *What Is Model Predictive Control?* URL: https://www.mathworks.com/help/mpc/gs/what-is-mpc.html.

– (2025f). *What Is Reinforcement Learning?* URL: https://se.mathworks.com/discovery/reinforcement-learning.html.

Mortensen, Kasper Juhl et al. (2025). *Neural Network Surrogate Modelling and Iterative Learning Control of Hydraulic Full-Bridge Oscillation Transformer*. Semester Project. Aalborg University.

FBoT states A

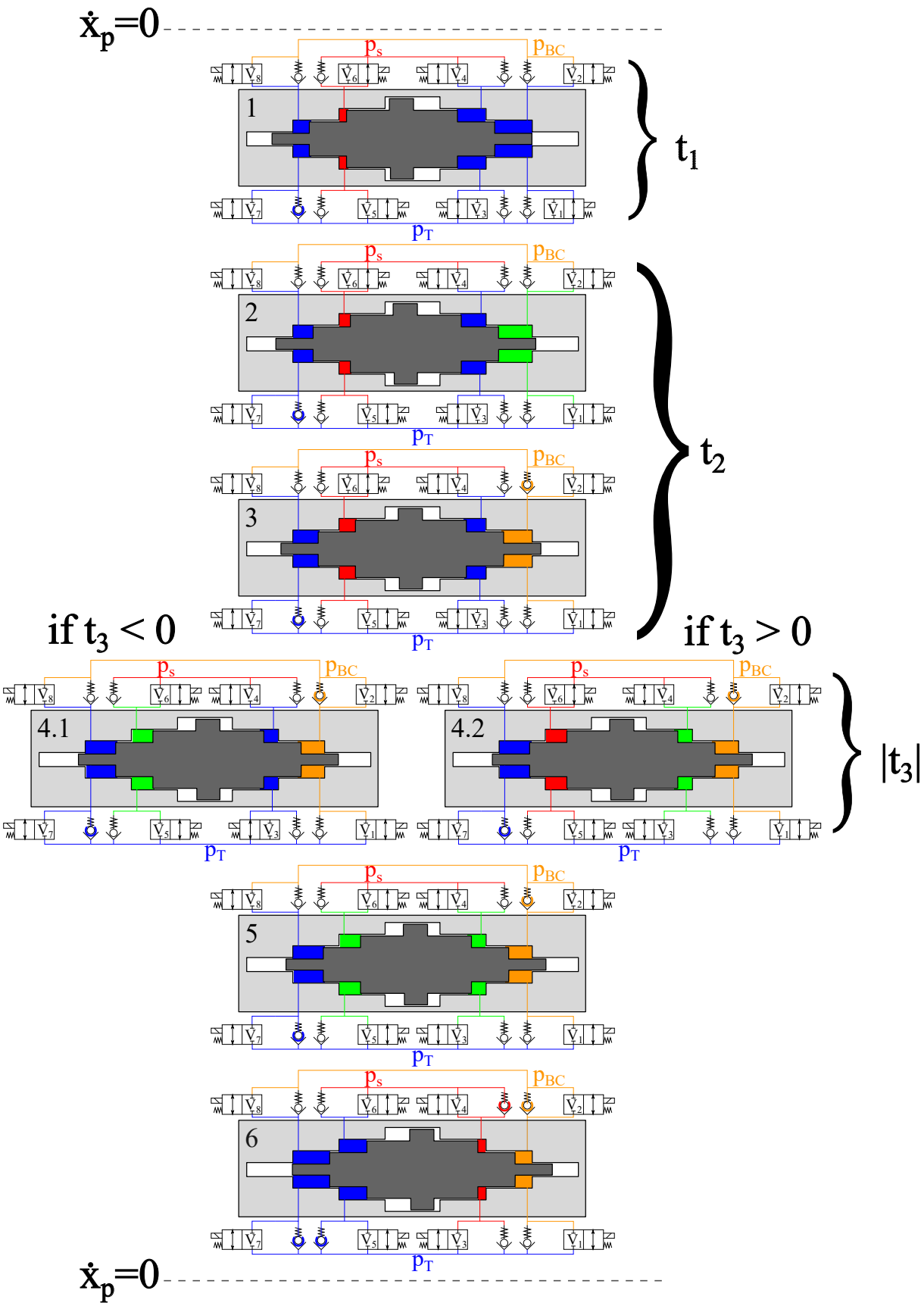


Figure A.1. Larger version of the states when moving from left to right.

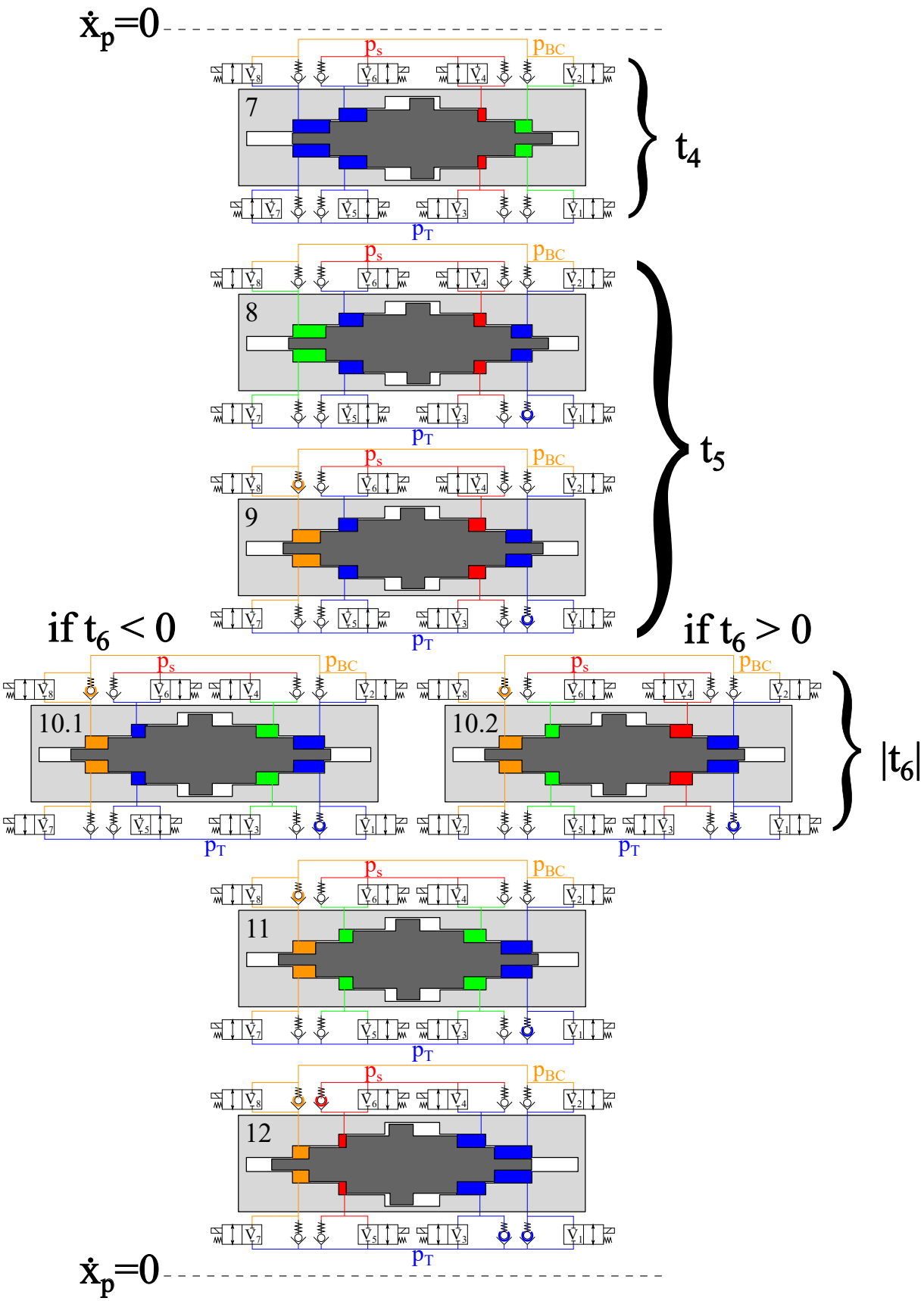


Figure A.2. Larger version of the states when moving from right to left

General Theory of Neural Networks B

A Neural Network is a general function approximation which consists of artificial neurons, also called units, gathered in layers. Signals travel from the input layer through the hidden layers to the to the output layer. A depiction of this can be seen in Figure B.1. The depth of the Neural Network is defined by the number of hidden layers. Neural Networks are generally classified as deep neural networks if there is more than one hidden layer. The width of a neural network is defined as the number of units in each hidden layer. The layers can be connected in multiple different ways

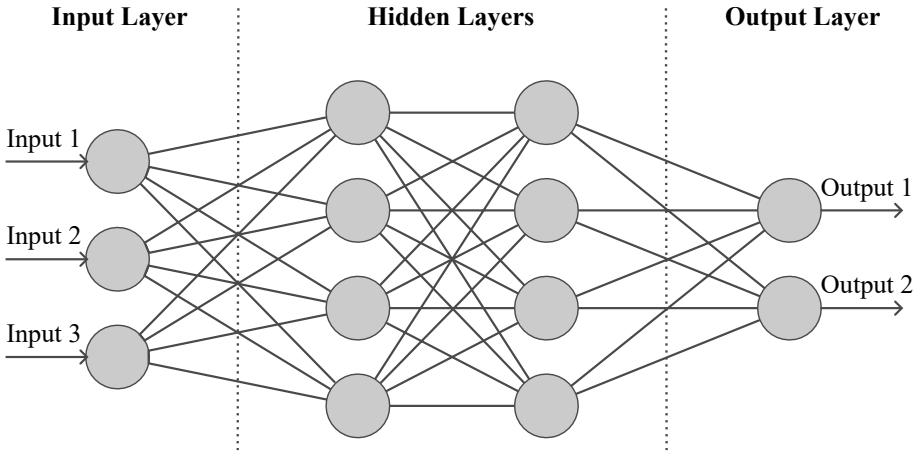


Figure B.1. Depiction of fully-connected feed-forward neural network with an input layer, two hidden layers and an output layer.

The output of each unit is called its activation. This is calculated for each layer with Equation (B.1). The function f is the activation function which is normally non-linear, as this is necessary for capturing non-linear relations.

$$f(\mathbf{W}\mathbf{a}^{(0)} + \mathbf{b}) = \mathbf{a}^{(1)} \tag{B.1}$$

Here, $\mathbf{W}$ is a $m \times n$ matrix of the weights where m is the number of units in the given layer and n is the number of units in the previous layer. $\mathbf{a}^{(0)}$ is a vector of activations from the previous layer, and $\mathbf{a}^{(1)}$ are the outputs from a given layer, $\mathbf{b}$ is a vector of biases.

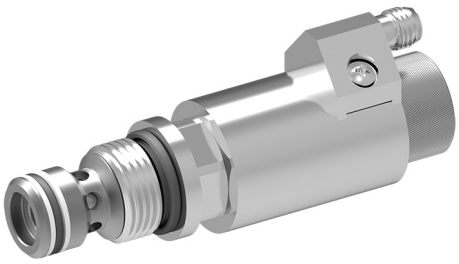
The two activation functions used in this project is the sigmoid activation function and the RELU activation function which is shown in Equation (B.2). The input x to the activation function is given as $\mathbf{W}\mathbf{a}^{(0)} + \mathbf{b}$

$$\sigma(x) = \frac{e^x}{e^x + 1}, \qquad ReLU(x) = \frac{x + |x|}{2} \tag{B.2}$$

Datasheet for on/off valves C

2/2 Cartridge Seat Valve, Size 5

Q_{max} = 30 l/min, p_{max} = 350 bar
Digital valve, bidirectional seat-valve shut-off, direct acting
Series WS22GD…/ WS22OD…

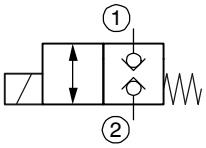


- For use in digital hydraulics
- With bidirectional seat-valve shut-off
- Compact construction for cavity type ALM – M20x1.5
- High switching performance
- Short response times
- All exposed parts with zinc-nickel plating
- High pressure wet-armature solenoids
- The slip-on coil can be rotated, and it can be replaced without opening the hydraulic envelope
- Can be fitted in a line-mounting body

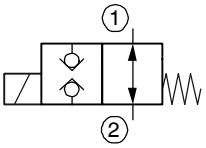
1 Description

These direct acting 2/2 solenoid operated directional seat valves, series WS22GD… / WS22OD…, are screw-in cartridges with a M20x1.5 or 3/4-16 UNF mounting thread. They are designed on the poppet/seat principle, and are therefore virtually leak-free in both directions of flow (bidirectional seat-valve shut-off). Over-excitation, preferably through an electronic switching device (booster), is required to operate the solenoid. Combined with the low mass of the moving parts, this results in short response times and high switching performance in a compact package. "De-energised closed" and "de-energised open" functions are available. The straightforward design delivers an outstanding price/performance ratio. The valves are used in applications in digital hydraulics, where fast response and long life with minimum size are vitally important. All external parts of the cartridge are zinc-nickel plated to DIN 50 979 and are thus suitable for use in the harshest operating environments. The slip-on coils can be replaced without opening the hydraulic envelope and can be positioned at any angle through 360°.

2 Symbol



WS22GD…



WS22OD…

3 Technical data

General characteristics	Description, value, unit
Designation	2/2 cartridge seat valve
Design	digital valve, bidirectional seat-valve shut-off, direct acting poppet/seat design (pressure balanced)
Mounting method	screw-in cartridge M20x1.5 or 3/4-16 UNF
Tightening torque	50 Nm ± 10 %
Size	nominal size 5, cavity type ALM M20x1.5 cavity type AL 3/4-16 UNF please contact BUCHER
Weight	0.20 kg

BUCHER
hydraulics

General characteristics	Description, value, unit
Mounting attitude	unrestricted
Ambient temperature range	-25°C…+80 °C

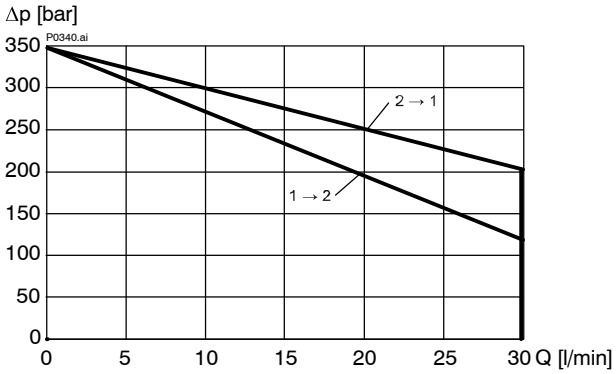
Hydraulic characteristics	Description, value, unit
Maximum operating pressure (ports 1 and 2)	350 bar
Maximum flow rate	30 l/min
Flow direction	1 → 2 / 2 → 1, see symbols
Hydraulic fluid	HL and HLP mineral oil to DIN 51 524; for other fluids, please contact BUCHER
Hydraulic fluid temperature range	-25 °C … +80 °C
Viscosity range	10…500 mm^2/s (cSt), recommended 15...250 mm^2/s (cSt)
Minimum fluid cleanliness Cleanliness class to ISO 4406 : 1999	class 20/18/15

Electrical characteristics		Description, value, unit
Excitation voltage		48 V DC (standard)
Length of over-excitation		4…5 ms
Supply voltage		12 V DC (standard)
Voltage tolerance		± 5 % (at ambient temperature < 60°C : ± 10 %)
Nominal power consumption		15 W at 12 V DC
Switching time	- model WS22G…	6 … 20 ms (energising) 10 … 30 ms (deenergising)
	- model WS22O…	6 … 30 ms (energising) 5 … 20 ms (deenergising) These times are strongly influenced by fluid pressure, flow rate and viscosity, as well as by the dwell time under pressure.
Relative duty cycle	- static	100 %
Duty cycle / switching frequency	- dynamic	see characteristics
Protection class to ISO 20 653 / EN 60 529		IP 65
Electrical connection:		3-pin plug M8x1
	- PIN 1	48 / 12 V DC
	- PIN 3	0 V
	- PIN 4	not used

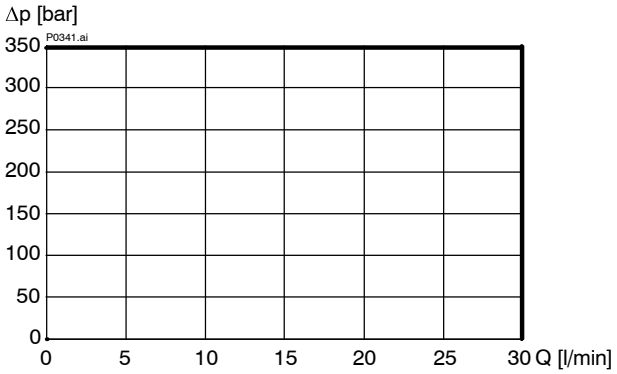
4 Performance graphs

measured with oil viscosity 33 mm²/s (cSt), coil at steady-state temperature and 10 % undervoltage

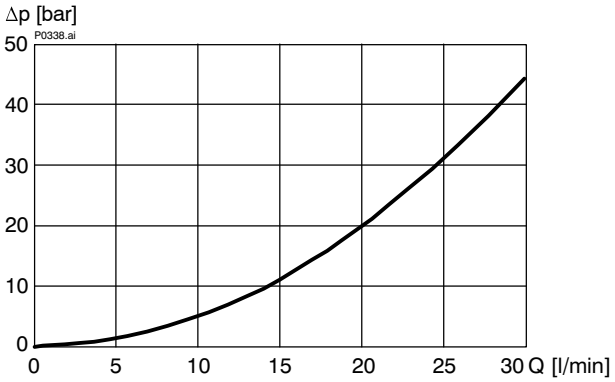
p = f (Q) Performance limits
[WS22GD…]



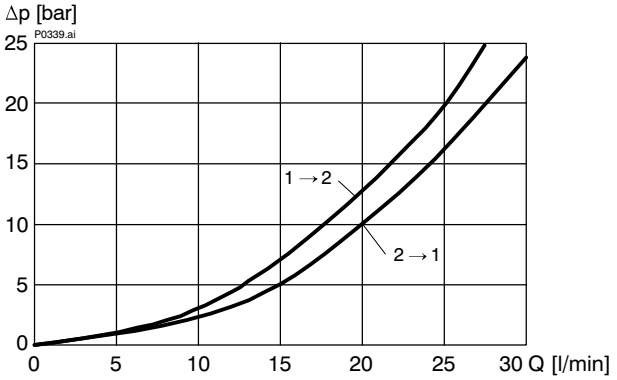
p = f (Q) Performance limits
[WS22OD…]



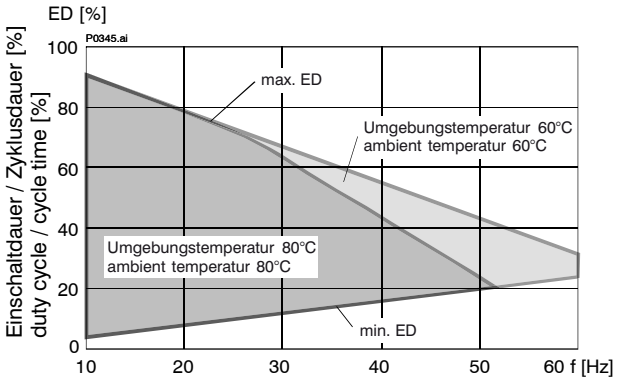
Δp = f (Q) Pressure drop - Flow rate characteristic
[WS22GD…]

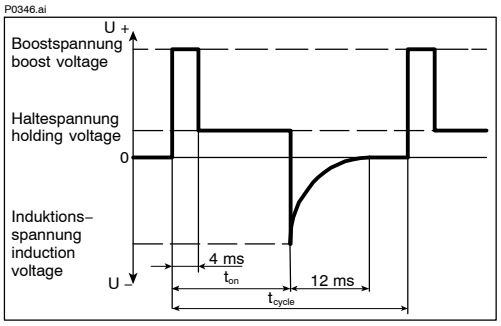


Δp = f (Q) Pressure drop - Flow rate characteristic
[WS22OD…]



ED = f (f) duty cycle - switching frequency - characteristic [at steady-state coil temperature]

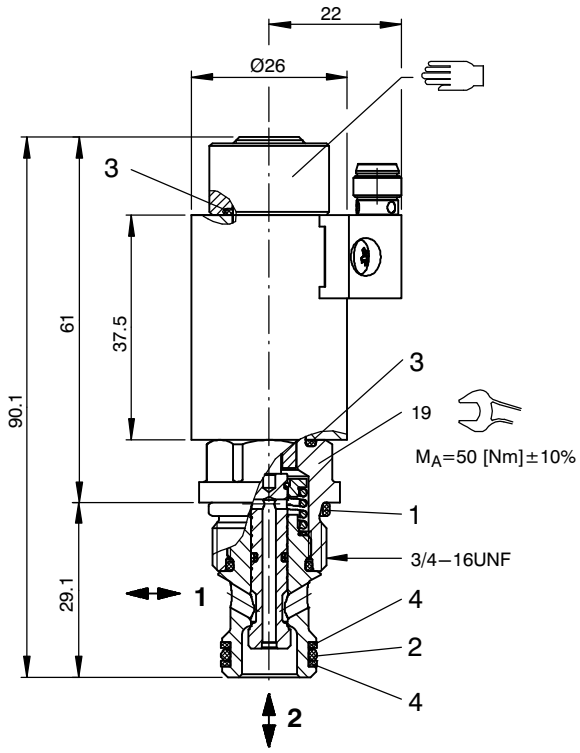




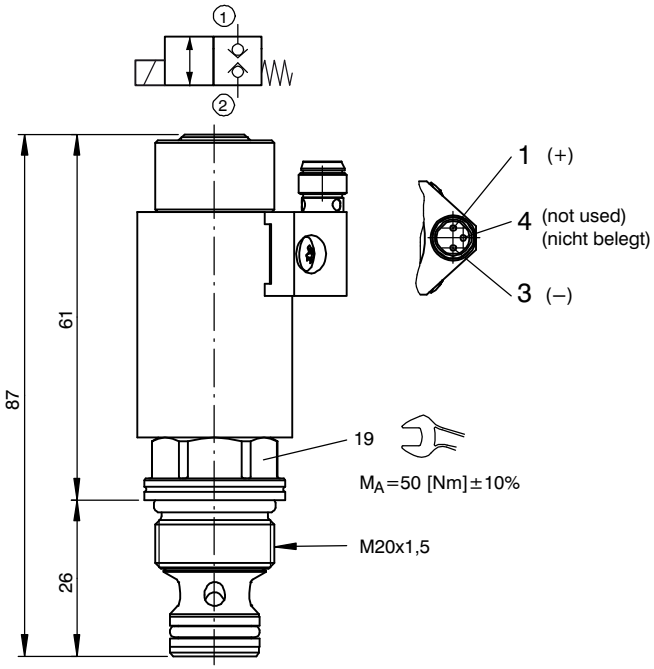
$$ED\ [\%] = \frac{t_{on}}{t_{cycle}} \times 100$$

5 Dimensions & sectional view

5.1 “Normally closed” design WS22GD…

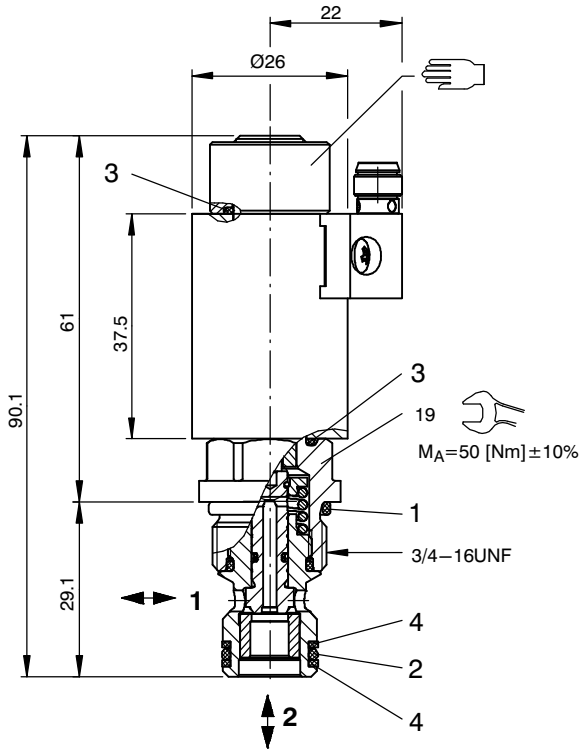


with 3/4-16 UNF thread – cavity type AL
please contact BUCHER

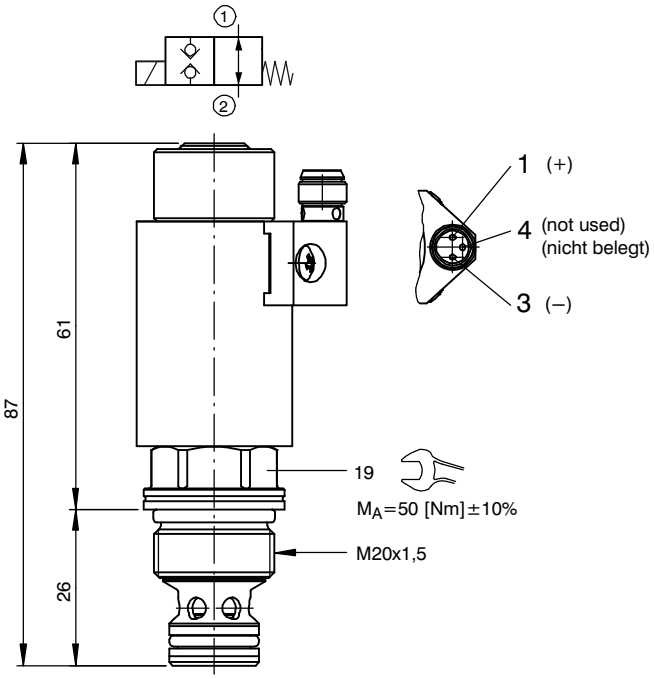


with M20x1.5 thread – cavity type ALM

5.2 “Normally open” design WS22OD…



with 3/4-16 UNF thread – cavity type AL
please contact BUCHER



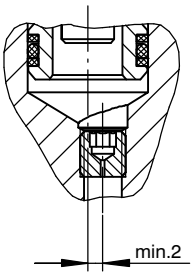
with M20x1.5 thread – cavity type ALM

6 Installation information

IMPORTANT!
When fitting the cartridges, use the specified tightening torque. No adjustments are necessary, since the cartridges are set in the factory.

ATTENTION!
Only qualified personnel with mechanical skills may carry out any maintenance work. Generally, the only work that should ever be undertaken is to check, and possibly replace, the seals. When changing seals, oil or grease the new seals thoroughly before fitting them.

ATTENTION!
If an orifice is fitted directly in port 2 close to the valve, and if the flow direction is from 2 to 1, it is important to ensure that the axis of the orifice drilling is offset from the valve axis by at least 2 mm!



3/4-16 UNF "A" – NBR seal kit no. DS-435-N [1)]

Item	Qty.	Description
1	1	O-ring no. 017 ∅ 17,17 x 1,78 N90
2	1	O-ring no. 014 ∅ 12,42 x 1,78 N90
3	2	O-ring ∅ 12.00 x 1.50 Viton
4	2	Backup ring ∅10.70 x 1,45 x 1,0 FI0751

IMPORTANT!
1) Seal kit with FKM (Viton) seals, no. DS-435-V

M20x1.5 "Z" - NBR seal kit no. DS-436-N [1)]

Item	Qty.	Description
1	1	O-ring no. 017 ∅ 17,17 x 1,78 N90
2	1	O-ring no. 013 ∅ 10,82 x 1,78 N90
3	2	O-ring ∅ 12.00 x 1.50 Viton
4	2	Backup ring ∅ 9.90 x 1,45 x 1,4 FI0751

IMPORTANT!
1) Seal kit with FKM (Viton) seals, no. DS-436-V

7 Ordering code

Ex. W | S | 22G | D | Z | 5 | _ | - | 1 | 12 | D | _

W	=	directional valve
S	=	seat valve, direct acting
22G	=	2/2 function, de-energised closed
22O	=	2/2 function, de-energised open
D	=	digital valve
Z	=	special features - with M20x1.5 thread (**standard**)
A	=	standard model - with 3/4 - 16 UNF thread (please contact Bucher)
5	=	nominal size 5
(blank)	=	NBR (Nitrile) seals (**standard**)
V	=	FKM (Viton) seals (special seals - please contact BUCHER)
1 ... 9	=	design stage (omit when ordering new units)
...	=	voltage e.g. 12 (12 V)
D	=	current DC
(blank)	=	M8x1 male connector (**standard**)
F	=	for flying leads (1000 mm), please contact Bucher

BUCHER
hydraulics

8 Related data sheets

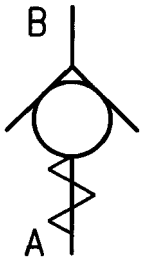
Reference	(Old no.)	Description
400-P-040011	(i-32)	The form-tool hire programme
400-P-040171	(i-33.10)	Cavity type AL
400-P-040201	(i-33.13)	Cavity type ALM
400-P-720101	(G-4.10)	Line-mounting body, type GALA (G 3/8")
400-P-720105	(G-4.11)	Line-mounting body, type GALMA (M20 x 1.5)

info.ch@bucherhydraulics.com www.bucherhydraulics.com

Classification: 430.300.-.305.305.300

Check valves D

Check Valves, Size 04 ... 16
Spherical Poppet-type, Screw-in Design
Series RKVG ... 80 l/min, 350 bar (500 bar)



1 General

1.1 Product description

Series RKVG units are screw-in check valves with mounting threads ranging from G 1/8" to G 3/4". For other thread forms, contact Bucher Hydraulics.

The valves prevent flow against the screw-in direction (A → B). In the opposite direction, the opening pressure is 0.2 to 1 bar.

The cavities are identical to those of the RVE/RKVE valves (REG-02 cavity only).

The units are spring-closed poppet valves with hardened poppets and seats. The poppet is fully guided, with a spherically shaped sealing surface.

A metal cutting lip seals the leakage path between the valve and cavity wall.

The valves can be used for pressure relief in the opening direction, but only to a limited extent (consult Bucher Hydraulics for such applications).

1.2 Advantages

- Virtually leak-free
- High pressure rating
- Compact construction
- Spring is enclosed

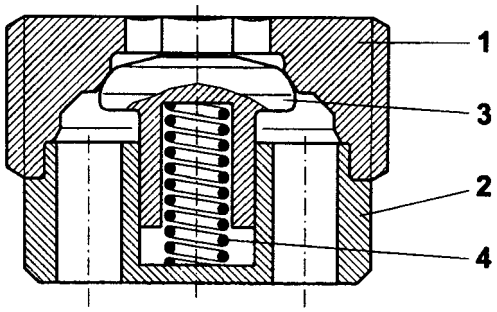
2 Main characteristics

Designation	check valve / non-return valve
Design	spherical poppet design
Mounting method	screw-in cartridge
Size	nominal 4...16 mm. See Table in section 5, Dimensions
Dimensions	see Table in section 5, Dimensions
Mounting attitude	unrestricted
No-flow direction	A → B (see symbol)
Operating pressure range	... 350 bar (for higher pressures, contact Bucher Hydraulics)
Opening pressure	0.2 … 1 bar
Flow rate, Q max.	... 80 l/min
Fluid	HL and HLP hydraulic oils to DIN 51524
Temperature range	-30°C... +80°C
Viscosity range	10... 500 mm²/s (cSt)
Min. fluid cleanliness	18/14 to ISO 4406 / CETOP RP70H, 8…9 to NAS 1638

For applications outside these parameters, please contact Bucher Hydraulics.

Classification: **4.30.30.40.16**
Reference: 170-P-051100-E-07 / 08.15

3 Schematic section

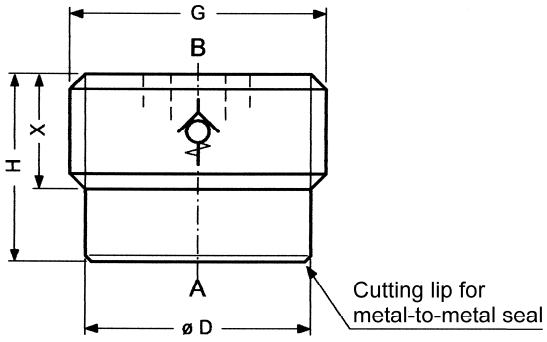


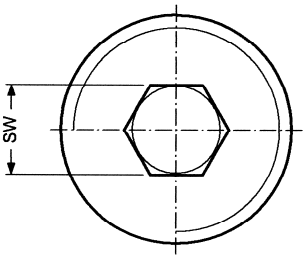
4 Components

Item	Qty.	Description
1	1	Valve seat
2	1	Valve body
3	1	Valve poppet
4	1	Spring

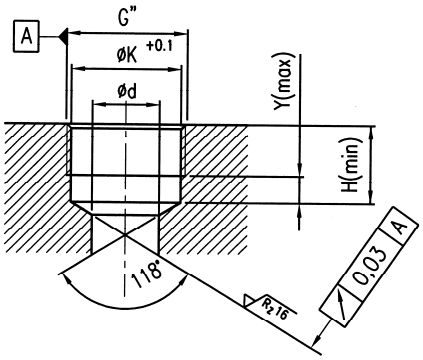
5 Dimensions

5.1 Valve





5.2 Dimensions – cavity type REG-02

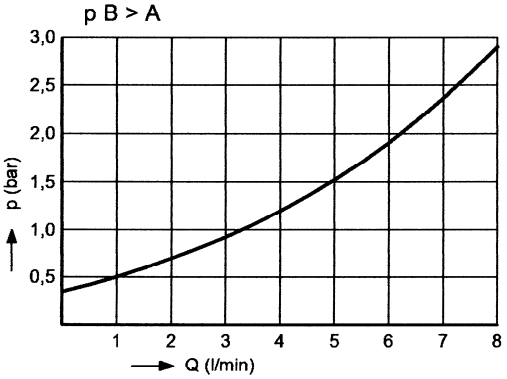


	Q Nom = Qmax (l/min)	G	ØD	H	X	SW	Mass (g)	ØK	Ød$_{max}$	h $_{min}$	Y$_{max}$	Tightening torque (Nm)
RKVG-04	8	G 1/8“	8.5	9.0	5.5	3,5	4	8.70	6.0	10.0	2.5	8
RKVG-06	15	G 1/4“	11.6	11.3	6.3	5	8	11.75	8.0	11.5	4.0	20
RKVG-08	30	G 3/8“	14.9	13.0	7.8	6	14	15.25	11.5	13.5	4.0	25
RKVG-10	50	G 1/2“	18.8	15.0	9.6	8	24	19.00	15.5	16.0	4.5	40
RKVG-16	80	G 3/4“	24.2	18.5	11.5	10	42	24.50	20.0	19.0	6.0	60

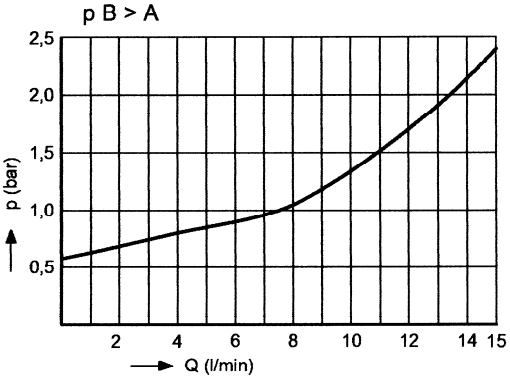
6 Performance graphs

measured with oil viscosity 33 mm²/s (cSt)

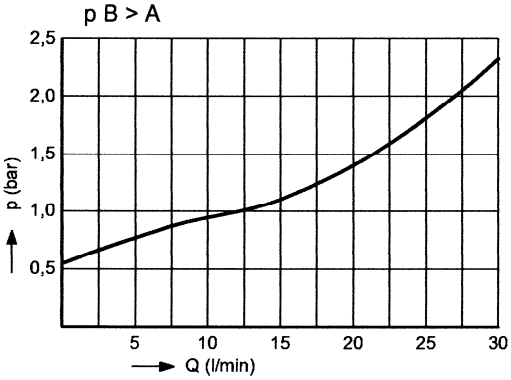
RKVG-04
p B > A



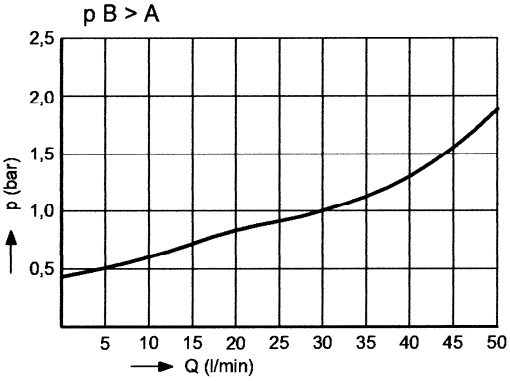
RKVG-06
p B > A



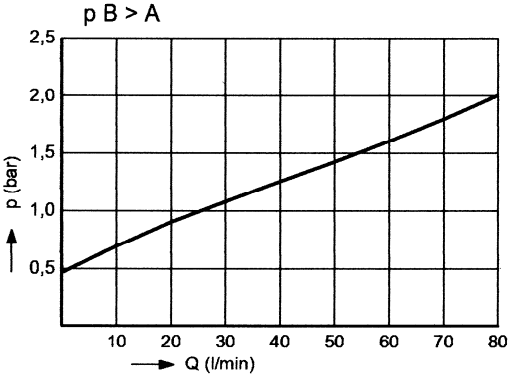
RKVG-08
p B > A



RKVG-10
p B > A



RKVG-16
p B > A



7 Ordering details

Model code key

Ex. R K V G - __ - __ - __

Check valve, screw-in type
K = spherical poppet
G = no-flow direction opposite to screw-in direction

Thread

Whitworth pipe thread	G
Metrical thread	M
UNF thread	U

Nominal size

04
06
08
10
16

Opening pressure

0.2 bar	02
0.5 bar	05
1 bar	1

Contact Bucher Hydraulics for further advice on:
- other opening pressures
- special materials
- customised designs

8 Design and installation notes

The installation dimensions and tolerances must be maintained.

We offer form tools for hire or sale.

Referring to the free-flow direction, nozzles and orifices must not be situated directly before the check valve (see Data Sheet 170-P-059000-E).

When fitting the valve, take particular care to ensure that:
- the valve cutting lip is firmly seated on the sealing surface
- valve components are not deformed by the use of excessive force

Use the specified tightening torque when fitting the valve.

9 Application notes

The maximum operating pressure must not be exceeded and any pressure peaks must be taken into consideration.

The specified nominal flow rate must not be exceeded.

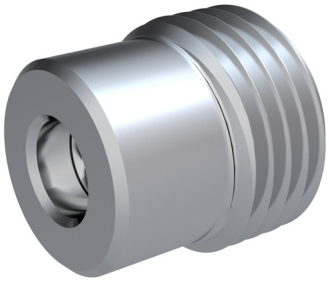
In applications such as accumulator circuits, where sudden pressure can be applied to the valve in the free-flow direction, ensure that the specified flow ratings are not exceeded. In dynamic accumulator circuits, use the internally damped valves.

Buyers bear the sole responsibility for ensuring that the selected products are suitable for their applications. Buyers normally establish this by undertaking qualification programs on test stands, or by evaluating the performance of prototype machines or systems.

info.dah@ bucherhydraulics.com **www.bucherhydraulics.com**

Check Valve, Size 04 … 16

Spherical poppet-type, screw-in design
Series RKVE …-VD 120 l/min, 350 bar



- for the same pressure differential, $Q_{max.}$ is around 50% higher than with standard RKVE valves
- no soft seal, therefore not temperature dependent
- same cavity as the RKVG and RVE series
- very low leakage
- with enclosed spring

1 Description

Series RKVE ...-VD screw-in cartridge check valves are furnished with G ⅛" … G ¾" threads, depending on their nominal size. Requests for other mounting threads will be subject to negotiation with the factory.

The valves prevent flow in the screw-in direction (B -> A) and open in the opposite direction. Opening pressures of 0.2, 0.5 and 1 bar can be supplied. For higher opening pressures, our RVVE preload valves with extended overall length are available (see data sheet 170-P-051010-E).

The cavity used is the REG-02 (118°), which can be manufactured by simple recessed thread tapping. Our RKVG and RVE series valves can also be used in this cavity. Installing the valves needs special fitting tools, which we can supply.

A metal cutting lip on the valve engages with the 118° bevel in the cavity, providing a metal-to-metal seal. By eliminating the soft seal, the valves can be applied without regard to temperature.

The units are spring-closed spherical-poppet valves. The body and seat are press-fitted together, with a guided poppet and an enclosed spring fitted between them. The valve seat, poppet and body are hardened. The properties of the sealing faces have been enhanced by precision mechanical processing.

Thanks to a fundamental redesign of the standard RKVE04…16 valves, we have achieved a 50% increase in the permissible flow rate of the new RKVE…-VD valve series for the same pressure differential. This means that in many cases it is possible to use a smaller valve and thus save installation space and costs.

The valves can be used for pressure relief in the opening direction, but only to a limited extent (please contact Bucher Hydraulics for such applications).

2 Symbol

BUCHER
hydraulics

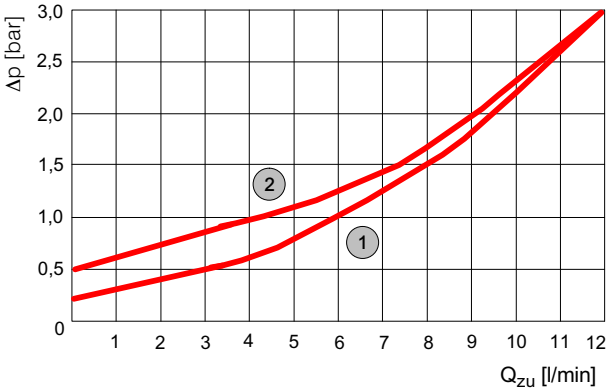
3 Technical data

General characteristics	Unit	Description, value
Type		check valve
Design		spherical poppet-type
Mounting method		screw-in cartridge
Size		nominal 04…16 mm (see table section 5 Dimensions)
Dimensions	mm	see table section 5: Dimensions
Mounting attitude		unrestricted
No-flow direction		B -> A (symbol see section 2)
Operating pressure	bar	350 (for higher pressures please contact Bucher Hydraulics)
Opening pressure	bar	0,2 / 0,5 / 1
Flow rate $Q_{max.}$	l/min	120
Fluid		HL and HLP hydraulic oils to DIN 51524, for other fluids please contact Bucher Hydraulics
Temperature range	°C	-30 … + 120
Viscosity range	mm^2/s [cSt]	10 … 500
Minimum fluid cleanliness		ISO 4406 code 20/18/15 (see section 11)

4 Performance graphs

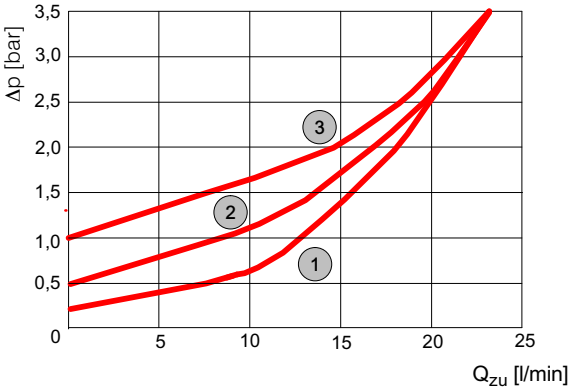
Measured with oil viscosity 33 mm^2/s (cSt)

4.1 RKVE-G-04-..-VD



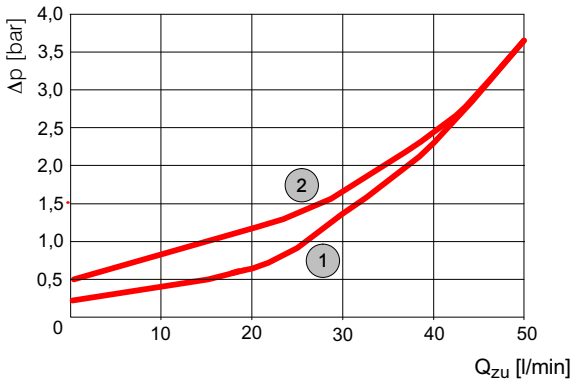
1	RKVE-G-04-02-VD
2	RKVE-G-04-05-VD

4.2 RKVE-G-06-..-VD



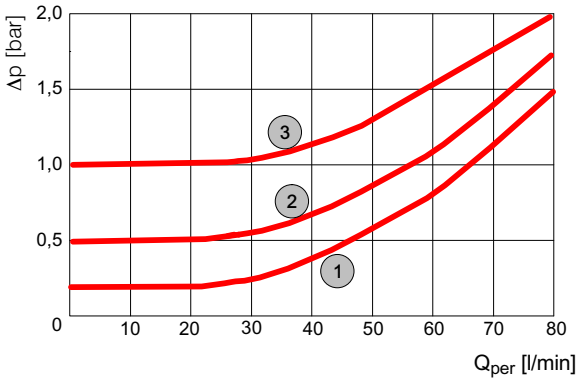
1	RKVE-G-06-02-VD
2	RKVE-G-06-05-VD
3	RKVE-G-06-1-VD

4.3 RKVE-G-08-..-VD



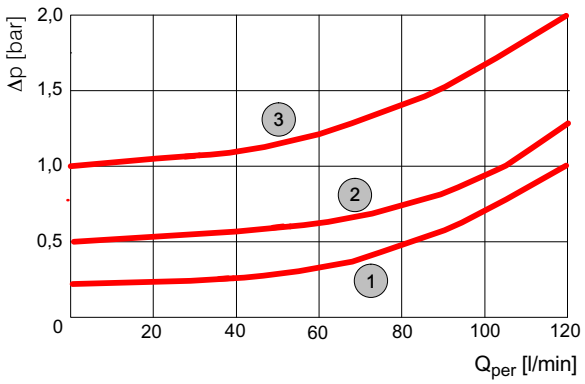
1	RKVE-G-08-02-VD
2	RKVE-G-08-05-VD

4.4 RKVE-G-10-..-VD



1	RKVE-G-10-02-VD
2	RKVE-G-10-05-VD
3	RKVE-G-10-1-VD

4.5 RKVE-G-16-..-VD

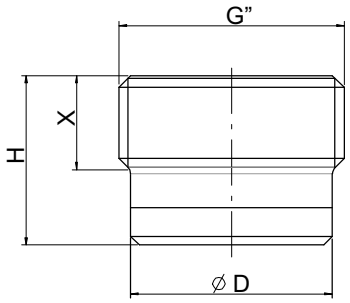


1	RKVE-G-16-02-VD
2	RKVE-G-16-05-VD
3	RKVE-G-16-1-VD

BUCHER
hydraulics

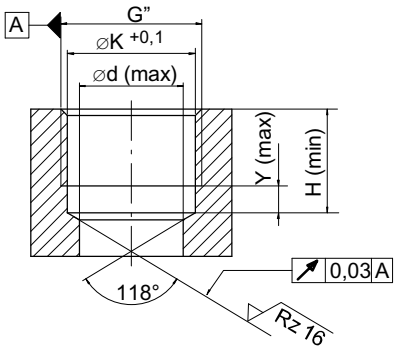
5 Dimensions

5.1 Dimensions - valve



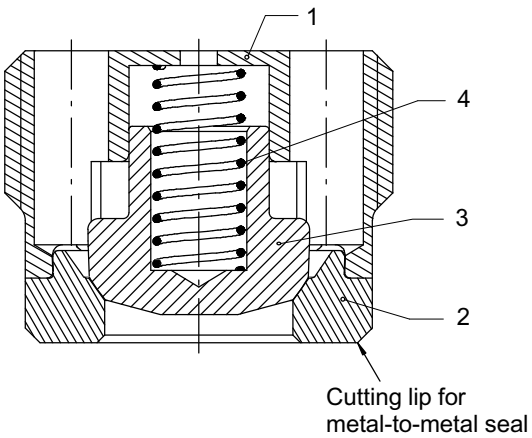
Type	$Q_{Nom}=Q_{max}$ [l/min]	G [mm]	⌀D [mm]	H [mm]	X [mm]	Tightening torque [Nm]	Fitting tool type
RKVE-04-…-VD	12	G1/8“	8,5	10,0	5,0	8	M-04
RKVE06-…-VD	25	G1/4“	11,5	11,3	5,5	20	M-06
RKVE-08-…-VD	50	G3/8“	14,9	13,3	7,0	25	M-08
RKVE-10-…-VD	80	G1/2“	18,8	15,9	9,0	50	M-10
RKVE-16-…-VD	120	G3/4“	24,3	18,9	10,5	60	MKS-16 / M-16

5.2 Dimensions - cavity type REG-02



Type	G	⌀ K [mm]	⌀ d [mm]	Y [mm]	H [mm]
RKVE-04-…	G1/8“	8,7	6,0	2,5	10,0
RKVE-06-…	G1/4“	11,75	8,0	4,0	11,3
RKVE-08-…	G3/8“	15,25	11,5	4,0	13,3
RKVE-10-…	G1/2“	19,0	15,5	4,5	15,9
RKVE-16-…	G3/4“	24,5	20,0	6,0	18,9

6 Schematic section



Item	Qty.	Description
1	1	Valve body
2	1	Valve seat
3	1	Valve poppet
4	1	Spring

7 Design and installation notes

IMPORTANT:

- Be sure to keep to the installation dimensions and tolerances
- Use the specified tightening torque when fitting the valve

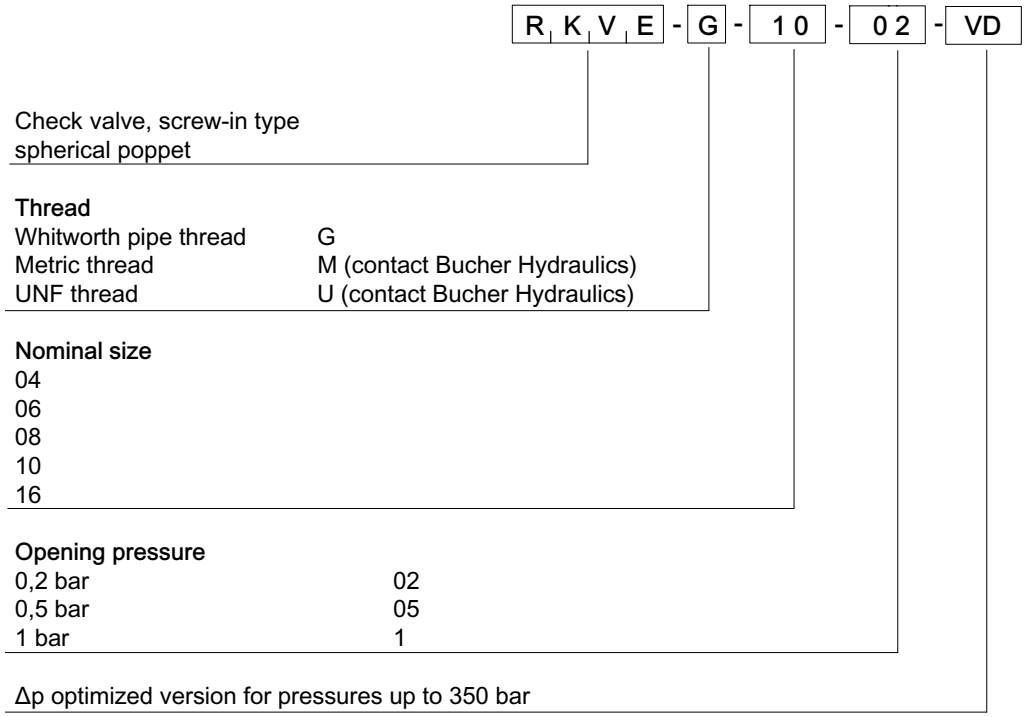
When fitting the valve, take particular care to ensure that:

- The valve is seated on the sealing surface
- Valve components are not deformed by the use of excessive force

Special fitting tools can be supplied.

- Do not situate nozzles and orifices directly before the check valve (referring to the free-flow direction) (see data sheet 170-P-059000-E)

8 Ordering code



9 Application notes

The maximum operating pressure must not be exceeded and any pressure peaks must be taken into consideration. The specified nominal flow rate must not be exceeded.

In applications such as accumulator circuits, where sudden pressure can be applied to the valve in the free-flow direction, ensure that the specified flow ratings are not exceeded.

Buyers bear the sole responsibility for ensuring that the selected products are suitable for their applications. Buyers normally establish this by undertaking qualification programs on the test stands or by evaluating the performance of prototype machines or systems.

Buyers bear the sole responsibility for ensuring that the selected products are suitable for their applications. Buyers normally establish this by undertaking qualification programs on the test stands or by evaluating the performance of prototype machines or systems.

10 Fluid

The oil for check valves RKVE must have a minimum cleanliness level of 20/18/15 to ISO 4406.

We recommend the use of fluids that contain anti-wear additives for operation with boundary lubrication. Fluids without appropriate additives reduce the service life of check valves. The user is responsible for maintaining, and regularly checking, the fluid quality.

11 Fluid cleanliness

Cleanliness class (RK) onto ISO 4406.

Code ISO 4406	Dirt particle number / 100 ml		
	≤ 4 µm	≤ 6 µm	≤ 14 µm
23/21/18	8000000	2000000	250000
22/20/18	4000000	1000000	250000
22/20/17	4000000	1000000	130000
22/20/16	4000000	1000000	64000
21/19/16	2000000	500000	64000
20/18/15	1000000	250000	32000
19/17/14	500000	130000	16000
18/16/13	250000	64000	8000
17/15/12	130000	32000	4000
16/14/12	64000	16000	4000
16/14/11	64000	16000	2000
15/13/10	32000	8000	1000
14/12/9	16000	4000	500
13/11/8	8000	2000	250

info.dah@bucherhydraulics.com www.bucherhydraulics.com

Classification: 430.315.340