10 Actuators

Technical processes are usually influenced or manipulated by actuators, which affect certain input variables. In most cases, electrical, hydraulic or pneumatic auxiliary power is required. The actuator's input variables, the manipulated variables, can be adjusted by feedforward or feedback control or by an operator. Actuators therefore act on the matter or energy flow of the process by means of information processing and are an important link between the signal level of the automation device and the technical process, Figure 10.1.

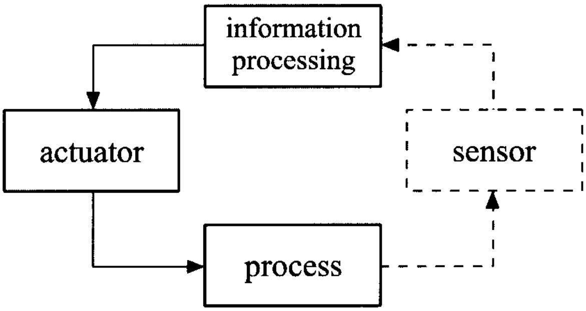


Figure 10.1. Actuator as connection between information processing and process

Actuators are applied in all fields of technology. Because of the manifold requirements, there exists a great variety of designs. This chapter deals with actuators for mechatronic systems. They normally have an electrical input and a mechanical output, *e.g.*, way, velocity or force, and use different kinds of auxiliary energies from electrical, pneumatic or hydraulic supplies. After a description of the basic structure, an overview of different actuator principles will be given. Actuators with different auxiliary energies will be presented and their properties and fields of application with regard to their selection will be compared. Mathematical models of electromagnets and electrical motors are already described in Chapter 5 and are not repeated here. However, detailed models will be derived for hydraulic and pneumatic actuators. After a comparison of the application areas of actuators, the design of fault-tolerant components as actuators and sensors is considered.

10.1 BASIC STRUCTURES OF ACTUATORS

Actuators usually transform low-power manipulated variables (*e.g.*, analog voltages 0...10V, applied currents 0...20mA or 4...20mA) into process input variables of a much higher power level. In mechatronic systems, the process input variable is often a flow of energy or matter. The power needed for actuating is provided by an auxiliary energy supply, which feeds the power amplifier integrated in the actuator. The auxiliary energy can be electrical, pneumatic or hydraulic. Many actuators have one of the basic structures shown in Figure 10.2. In most cases, the manipulated variable U_i is an electrical variable that is transformed into a controlled variable U_1 suitable for controlling the actuator drive by means of a *signal transformer*, Figure 10.2a. For an electrical actuator drive, this control variable could be another voltage or a voltage clock signal. For a pneumatic drive, it could be an air flow and for a hydraulic system an oil flow. The low-power-controlled variable U_1 is used as an input into the subsequent *actuator servo-drive* and is amplified into a higher power output signal U_2 by means of the auxiliary energy. For translational motions, the variable U_2 can be a force, a displacement or a velocity, for rotational motions it can be a torque, a rotational angle or a rotational speed.

Consequently, using the terminology of Chapter 2, this servo-drive is an active transformer or an active converter. The actuator drive's output variable U_2 often has to be transformed into a suitable range for the subsequent component. The corresponding element is called an *actuator transformer* with output variable U_3. It could be a lever transformer for displacements, a gear for angles and torques, or it can transform rotational motion into translational motion, like a spindle. Thus, a controlled actuator with an input variable U_i and an output variable U_3 is created, Figure 10.2a.

The relation between U_i and U_3 can be influenced by disturbances or by properties like friction, backlash, electromagnetic hysteresis or changes of the gain because of aging and wear. For this reason, the output variable U_3 is sometimes *feedback-controlled* if actuators of higher precision are required, Figure 10.2b. This requires an additional sensor in the actuator, which measures displacements, velocities or forces. An analog or digital position controller R_1 changes the manipulated variable U_i in such a way that the output variable U_3 conforms with the new input variable, the reference input variable U_{3r}. Hereby, it is possible to achieve a better static correspondence between U_{3r} and U_3 and a better dynamic behavior.

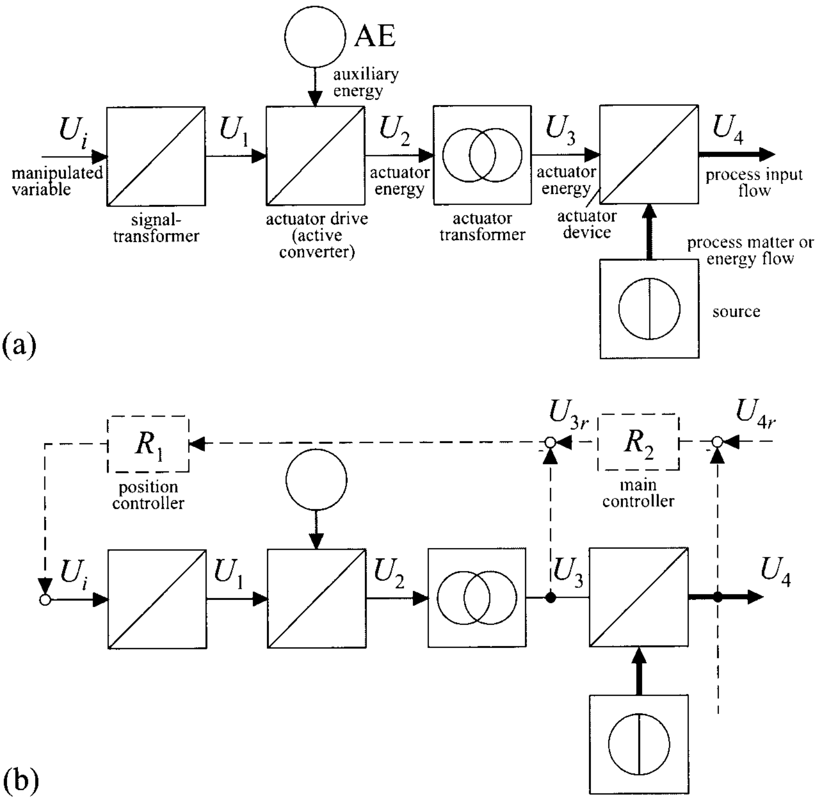


Figure 10.2. Basic structures of actuators: (a) open-loop-controlled actuator; (b) closed-loop-controlled actuator

In many cases, the variable that actuates on the technical process is not the output variable U_3 of the actuator drive but the matter or energy flow of a subsequent actuating device or valve, Figure 10.2a. This could be the air flow (throttle valve) or the fuel flow (injection valve) of a spark-ignition engine, the oil flow into a hydraulic work cylinder (main valve), the momentum flow (force) of the rudder of an airplane, the warm water flow of a heat exchanger (regulating valve) or the electrical current of an electrical motor (thyristor-current converter). The matter or energy flow provided by the actuating device lies in the power level of the process input and is usually significantly higher than the power level of the auxiliary energy. Therefore, the actuating mechanism is a second power amplifier and thus a second active element (transformer or transducer), so that the whole actuator (or the actuator system) contains at least two power amplifiers.

1. actuator: the manipulated variable controls the actuator power;
2. actuator device: the actuator power controls the power flow.

It is also possible to speak of a "primary actuator" and a "secondary actuator". Between them, there might be an additional energy or power amplifier, called an *actuator transformer* or converter. Because of the disturbances and non-linearities in the actuating mechanism, the output variable U_4 can be controlled by a second controller R_2. This actuator main controller (matter or energy flow controller) operates on the reference input of the position controller, which leads to a *cascade control system*, Figure 10.2b.

Hence, Figure 10.2 shows that actuators consist of a chain of energy or power amplifiers, Oppelt (1980, 1986), and that they can have several feedbacks. One can distinguish between the actuator drive, the mechanical actuator device and sensors, as shown in Figure 10.3, with an energy flow in the forward direction and an information flow in the backward direction. A comparison with Figure 1.1 shows that actuators are mechatronic systems themselves.

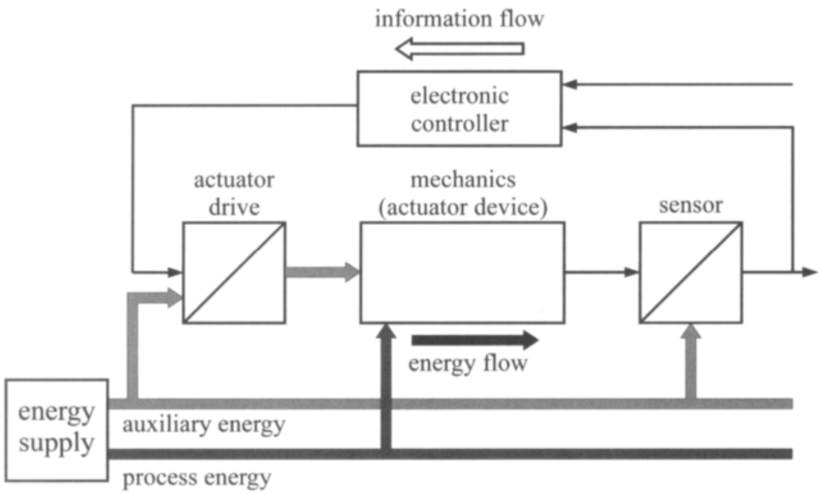


Figure 10.3. Electromechanical actuator as mechatronic system

In the following section, the technical implementations of actuators will be discussed. The descriptions will be limited to actuator drives and actuator devices with an electrical input, a mechanical output and powers up to about 5 kW.

10.2 GENERAL SURVEY OF ACTUATORS

To obtain a general view of the different types of actuator principles, this section will present the most important kinds of auxiliary energy and the basic transient responses, Raab (1990).

10.2.1 Types of Auxiliary Energy

Figure 10.4 schematically shows different kinds of auxiliary energies and force generation.

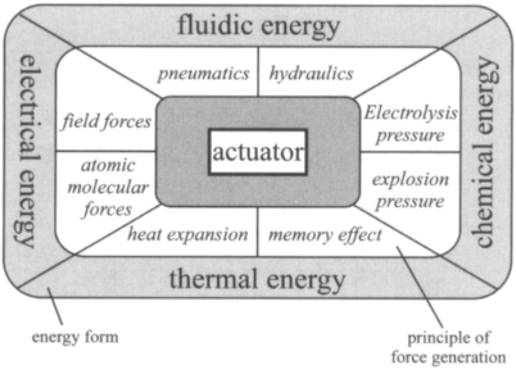


Figure 10.4. Auxiliary energies and forces to generate mechanical output variables

a) Electricity

In most cases, electrical energy already exists and is decentrally available. Its unproblematic generation and battery storage in combination with good conversion and transmission capability provides high flexibility. This is supported by an easy manipulation of energy flows with relatively low-priced semiconductor elements. The signal conversion and the actuator drive can be operated with the same kind of energy and maybe even with the same potential. Because of these advantages, the electrical auxiliary energy is usually given preference over other kinds of energy. Exceptions are made in the case of too-high actuating forces, high temperatures or for safety reasons.

b) Hydraulics

The oil flow of a hydraulic circuit usually has to be provided by an additional auxiliary energy supply. The operation pressures are relatively high (100 to 400 bar). The resulting advantages are high positioning forces and robust and compact actuator drives with a very high power–weight ratio.

c) Pneumatics

Pneumatic systems are realized both with vacuum (especially in automobiles) as well as with over-pressure relative to atmospheric pressure. The supply pressures are limited to 6–8 bar, in the process automation to 1.4 bar, which leads to larger devices compared to hydraulics. Furthermore, careful air conditioning is indispensable. Nevertheless, pneumatics provide a series of advantages such as a robust design and a reliable and safe operation, especially in a warm and explosive environment.

Table 10.1 shows different properties and characteristics of actuators used in automobiles for the discussed types of auxiliary energies.

Depending on these three main auxiliary energies, the following actuator principles can be distinguished:

- electromechanical actuators;
- fluid energy actuators;
- unconventional actuators.

Table 10.1. Properties and characteristics of auxiliary energies for actuators in vehicles, Raab (1990)

auxiliary energy	potential	average rated power	power/ weight ratio*	suitable for translational motion	suitable for rotational motion
		W	W/kg		
electricity battery generator	 12–24 V 14–26 V	 < 100 < 500	40–130	medium	good
hydraulics engine pressure hydr. system	 1–5 bar 30–200 bar	 < 100 > 1000	1000–2500	good	medium
pneumatics vacuum over pressure	 0.1–0.8 bar 6–8 bar	 < 100 > 1000	 5–25 200–400	good	medium

* without power supply

Figure 10.5 shows another classification into different realizations, Raab, Isermann (1990). Their characteristic properties and typical ranges of application will be described in the following subsections.

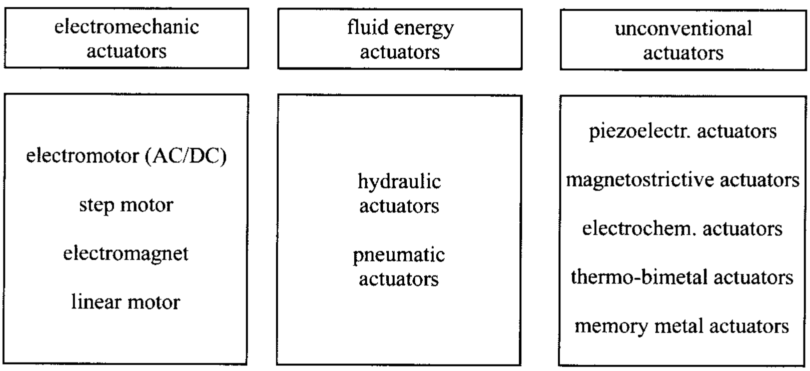


Figure 10.5. Classification of different actuator principles

10.2.2 Actuator Time Behavior and Control

The different actuator drives or complete actuators can also be distinguished by their transient behavior. If for a step increase of the controlled variable ΔU_1 the output variable ΔU_2 of the actuator drive, Figure 10.2a, changes proportionally to ΔU_1, such that

$$\Delta U_2(t) = K_p \Delta U_1(t) \tag{10.2.1}$$

the transient behavior is called *proportional action* (Figure 10.6a). If the output variable changes with a constant rate

$$\frac{dU_2(t)}{dt} = K_I \Delta U_1(t) \text{ or } \Delta U_2(t) = K_I \int_0^t \Delta U_1(t')dt' \tag{10.2.2}$$

then the transient behavior is called *integral action*, Figure 10.6b. The static behavior of proportional action actuator drives is represented by characteristic curves. Figure 10.7a gives examples of unique linear and non-linear characteristics. An example of an ambiguous characteristic is the hysteresis, Figure 10.7b, which is generated by dry friction or backlash. If $\Delta \dot{U}_2(t)$ is drawn above $U_1(t)$, one gets corresponding characteristics for integral action actuators.

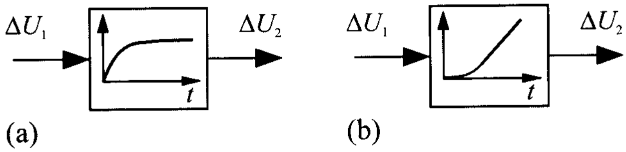


Figure 10.6. Transfer behavior of actuator drives: (a) proportional acting;(b) integral acting

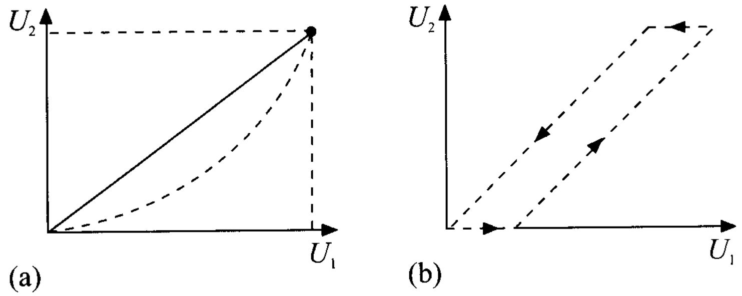


Figure 10.7. Static behavior of proportional acting actuator drives: (a) unique characteristics; (b) ambiguous characteristic (hysteresis)

Table 10.2 shows the typical static behavior of some actuator drives. In Figure 10.8, the block diagram of an actuator drive is shown, which is typical for several designs. The actuator force generation often follows a non-linear characteristic (*e.g.*, electromagnetic, pneumatic or hydraulic cylinder) with a dynamic behavior of low order (often almost linear and of first order). In this case, the positioning force acts on the mechanical converter (*e.g.*, guide bar with bearing and return spring). This mechanical part includes a mass m, a return spring constant c and a viscous and dry friction. The dry friction leads to another non-linearity, which is ambiguous (hysteresis).

Therefore, high-quality actuator drives are supplied with a cascade control like in Figure 10.9, with an underlying force control (current control for electromagnets, differential pressure control for fluidic cylinders) and an overlying positioning control. The non-linear characteristic of the force generation can be compensated for by an inverse characteristic, resulting in an approximately linear behavior of the force control. Furthermore, the non-linear behavior of the mechanical converter for the positioning control can be cancelled by compensation of the dry friction. With these possibilities of a software-based influence, the negative properties of actuator drives can be significantly improved, see Isermann, Keller (1993) and Section 10.5.3.

Table 10.2. Static transfer behavior of some actuators

actuator principle	proportional acting			integral acting		
	characteristic			characteristic		
	unique		non-unique	unique		non-unique
	linear	non-linear	non-linear (hysteresis)	linear	nonlinear	non-linear (hysteresis)
electro-mechanical actuators	stepper motors		electro-magnets	DC motors	AC motors (with switch)	electrical drives with friction or backlash
fluidic actuators			pneumatic diapraghm with return spring		hydraulic actuators	pneumatic cylinders
unconven-tional actuators			piezoceramic acutator magnetostrictive actuators memory metal actuators			

The actuator device following the actuator drive, Figure 10.2, which frequently manipulates a matter or energy flow and often has proportional behavior. The layout of the static characteristic of this actuator device depends on the static behavior of the subsequent process, such that the combined action of actuator and process defines a certain *overall operating characteristic.*

This operating characteristic should be approximately linear in closed-loop control, if this leads to stability with constant control parameters over the whole operating range. If the process has a linear characteristic, then the characteristic of the actuator also has to be designed as linear. For a non-linear process characteristic, a linear operating characteristic can be achieved by a corresponding inverse non-linear actuator characteristic (*e.g.*, for a flow valve). The adaptation to an operator sometimes makes non-linear actuator characteristics necessary, *e.g.*, the non-linear drive characteristic for combustion engines in automobiles and the adjustment of stick-rudder behavior in airplanes.

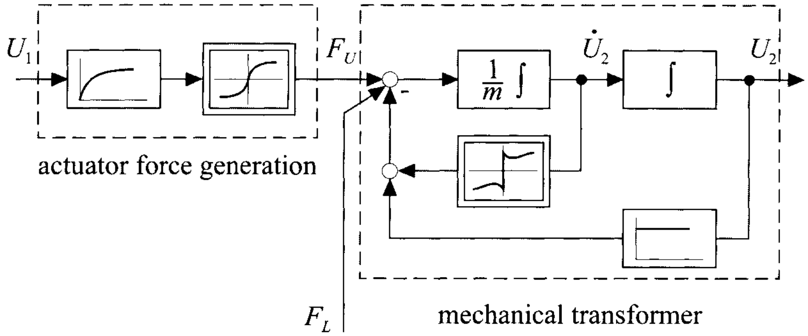


Figure 10.8. Block diagram of a proportional acting actuator drive with non-linear force generation (electrical, hydraulic, pneumatic) and mechanical transformer with viscous and dry friction for generating a speed and displacement: U_1: input of the actuator; U_2: output (displacement) of the actuator; F_U: actuation force; F_L: load force (from subsequent actuator device)

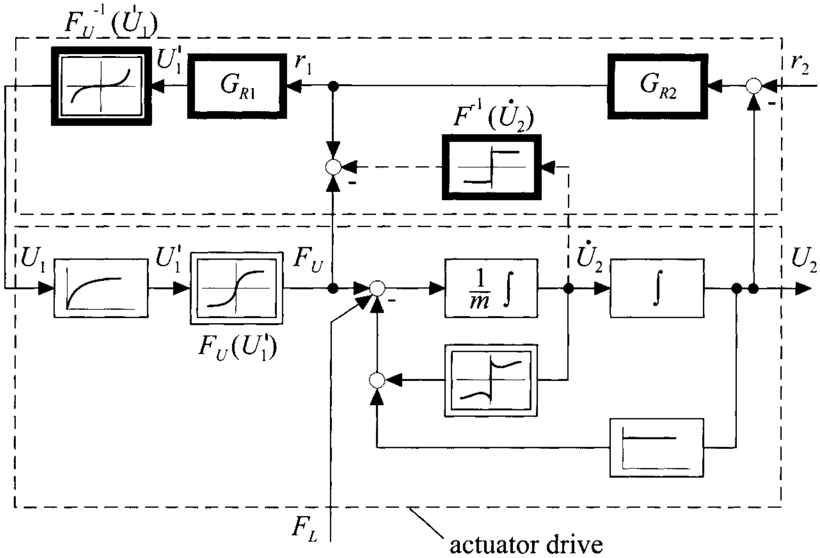


Figure 10.9. Block diagram of the cascade control of an actuator drive as Figure 10.8 with G_{R1}: force controller; G_{R2}: position controller; $F_U^{-1}(U_1')$: compensation of static non-linearitiy; $F^{-1}(\dot{U}_2)$: compensation of dry friction

10.2.3 Requirements for Actuators and Servo-drives

Unlike power-generating or power-consuming machines, the drives of actuators are not operated continuously but only for short periods of time, and have to be exact in positioning. Therefore, special drives were developed for this task, so-called *servo-drives*. They have to fulfil the following requirements:

- functioning in the four-quadrant operation (drive and brake in both directions);
- high overload capability;
- high resolution for exact positioning;
- good static transfer properties (as linear as possible, little friction, no backlash);
- fast, well-damped dynamic properties (little time constants, no overshoot);
- high range of longitudinal speed and rotational speed;
- high force generation or torque generation and little wear in standstill (breaking-off and holding);
- appropriate interfaces to the signal transformer of the manipulated variable.

10.3 ELECTROMECHANICAL ACTUATOR DRIVES

Electromechanical actuating units are widely used and are characterized by a large variety of different designs. The variety of types, especially of motor-driven actuators, permits a flexible customization for a wide range of applications, as can be seen in Table 10.3.

Table 10.3. General features of electromechanical (in particular electromotive) actuators

advantages	disadvantages
• good response characteristics; • highly dynamic; • flexible drive concepts; • high overall efficiency; • condition can be monitored well.	• restricted power density; • energy consumption during static operation; • restricted thermal range of operation; • high percentage of moving mechanics.

Electrical drives play a predominant role if small to medium actuating power must be provided. This fact can, in part, be attributed to the high availability of electric power and the number of ways to convert it. Additionally, electrical drives deliver a high degree of positioning accuracy and good dynamic performance. The high efficiency of the total system surpasses the efficiency of comparable pneumatic or hydraulic systems.

However, high dynamic requirements combined with a demand for higher actuating power impose limits on the applicability of electromechanical actuators. Demanding high actuating power results in large drives since the power density is physically limited by saturation effects of the magnetic materials.

Major disadvantages arise from the mechanical construction. If mounted directly onto other machines, electrical drives can be subject to severe strain due to shaking, which might necessitate extensive constructive countermeasures. High ambient temperatures can also affect the operation from effects such as degaussing or impairing the isolation of the windings. The utilization and lifespan of electromechanical systems is thus limited in the presence of high ambient temperatures or vibrational strain, Wietschirke, v. Willich (1986).

Electromechanical drives can primarily be divided into translatory (electromagnets, linear motors) and rotary (electrical motors) transformers. The latter will usually be used in conjunction with a gear or feeding mechanism (actuation transformer) in order to generate a different rotary or linear motion. In the following, the different types of actuator drives will be discussed in detail. Based on the general description of electrical motors as well as their mathematical models in Chapter 5, here only the special properties for actuator drives are outlined and compared.

10.3.1 Electrical Motors

A large number of the actuators employed in mechatronic systems are electrical motors. The general structure of the different classes of electrical motors along with their primary features has already been presented in Chapter 5. In contrast to the electrical drives used in power-generating or power-consuming machines, the electrical motors employed as actuators are running neither continuously nor in a preferred rotational direction. Actuators are used as positioning devices, which results in requirements different from those for prime movers. These different requirements include the capability to sustain larger electrical and mechanical overloads, high positioning accuracy, very good dynamic behavior and therefore a small moment of inertia, low-wear generation of a certain holding torque, and a large region of acceptable rotational velocities. The power output ranges from a few watts for sub-miniature motors up to several kilowatts.

Since, for most applications, different types of motors can be taken into consideration, a comparison of the different constructional principles will be presented in the following. The interested reader is also referred to the survey articles by Jung, Schneider (1984), Weck (1989), (1990), Henneberger (1989), von Bechen (1989), and Janocha (1992) among others.

With regard to their key features, electrical motors can be divided into self-commutated and externally commutated machines, compare Section 5.7. For the former, the windings will be energized as controlled by the angular position of the rotor, whereas for the latter, the windings will be energized as determined by either the phase sequence of the feeding power supply system or an external control circuit.

a) Self-commutated motors

Self-commutated motors with a mechanical commutator are characterized by two features. First, the torque-generating stator field is created by permanent magnets or by a field coil. Secondly, the torque-producing current in the rotor windings is supplied to the rotor via brushes and a commutator. These brushes and the commutator form a switch, which toggles the current in the armature windings such that the torque is always created under the most favorable conditions. Table 10.4 presents a survey of the most common commutator motors with a power output of up to 1 kW, see also Chapter 5.

Mechanically commutated (brush) DC motors

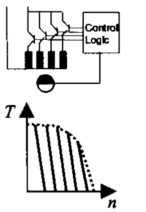
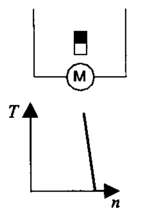
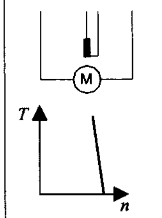
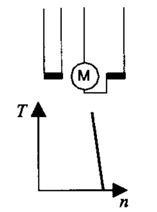
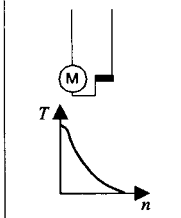
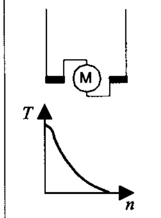
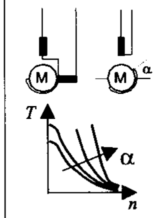
For permanently excited DC motors with mechanical commutation, the stator will consist of permanent magnets while the rotor is made up of a winding connected to a commutator. These motors typically have a voltage rating below 42 V and for battery-powered applications their use is widespread due to their high efficiency. These areas of application include road vehicles, household appliances, garden tools, medical and laboratory equipment as well as office equipment.

If the permanent magnets are replaced by a stator winding, the resulting motor will be an electrically excited DC motor. For series-wound motors, the stator winding and the armature winding are connected in series. The rotational velocity drops disproportionatly to the load torque, but, on the other hand, the break-away torque can be two to six times larger than the nominal torque. Since series-wound motors can also be supplied with alternating current, they are also referred to as *universal motors*. However, their efficiency is higher if they are fed with direct current.

For shunt-wound motors, the field coil and the armature coil are connected in parallel, which causes the rotational velocity-torque relation to be as stiff and as linear as that of the permanent-magnet excited machines. Shunt motors can only be fed with direct current. Operation with an AC supply is only possible if the voltage is rectified prior to being supplied to the motor.

Compound motors consist of both a shunt winding and a series winding, thereby combining the advantages of both types of winding, namely the high break-away torque and the stiff rotational velocity behavior. These motors can only be supplied with direct current and are quite expensive due to the elaborate stator winding.

Table 10.4. Survey of self-commutated motors with power ratings up to 1 kW

type of power supply		DC					AC (single phase)	
type of motor		brushless DC motor	permanent magnet motor	shunt-wound motor	compound motor	series-wound motor	universal motor	repulsion motor
type of commutation		electronically	mechanically	mechanically	mechanically	mechanically	mechanically	mechanically
basic circuit diagramm rotational speed versus torque								
nominal speed	rpm	< 60000	< 30000	< 12000	< 6000	< 15000	3000–30000	< 3000
power rating	W	1–1100	0-001–1000	0.2–1000	20–1000	8–1100	5–1000	< 500
voltage rating	V	< 400	< 250	< 600	< 220	< 600	< 220	< 220
efficiency	η	0.4–0.7	0.4–0.8	0.3–0.7	0.3–0.7	0.3–0.7	0.3–0.7	0.3–0.6
max. torque/ rated torque			< 10	< 6	< 6	< 5	< 5	< 2.5
speed control		electronically	series resistor or third brush		resistor in parallel with armature winding tapped winding series resistor phase control pulse width modulation			variation of brush angle α
legend		■□ permanent magnet stator ◓ permanent magnet rotor Ⓜ rotor with cage or commutator winding			○ reluctance rotor ○ hysteresis rotor transistor			

The key features of mechanically commutated DC motors are summarized in Table 10.5. The main disadvantages compared to other types of electrical drives emanate from the primary structure. The heat developing in the armature windings cannot be carried off well. Therefore, thermal aspects must be considered even if the motor is in overload for only a short amount of time. Furthermore, the mechanical commutation also limits the maximum motor current during standstill ("burn through") as well as at high rotational velocities ("brushfire"). Since the coal brushes wear out over time, low-cost drives require a certain amount of maintenance work.

Table 10.5. Characteristics of mechanically commutated DC motors

advantages	disadvantages
• good response characteristics; • small volume and weight; • good dynamic behavior; • very high degree of synchronism; • large speed range.	• wear of commutator and brushes (→ short lifetime →maintenance); • dynamics and standstill torque are restricted by commutator; • bad heat discharge.

Scope of application:
- small to medium actuating torque/force;
- precise positioning tasks;
- standard applications (areas of application include auxiliary drives for road vehicles, laboratory equipment, office equipment).

Mechanically commutated (brush) AC motors

In contrast to the repulsion motor, whose use is hampered by both the complicated speed control (by means of a variation of the angular position of the brush) and the limited speed range, the universal motor plays an important role as an electrical drive. Due to the high maximum rotational velocity, these motors can be designed such that they offer a good power-weight ratio. The speed control for universal motors is less complicated and allows for a wider range of rotational velocities than any other type of electrical motor. This makes the universal motor the predominant choice for household appliances and machines tools, Table 10.6.

Table 10.6. Characteristics of mechanically commutated AC motors

advantages	disadvantages
• good response characteristics; • very high power-weight ratio; • large speed range; • high maximum speed.	• wear of commutator and brushes (→ short lifetime →maintenance); • bad heat discharge.

Scope of application:
- small to medium actuating torque/force;
- standard applications (areas of application include household appliances and machine tool drives).

Electronically commutated (brushless) motor

Until a few years ago, mainly mechanically commutated DC motors were chosen as highly dynamic servo-drives. In the range of small to medium actuating powers, they are increasingly being replaced by electronically commutated brushless DC motors. These consist of a permanent magnet rotor, multiple stator windings and an electronic control circuit, which cyclically energizes the stator windings depending on the angular position of the rotor. This working principle makes these motors as robust and as quiet-running as asynchronous drives. Further advantages are the maintenance-free operation and the higher overload capacity due to the absence of a mechanical commutator, Table 10.7.

Table 10.7. Characteristics of electronically commutated DC motors

advantages	disadvantages
• as robust and as quiet-running as asynchronous drives; • good response characteristics due to linear current-torque relation; • high overload capacity; • maintenance-free; • smaller moment of inertia and better power-weight ratio than mechanically commutated DC motors.	• sensor system and complex control circuitry; • higher cost of system than for normal DC drive; • typically: restricted degree of synchronism (torque ripple).

Scope of application:
- small to medium actuating torque/force;
- high-value applications.

The big advantage of brushless motors is that the heat, solely originating in the stator windings, can be carried off well. This also results in a better performance-to-weight ratio compared to the corresponding traditionally commutated drives. On the other hand, these beneficial effects are at the expense of both more complex control circuitry and a more comprehensive sensor system. Moreover, the rotor speed and generated torque exhibit ripples, an effect that could be reduced by employing modern control concepts such as a sinusoidal current control, Wilke (1988). Due to these ripples, the degree of synchronism is limited, a problem that is especially severe at low rotational velocities.

Servo-motors

The lower end of the power range of servo-motors allows for the use of servo-motors as small actuators. These servo-motors are used for positioning tasks with a specified time-frame and accuracy. Quite often, several motors have to operate with a high degree of synchronism. These requirements are typical for machine tool drives, robots and flap- and valve-drives. Thus, restrictive requirements are imposed on the dynamics, positioning accuracy, maximum torque, degree of synchronism, and efficiency to name a few (see Section 10.2.3). In the lower power range, electrical servo-drives have prevailed against hydraulic and pneumatic actuators. The reasons have been listed in Table 10.8.

Table 10.8. Comparison of hydraulic and electrical servo-drives

hydraulic servo-drives	electrical servo-drives
• higher energy density; • lower power-weight ratio; • smaller installation space; • faster acceleration; • lower cost; • simple generation of a linear motion (cylinders); • very robust.	• better control behavior; • better accuracy; • higher efficiency; • simple diagnosis and maintenance; • matching to different applications quite simple; • electrical energy used for sensors, controller and actuator.

b) Externally commutated asynchronous drives

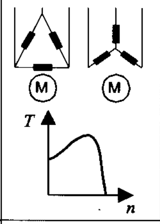
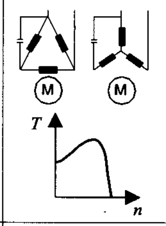
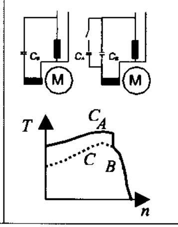
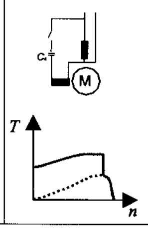
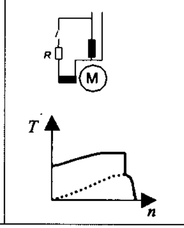
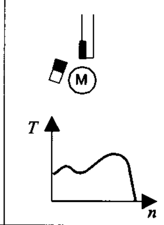
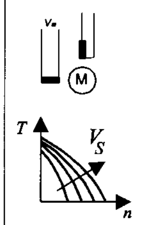
Asynchronous drives are employed for household appliances, garden tools, laboratory and medical equipment, and machine tool drives among other areas of application. A survey of commonly used asynchronous drives for small power output is shown in Table 10.9. The *three-phase asynchronous drive* exhibits the most favorable power-weight ratio of all asynchronous drives. This is exploited for drives for portable tools, where the motor is supplied with alternating current of a frequency of between 200 and 300Hz. Here, rotational velocities between 12000 and 18000rpm can be reached. This speed range is comparable to that of universal motors. It is possible to connect a three-phase asynchronous machine to a single-phase power supply system. In this case, two terminals are connected directly to the power supply system, whereas the third is connected to a capacitor which in turn is connected to the power supply system. This design is called a *three-phase capacitor motor*.

However, two windings are sufficient to generate a rotary field, provided they carry phase-shifted currents. This phase shift is generated with the help of a capacitor (*two-phase capacitor motor*), by increasing the ohmic resistance of one winding (*resistance-start AC induction motor*) or by shortening one of the windings. In the last case, an induced current will flow in the shortened winding (*shaded-pole motor*). *Resistance-start AC induction motors* are even more robust than capacitor motors and are typically used when the motor is switched on and off frequently (*e.g.*, cooling units). *Shaded-pole motors* are very cheap to build, but have a very bad efficiency and are solely used for very simple appliances.

c) Externally commutated synchronous motors

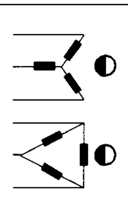
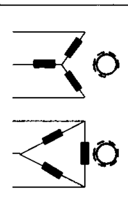
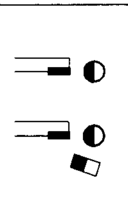
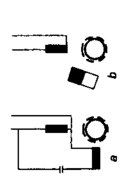
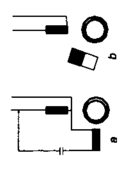
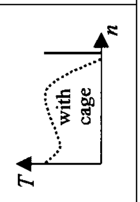
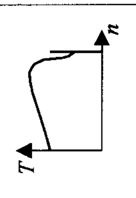
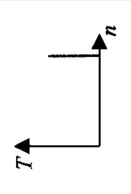
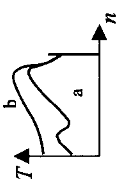
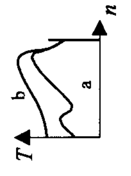
For synchronous motors, the stator field is generated the same way as for asynchronous motors. For the latter, the rotor always has a smaller rotational velocity than the rotary field. On the other hand, for synchronous motors, the rotor is turning at the synchronous speed of the rotating field. Due to this characteristic, synchronous motors are predominantly used as drives for high-precision speed control. Because of their simple design, they are also used in household appliances.

Table 10.9. Survey of externally commutated asynchronous motors with small power range

type of power supply		three-phase	single-phase					
type of motor		brushless DC motor	capacity motor C_A= starting capacitor C_B= running capacitor			resistance-start AC induction motor	shaded-pole motor	Ferraris motor
number of phases in the motor		3	3	2	2	2	2	2
basic circuit diagram rotational speed versus torque								
nominal speed	rpm	< 6000	<3600				3000–30000	< 3000
power rating	W	0.06–1100	0.2–1100				5–1000	< 500
voltage rating	V	12–800	0.2–500					
efficiency	η	0.5–0.8	0.3–0.7	0.4–0.7	0.4–0.7	0.3–0.7	0.05–0.4	0.2–0.5
break-away torque / rated torque		1–3	1–2	C_A: 2–4 C_B: 1–2	2–4	0.2–1	0.2–1	< 2
max. torque / rated torque		1.5–7	< 1.50	< 1.56	< 1.5	< 1.5	< 1.2	< 1.2
speed control		reduced stator voltage change of number of poles		reduced stator voltage change of number of poles tapped winding			choke coil	change of control voltage V_S
legend		■□ permanent magnet stator ◐ permanent magnet rotor Ⓜ rotor with cage or commutator winding				○ reluctance rotor ○ hysteresis rotor transistor		

Synchronous motors can be constructed with two or three independent coils. If connected to a single-phase power supply system, capacitors or resistors can be used to generate a phase-shifted current, see Table 10.10. These are the same working principles that are also used for single-phase asynchronous motors.

Table 10.10. Survey of externally commutated synchronous motors

type of power supply		three-phase alternating current		single-phase alternating current		
type of motor		permanent magnet motor	reluctance motor	permanent magnet motor	reluctance motor	hysteresis motor
number of phases in the motor		3	3	2	2	2
basic circuit diagram						
rotational speed versus torque						
nominal speed	rpm	< 33000		< 6600		
power rating	W	1–1100		0.01–1100		
voltage rating	V	1–800		12–500		
efficiency	η	0.3–0.6		< 0.05–0.6	< 0.05–0.6	< 0.05–0.4
break-away torque / rated torque		with cage < 3 without cage < 1	< 4	< 1	0.5–4	0.2–2
max. torque / rated torque		< 1.5	< 1.3	< 1.5	< 1.3	< 1.5
speed control		change of number of poles		choke coil		
legend		permanent magnet stator; permanent magnet rotor; (M) rotor with cage or commutator winding		reluctance rotor; hysteresis rotor; transistor		

Typical for synchronous motors is their run-up behavior. Permanently excited synchronous motors can only be run up by gradually increasing the frequency until the nominal frequency has been reached or by means of a pole-changed winding. In contrast to this, the hysteresis motor runs up autonomously. The magnetization of the hysteretic material is constantly reversed, thereby softly accelerating the rotor. For the reluctance motor, the magnetic resistance of the rotor (reluctance) changes along the circumference of the rotor according to the pole number. Table 10.11 lists some of the typical advantages and disadvantages of externally excited synchronous motors.

Table 10.11. Characteristics of synchronous motors

advantages	disadvantages
• cost-effective drive concept; • maintenance-free; • speed is in proportion to the power supply system's frequency.	• run-up issues; • stall under overload; • susceptible to oscillation build-up under load changes.
Scope of application:	• mostly small actuating torque/force; • standard tasks (*e.g.*, household appliances); • control engineering.

Many positioning tasks demand stepwise rotating motors instead of continuously turning motors. For small actuating power (< 500 W) *stepper motors* represent a cost-effective alternative compared with electronically commutated DC motors. Stepper motors also commutate electronically, but with a constant frequency determined by the external control logic. Stepper motors are thus synchronous motors and exhibit all their characteristic features.

There exists a vast number of stepper motor types, which in conjunction with the integration of control circuitry allows for an easy design of controlled positioning devices, see Section 5.3.7 and, *e.g.*, Traeger (1979), Kreuth (1988). The feedforward-controlled stepper motor is limited in its application as a positioning drive, since reliable operation necessitates good knowledge about the applied load. The assumption that load variations, break-away forces and vibrations do not cause stepping errors can only be justified up to certain maximum loads. To account for this, stepper motors are most of the times overrated. If stepping errors cannot be accepted, the motor must be controlled in closed-loop, Gfröer (1988). However, this annuls the principal advantages over other types of motors. It must also be pointed out that the efficiency of stepper motors is lower and that they are not capable of withstanding large overload. Also, the stepper motor must constantly be supplied with the maximum current, which means that the power consumption is unnecessarily high, Höfer (1991). However, the stepper motor offers a very high degree of reliability for a known load and maintenance-free operation and Table 10.12 shows some specifications for stepper motors. Table 10.13 summarizes the key aspects of stepper motors.

Table 10.12. Some characteristic values of stepper motors and linear motors

type of motor		stepper motor			linear motor / electrical cylinder		
current type		DC	AC, single pha-se	AC, three phase	DC	AC, single phase	AC, three phase
holding torque	Nm	0.001–30	0.2–7	0.,015–5	---	---	---
nominal torque	Nm	0.003–20	0.2–7	0.23–28	---	---	---
stepping angle	°	0.003 – 400	0.03–2	0.03–120	---	---	---
stepping frequency	Hz	0–250000	0–250000	0–400000	---	---	---
supply voltage	V	1–310	0–310	3–310	12–750		
efficiency							
power rating	W	< 500			up to 10000		
maximum stroke	mm	---	---	---	< 5000	linear motor: < 20000 electrical cylinder: < 5000	
maximum force	kN	---	---	---	linear motor: < 1000 electrical cylinder: < 600		
lifting speed	mm/s	---	---	---	linear motor: < 20000 electrical cylinder: < 2000		

Table 10.13. Characteristics of stepper motors

advantages	disadvantages
• direct digital control via integrated control circuitry and drives; • maintenance free; • cost-effective drive concept; • feedforward-controlled operation.	• load range must be known → overrating necessary; • rather small power density; • danger of stepping errors on feedforward-controlled operation; • comparatively small actuating dynamics.
Scope of application: • small actuating torque/force; • simple positioning task provided load is known.	

Subsequent mechanical converters (matching gears)

Electrical motors are primarily generating a rotary motion, which in general will not match the rotational velocity and torque requirements of the downstream actuator devices. In many cases, the motion must also be converted into linear motion. These demands can be satisfied by matching gears, which can perform the desired conversion. The use of these gears will also alter the system characteristics significantly. Details of these implications are pointed out in Table 10.14.

Among the gears employed in mechatronic systems, *gear drives* or *transmissions*, Tables 4.1 and 4.2, play a predominant role. Due to the form-locked force transmission, they are predestined for large power/large torque applications. Gear drives are built as gears with fixed ratio or stepwise variable ratio. They can also be combined with continuously changeable gearboxes.

Table 10.14. Characteristics of mechanical matching gears

advantages	disadvantages
• match rotary motion to load and rotational velocity requirements; • mediate generation of linear motion; • self-locking gears reduce the power consumption during holding phases; • expands connection possibilities.	• time-invariant friction and gear play (backlash); • increased total moment of inertia; • power loss in gear necessitates stronger drive and reduced overall efficiency; • increases overall size.

The most simple design of a gear drive is the *spur gear*. This gear type is characterized by the teeth, which are perpendicular to the face of the gear. The two meshed gears are termed gear and pinion. The shafts are supported by friction bearings or rolling bearings. A V drive, that is, a gear where the driving shaft and the driven shaft intersect, is typically realized as a *bevel gear* or *worm gear*. The latter have a higher share of viscous friction than spur gears or bevel gears. Due to this, they are quiet-running, but they also operate at a lower efficiency and will thus heat up more extensively for the same power throughput. One important property of worm gears is the irreversibility. When a worm gear is driven, the meshing spur gear will turn, but turning the spur gear will not turn the worm gear. Thus, the resulting gear is self-locking.

A special form of the spur gear is the *planetary gear*, which offers one additional degree of freedom. It is in general constructed such that a central sun gear is surrounded by a number of planetary gears, see Table 4.1. These are mounted on a plate, the planet carrier. This carrier has the same center of rotation as the sun gear. All the planetary gears also engage the inside of the ring. The ring also has the same center of rotation as the sun and the planet carrier. The input and the output of this gear can be chosen arbitrarily, thus different gear ratios can be realized with the same gear.

If a very high gear ratio must be realized with few gearbox stages, a special and quite expensive design, the *harmonic drive*, comes into play. In general, these gears consist of an internal gear and an eccentrical positioned gear. This gear can rotate freely and is mounted on an eccentric carrier, which itself is mounted coaxially with the internal gear. Despite the high gear ratio, which is determined by the difference in teeth of the two gear wheels, harmonic drives are compactly built and have a coaxial arrangement of the driving and the driven shafts. The rotational motion of the driven shaft is homogenous.

One application that is quite common is the conversion of rotary into translatory motion. This task is accomplished by *straight gears*, which include *rack and pinion gears*, *toothed belt drives* and the *recirculating ball spindle*.

Transmission belts are used if a certain distance between the driving shaft and the driven shaft must be bridged. They can also be used to drive multiple shafts in the same or even in the opposite sense of rotation (*e.g.*, the internal combustion engine). They outperform gear drives with their simple and thus cheap design and only a little amount of maintenance is required.

Hydraulic transmissions can be divided into *hydrostatic* and *hydrodynamic transmissions*. The former are arbitrarily changeable in their gear ratio, whereas for the latter, the gear ratio will usually depend on the load and torque. Hydrostatic transmissions are composed of a pump P, a hydraulic motor M, and some sort of control unit, see Table 10.22. This transmission system can be controlled by either manipulating the geometry of the pump or of the motor (primary or secondary control) or by controlling both. Compared to continuously changeable mechanical gears, they have a lower efficiency, but offer tremendous overload capacities and a high power density.

Hydrodynamic transmissions are based on the Föttinger principle. In contrast to the original design, they nowadays consist of an additional reactor that can support part of the torque delivered, thereby allowing adjustment of the torque transmitted through the unit. Different geometries of the impeller I, reactor R and turbine T allow the overall behavior adjustment of the hydrostatic transmission, see Table 4.2.

10.3.2 Electromagnets

Electromagnets are electrical-magnetic-mechanical converters. They convert electrical energy into mechanical energy via the intermediate form of magnetic energy. They are paramountly used for the generation of limited, alternating and latching, translatory or rotary motion. The latching positions correspond to the stable positions, the magnet settles in if the coil is energized and switched off. Because of the energy conversion principle of electromagnets, the reluctance principle, it is characteristic for electromagnets that the magnet armature can only be displaced in one direction (active direction of motion). An external force (*e.g.*, spring load) must be used to return the magnet armature to its initial position if the magnet is switched off, see Chapter 5.

Because of their operating principle, electromagnets are directly acting actuators, which means that there is no mechanical converter or transmission element. This also requires that the electromagnet's specifications are matched to each particular application. The best adaptation can be accomplished if the electromagnet and the actuated element form one functional unit (*e.g.*, magnetic valves, relay, magnetic clutch). Hereby, elements arise that offer superior characteristics (smaller unit volume, faster switching times, longer lifespan) and at the same time allow for cost savings. The vast majority of electromagnets that are produced today are specialized designs. A classification scheme for electromagnets is suggested in Table 10.15. For this scheme, the electromagnet has been considered both as a separate unit and as part of a functional group.

Table 10.15. Classification criteria for electromagnets, Kallenbach *et al.* (1994)

function	actuating magnets (operating magnets)	lifting magnet pull-type electromagnet rotary magnet valve magnets: – switching valve magnets – proportional valve magnets
	holding magnets (armature-free magnets)	clamping magnet lifting magnet magnetic separator
	force transmission magnets	electromagnetic clutch electromagnetic brake
iron	basic shape	pot shape (axially symmetric) U-shape E-shape
	assembly	solid laminated sintered
type of excitation	mode of operation power-on time	AC DC DC with permanent pre-magnetization
field coil	shape	cylindrical coil rectangular coil shaped coil
	assembly	wire-wound coil: – orthocyclical-wound – random-wound thin film coil
	number of coils	single coil multiple coils
magnet gap	position of magnet gap	within coil outside coil
	number of gaps	one magnet gap multiple magnet gaps
	size	short stroke magnet medium stroke magnet long stroke magnet
type of motion	translatory	pull-type push-type alternating
	rotary	single-acting double-acting double-acting with center position

Further adaption possibilities arise from the application of electronic driver circuits, by which the static and dynamic behavior of electromagnets can be influenced to a high degree. These circuits allow for the adjustment of the non-linear magnetic-force curve, Raab (1993), compensation of friction, Maron (1990), and fast positioning even under the demand for relatively large actuation forces, all by means of optimal control strategies, see Section 5.2.

10.4 HYDRAULIC ACTUATORS

The next two sections discuss hydraulic and pneumatic actuators, which both belong to the class of *fluidic actuators*. They are characterized by their rugged design and their high power-to-weight ratio. In addition, they also offer the opportunity to generate linear motion easily and directly by employing hydraulic cylinders and diaphragm drives. Further, fluidic actuators can be designed such that they only dissipate energy during dynamic operation. Thus, during static phases, high reaction forces can be generated with little energy consumption. The ongoing fusion of fluidic and (micro-)electronic components ("fluidtronics") allows the design of actuators that offer both precise positioning and fast dynamics, Backé (1992), Anders (1986).

Besides these advantages, there are also some drawbacks. The efficiency of the overall systems is below that of electrical drives and the required auxiliary energy may not be available. Furthermore, the maximum positioning accuracy is limited to some 10 μm. The major advantages and disadvantages are summarized in Table 10.16.

Table 10.16. Characteristics of fluidic actuators

advantages	disadvantages
• high actuating force; • large displacement; • high power density; • direct generation of linear motion; • no power consumption during standstill; • rugged design.	• (additional) auxiliary power unit necessary; • requires complex system structure; • partly expensive servo-components (*e.g.*, valves) • limited positioning accuracy; • noise.

Table 10.17 shows a classification of fluidic actuators. Mainly, they can be divided into two groups. The first comprises the pneumatic actuators and is further divided into pressure actuators and vacuum actuators. The second group contains the hydraulic actuators with the typical translational and rotary drives.

Table 10.17. Overview of important fluidic actuators (standard: translatory motion, italic: rotary, angular motion)

force generation	energy conversion	implementation
pneumatic	pressure actuator	cylinder diaphragm drive *air motor*
	vacuum actuator	diaphragm drive
hydraulic	pressure actuator	cylinder *hydraulic motor*

10.4.1 Hydraulic Actuation Systems

Hydraulic actuators are usually applied to very large actuation forces and actuation power respectively. However, they require a special hydraulic power supply. Therefore, they are an interesting option if several servo-axes for medium actuation forces can be replaced simultaneously. Hydraulic actuators outperform comparable electrical drives with respect to their system dynamics. The fact that the moving mass of hydraulic actuators is rather small despite the large actuation power facilitates dynamically fast positioning. Hydraulic actuators surpass pneumatic actuators in their high stiffness and shock resistance. The major advantages and disadvantages of hydraulic actuators are listed in Table 10.18.

Table 10.18. Characteristics of hydraulic actuators

advantages	disadvantages
• small dimensions; • fast dynamics and high power density; • high stiffness; • high working capacity.	• possibly high system cost; • two line system; • possibly oil conditioning necessary; • friction and complex dynamics exacerbate control.
scope of application:	• medium to high actuating force; • medium to high displacement range; • limited space available; • highly dynamic systems.

The dynamic behavior of hydraulic control circuits is, above all, characterized by the weak damping, which is furthermore dependent on both the piston displacement and the external load. Nevertheless, in conjunction with modern control concepts, one can design servo-hydraulic actuators that offer both good positioning accuracy and good dynamic behavior, Saffee (1986), Scheffel (1989), Glotzbach (1996).

Figure 10.10 shows a typical hydraulic actuator for linear motion. The hydraulic power supply consists of an oil tank and an electrically-driven piston pump connected to a proportional acting electro-hydraulic servo-valve. Depending on the position of the servo-valve, the cylinder moves with the corresponding direction and speed. The connection pipe between the pump and the servo-valve usually contains a directional (non-return) valve, avoiding back-flow of the oil, a pressure relief valve as a safeguard against excessive pressures, and a hydraulic accumulator to dampen pressure oscillations from the pump.

The next section will describe the principles and modeling of these hydraulic components. At the end of this section, a hydraulic servo-axis will be considered in detail.

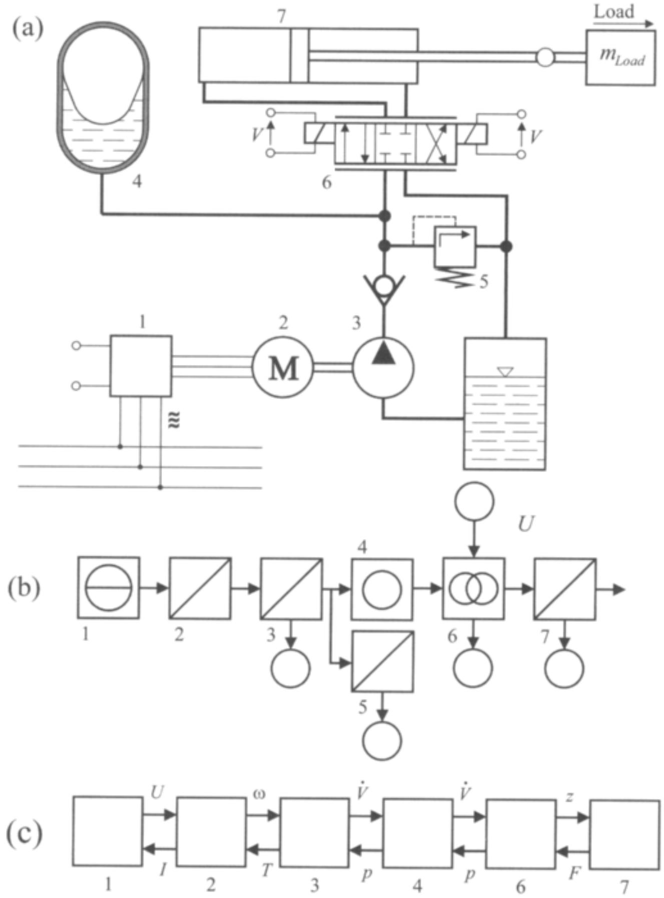


Figure 10.10. Hydraulic actuator for linear motion with power supply: (a) schematic; (b) energy flow scheme; (c) two-port representation. 1 power electronics; 2 AC motor; 3 radial piston pump; 4 accumulator; 5 check valve; 6 proportional valve; 7 cylinder

10.4.2 Hydraulic Components and their Models

For modeling of hydraulic components, a short summary of fluid dynamic principles is presented. Then, some material properties of the hydraulic fluid will be discussed. Next, hydraulic components will be considered. These include solenoid valves, as well as locking valves, throttle valves and check valves. Further on, hydraulic transmission lines, accumulators and motors are considered.

a) Some fluid dynamic properties

Hydraulic systems use fluids to transmit power between different components. In this section, some laws are considered for hydraulic power systems in the one-dimensional form.

The *mass balance* of a fluid mass $m_s = \rho V_S$ becomes, according to (2.3.4)

$$\frac{\partial}{\partial t}m_s(t)=\frac{\partial}{\partial t}\left[V_s(t)\rho(t)\right]=\frac{\partial V_s(t)}{\partial t}\rho(t)+\frac{\partial\rho(t)}{\partial t}V_S(t)=\sum_i\rho_i\dot{V}_i(t) \quad (10.4.1)$$

where $\dot{V}_i(t)$ are the inward volume flows across the volume surface. If the density can be assumed to be constant, the mass balance reads

$$\frac{\partial V_s(t)}{\partial t}=-\sum_i\dot{V}_i(t) \quad (10.4.2)$$

and if the volume is constant, it holds that

$$\frac{\partial\rho(t)}{\partial t}V_S=-\sum_i\rho_i\dot{V}_i(t) \quad (10.4.3)$$

Applying the momentum theorem to a fluid element of length dz, constant cross-sectional area A_F, density $\rho(z,t)$ and friction force $F_f(z,t)$leads to

$$F_1(t)+g\,dm\sin\alpha-dF_f-F_2(t)=\frac{d}{dt}\left(A_F\rho\,v_z dz\right) \quad (10.4.4)$$

See Figure 10.11.

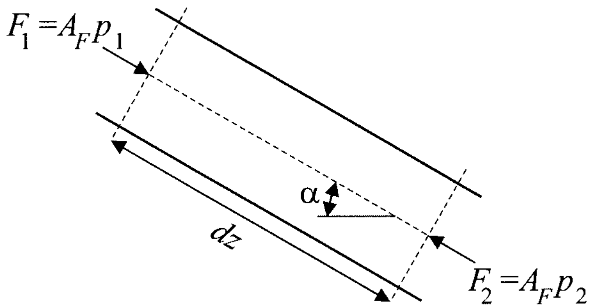


Figure 10.11. Fluid element in a tube

Using the pressure, this equation becomes

$$p_1(t)+g\rho\,dz\sin\alpha-\frac{1}{A_F}dF_f-\left(p_1(t)+\frac{\partial p}{\partial z}dz\right)=\frac{d}{dt}\left(\rho\,v_z dz\right)$$

$$-\frac{\partial p}{\partial z}+\rho g\sin\alpha-\frac{1}{A_F}\frac{\partial F_f}{\partial z}=v_z\left(\frac{\partial\rho}{\partial t}+v_z\frac{\partial\rho}{\partial z}\right)+\rho\left(\frac{\partial v_z}{\partial t}+v_z\frac{\partial v_z}{\partial z}\right)$$
$$=\frac{D(\rho v_z)}{Dt} \quad (10.4.5)$$

For laminar flow, this is known as the *Navier-Stokes equation*. This equation can be simplified for an ideal fluid with ρ = const. and $\alpha = 0$

$$\frac{Dv(z,t)}{Dt}=\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial z}=-\frac{1}{\rho}\frac{\partial p}{\partial z} \quad (10.4.6)$$

which is known as the *Euler differential equation.*

Now, a prismative fluid column of length l is considered and lumped parameters are assumed (like a rigid body). Then, it follows that

$$\rho\, l \frac{dv(t)}{dt} = -\Delta p(t) = -\left(p_1(t) - p_2(t)\right) \tag{10.4.7}$$

If the flow is stationary and the fluid incompressible, see Figure 10.12, integration of (10.4.6) leads to Bernoulli's equation

$$\left(p_1 + \frac{\rho v_1^2}{2}\right) - \left(p_2 + \frac{\rho v_2^2}{2}\right) = \Delta p_{Loss} \tag{10.4.8}$$

where Δp_{Loss} denotes the pressure loss due to friction effects.

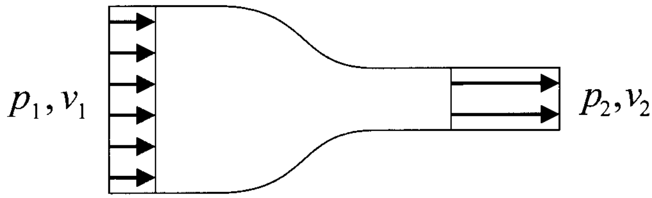


Figure 10.12. Fluid with changing tube area (contraction)

With regard to the friction in the pipe, two different types of flow can be distinguished, laminar and turbulent flow, see also Section 2.3.

Laminar flow is characterized by an orderly movement of the individual fluid layers, which slide on one another. Laminar flow losses can be described by the linear relationship

$$\Delta p = \frac{1}{G}\dot{V} = f\dot{V} \tag{10.4.9}$$

compare (2.3.62). G is termed the hydraulic conductance and f the friction factor. For a pipe with a circular cross-section, G evaluates to

$$G = \frac{\pi D^4}{128 \nu \rho\, l} \tag{10.4.10}$$

which can be derived from the circular Poiseuille flow. In this equation, l denotes the length of the pipe with diameter D. For a rectangular cross-section, the hydraulic conductance is given by

$$G = \frac{b h^3}{10 \rho\, \nu l}\left[1 - \frac{192 h}{\pi^5 b}\tanh\frac{\pi b}{2h}\right] \tag{10.4.11}$$

where b and h describe the geometry of the rectangle. Other hydraulic conductances can be found in Merrit (1967).

Turbulent flow is composed of both a general movement in direction of the pipe and a superposed, inordinate, orthogonal movement. Therefore, the path of fluid particles is not known in advance for a turbulent flow field. The motion of individual particles is random and can only be described by statistical parameters such as the mean velocity $\bar{v}$. Experiments show that the pressure loss is

$$\Delta p = f \frac{l}{D} \frac{\rho \bar{v}^2}{2} \tag{10.4.12}$$

In general, the friction factor f depends on the surface roughness and the Reynolds number Re, which is defined as

$$Re = \frac{D\bar{v}}{\nu} \tag{10.4.13}$$

where ν_f is the kinematic viscosity.

For smooth pipes and $Re \leq 80000$, *Blasius' law* can be used to calculate the friction factor as

$$f = \frac{0.3164}{Re^{0.25}} \tag{10.4.14}$$

Laminar flow can also be covered by (10.4.12) if the friction factor for the laminar regime is set to

$$f = \frac{64}{Re} \tag{10.4.15}$$

In the transition area between laminar and turbulent flow, a third relationship is defined to avoid discontinuities. One possible set of equations is given as

$$f = \left\{ \begin{array}{llll} \dfrac{64}{Re} & \text{for} & Re < 1404 & \text{(Poiseulle)} \\ 0.0456 & \text{for} & 1404 \leq Re < 2320 & \text{(transition)} \\ \dfrac{0.3164}{Re^{0.25}} & \text{for} & 2320 \leq Re & \text{(Blasius)} \end{array} \right\} \tag{10.4.16}$$

For lumped losses, it is reasonable to assume turbulent flow in most cases, also because of elements with higher resistances like orifices or tube bends.

For turbulent flow, the pressure loss-volume flow rate relation at the orifices is given by

$$\dot{V} = \alpha_D A \sqrt{\frac{2}{\rho}} \sqrt{|\Delta p|} \operatorname{sign}(\Delta p) \tag{10.4.17}$$

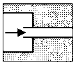
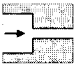
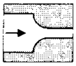
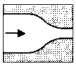
where α_D incorporates the contraction of the jet upon passing through the orifice. The point along the jet where the cross-sectional area of the jet reaches its minimum is called *vena contracta*. α_D and A are combined into one parameter, which is obtained by experiments.

Turbulent flow losses are also evoked by sudden changes in the flow's cross-sectional area, by bends and other obstacles. For these cases, the pressure loss is defined by

$$\Delta p = \zeta \frac{\rho}{2}\left(\frac{\dot{V}}{A}\right)^2 \tag{10.4.18}$$

where ζ is the loss coefficient, which depends on the geometry, see Table 10.19.

Table 10.19. Loss coefficients for typical orifices, Merrit (1967)

	type of entrance	ζ
	inward projecting pipe entrance	1 - 1.3
	sharp edged entrance	0.3 - 0.4
	rounded entrance	0.012 - 0.2
	perfectly rounded entrance	approximately 0

Material properties of the hydraulic oil and surrounding vessels play an important role in the overall performance of hydraulic control systems. The equation of state relates by definition the pressure, temperature and density of any solid, liquid or gaseous phase. While the equation of state for an ideal gas can be derived by applying first principles, it is not possible to do so for a liquid fluid. In general, the density increases with an increase in pressure or a decrease in temperature. The density increase due to pressure, the so-termed compressibility, is more than 100 times larger than that of steel. Therefore, this effect cannot be neglected. The *bulk modulus* or *compressibility module* is defined as

$$\beta = -V\left(\frac{\partial p}{\partial V}\right)_{T=\text{const.}} \tag{10.4.19}$$

The compressibility κ of a liquid is defined by

$$\kappa = \frac{1}{\beta} = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_{T=\text{const.}} \tag{10.4.20}$$

The relative change of length l of a liquid column becomes, for constant cross-sectional area A

$$\frac{dl}{l} = -\frac{dV}{V} = \frac{1}{\beta}dp = \frac{1}{\beta}\frac{dF}{A} \tag{10.4.21}$$

This corresponds to the elasticity ε of a solid material and a comparison with (4.1.1) and (4.1.3) shows that β is equivalent to the modulus of elasticity E. Therefore, the stiffness of a liquid column of length l becomes,

$$c_l = \frac{dF}{dl} = \beta \frac{A}{l} \tag{10.4.22}$$

The compressibility is not only affected by the fluid itself. It is also altered by air trapped in the liquid. Further, if thin-walled vessels are pressurized, then these vessels will expand as their walls give way. This increases the volume that the enclosed liquid can fill and thus causes a decrease in both pressure and density of the encompassed fluid. This interaction also effectively changes the bulk modulus. All these influences are combined into one parameter, the effective bulk modulus, usually denoted by β'.

One point that divides solids from fluids is the fact that the latter deform unlimitedly if subject to a shear force. Different material-specific laws have been formulated to capture this behavior. Hydraulic fluids are most often modeled as Newtonian fluids, for which the shear stress on the surfaces is in proportion to the velocity gradient

$$\tau = \mu \frac{\partial v}{\partial y} \tag{10.4.23}$$

with the absolute viscosity μ being the proportional constant. Depending on the experimental set-up used to measure the viscosity, two different parameters can be acquired: The absolute viscosity μ and the kinematic viscosity ν_f. They can be converted into one another by

$$\nu_f = \frac{\mu}{\rho} \tag{10.4.24}$$

Some properties of standard hydraulic oil have been collected in Table 10.20 to give an idea of the magnitude of the respective values.

Table 10.20. Characteristic properties of standard hydraulic oil

variable	density	specific volume	specific heat	bulk modulus	kinematic viscosity	thermal expansion coefficient	vapor pressure
symbol		v	c_p	β	ν_f	ε_T	p
dimension	$\frac{kg}{m^3}$	$\frac{m^3}{kg}$	$\frac{J}{kgK}$	$\frac{N}{m^2}$	$\frac{mm}{s^2}$	$\frac{1}{K}$	bar
defined at	15°C	15°C	20°C		40°C		50°C
value	870	$1.15\cdot10^{-3}$	1885	$2\cdot10^{9}$	15...70	$7\cdot10^{-4}$	37536

Hydraulic control systems consist of different components, which will now be modeled. As has already been stated in Chapter 2 (Table 2.3), the pressure p (effort variable) and volume flow rate $\dot{V}$ (flow variable) are usually chosen as power variables for hydraulic systems.

b) Hydraulic control valves

For manipulating the oil flow in tubes, hydraulic control valves are used. Based on their task, valves can be divided into four classes: solenoid valves, locking valves, pressure control valves and flow valves.

A schematic view of a spool valve is shown in Figure 10.13. Here, the spool is driven by a combination of two electromagnets, each coil being responsible for displacing the spool in one direction.

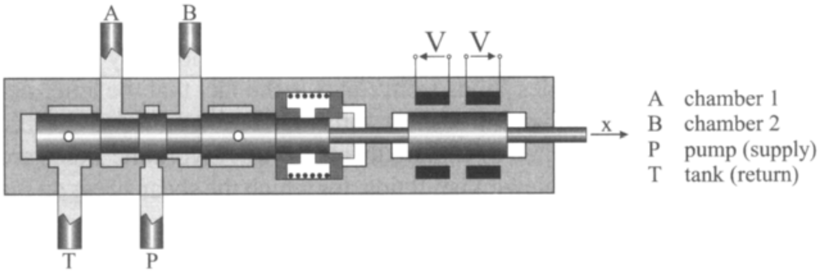


Figure 10.13. Schematic view of a hydraulic spool valve

In proportional valves, the spool is driven by one or more electromagnets. Electromagnets have been discussed in detail in Section 5.2. Here, only the most important equations will be repeated. The electromagnet can be modeled using (5.2.42)

$$V(t) = RI(t) + L_d \frac{dI(t)}{dt} + c_Y \frac{dY}{dt} \tag{10.4.25}$$

The force exerted by the electromagnet is given as (5.2.43)

$$F_M(t) = c_Y I(t) \tag{10.4.26}$$

The position dependency is neglected. Under this assumption, one can combine the equations (10.4.20) and (10.4.21), yielding

$$\dot{F}_M(t) + \alpha_M F_M(t) = k_M V(t) \tag{10.4.27}$$

The quantity V denotes the terminal voltage, α_M and k_M designate the time constant and gain respectively. This magnet exerts a force F_M on the spool. The spool is modeled as a second order system

$$m_s \ddot{z}(t) + d_s \dot{z}(t) + c_s z(t) = F_M(t) + F_{Ext}(t) \tag{10.4.28}$$

where the coefficients m_s, d_s and c_s denote the mass, damping and spring constant respectively, and F_{Ext} is the resultant of all external forces acting on the spool. These forces could stem from friction as well as from the fluid flow inside the valve. Most often, these forces are neglected, mainly because their influence is relatively small.

Flow through the valve orifices is modeled as turbulent flow, described by (10.4.15). The area of the orifice is a function of the spool displacement and is given by

$$A = \begin{cases} A_V'(z - z_0) & \text{for} \quad z \geq z_0 \\ 0 & \text{for} \quad z < z_0 \end{cases} \tag{10.4.29}$$

A_V' denotes the cross-sectional area of the valve opening per unit displacement of the spool and is called the opening or area gradient of the valve. The function in (10.4.24) depends on the geometry of the valve and will vary with different designs. The parameter z_0 is determined by the valve lapping, *i.e.*, $z_0 < 0$ for an underlapped valve, $z_0 = 0$ for a zero-lapped valve and $z_0 > 0$ for an overlapped valve. Figure 10.14 presents the different configurations along with the characteristic curves for the flow. For the underlapped valve, two flow paths can be open at the same time. This is indicated by the characteristic curve having two branches.

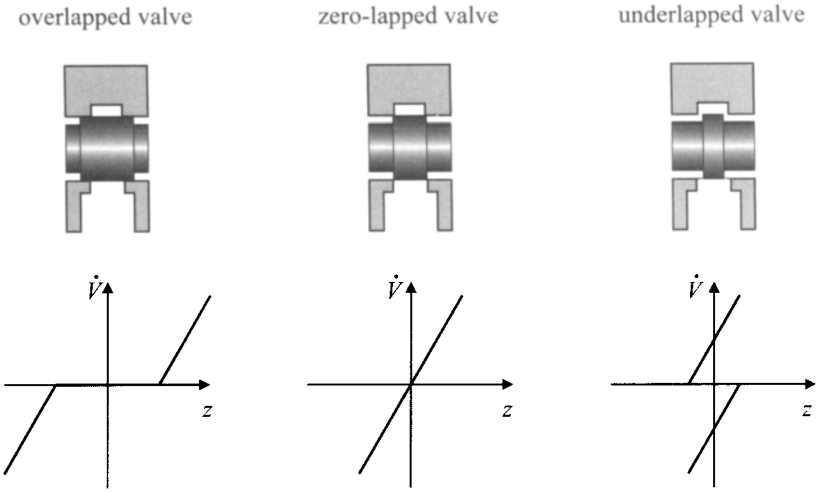


Figure 10.14. Spool displacement flow curves for differently lapped proportional valves

There exists a wide variety of different control valves, differing in the number of hydraulic connectors, arrangement of the flow paths and actuation principle. Some examples are given in Figure 10.15 along with the graphical representation used in hydraulic circuit diagrams. In the case of switching valves, a segment is used for each switching position. Finite switching valves are specified in terms of the number of connections and the number of positions they have. Thus, a 4/2 valve is one with four ports and two positions. Two-way valves can be constructed as seat valves, which offer superior (leakage-free) sealing. The other valves are typically constructed as sliding valves, with the spool type being the predominant construction principle.

Two-way valves are also manufactured with internal, mechanical control circuits, controlling the pressure or flow through the valve. Examples for these types of valves are given in Table 10.21. The locking valves are constructed as seating valves with a ball being used to tightly seal the opening. The check valves shown in Table 10.21 are constructed as spool-type valves. The simple non-return valve can be used to bypass a throttle valve in one flow direction. Thus, the dynamics of the pressure build-up and pressure release can be influenced independently. This technique is exploited for hydraulic braking systems of road vehicles. Pilot-operated non-return valves are typically used in conjunction with solenoid valves, where they tightly seal the cylinder chambers whenever the spool is in its center position. Throttle valves change the pressure-flow relation as they dissipate hydraulic power. Pressure relief valves are safety devices that limit the maximum level of the hydraulic pressure in a system. Their construction is almost identical to that of sequence valves. The latter only pressurize their secondary hydraulic port if the pressure at the primary port exceeds a certain threshold. The pressure threshold can also be set by means of an external control pressure. Other valves are used to maintain a certain pressure ratio between the primary and the secondary port. The functional principle is the force balance at the piston. The piston is constructed such that the areas on which the primary and secondary pressure act are different. Since the force is the product of pressure and area, the pressure ratio required for a balance of forces will be the reciprocal of the active piston areas.

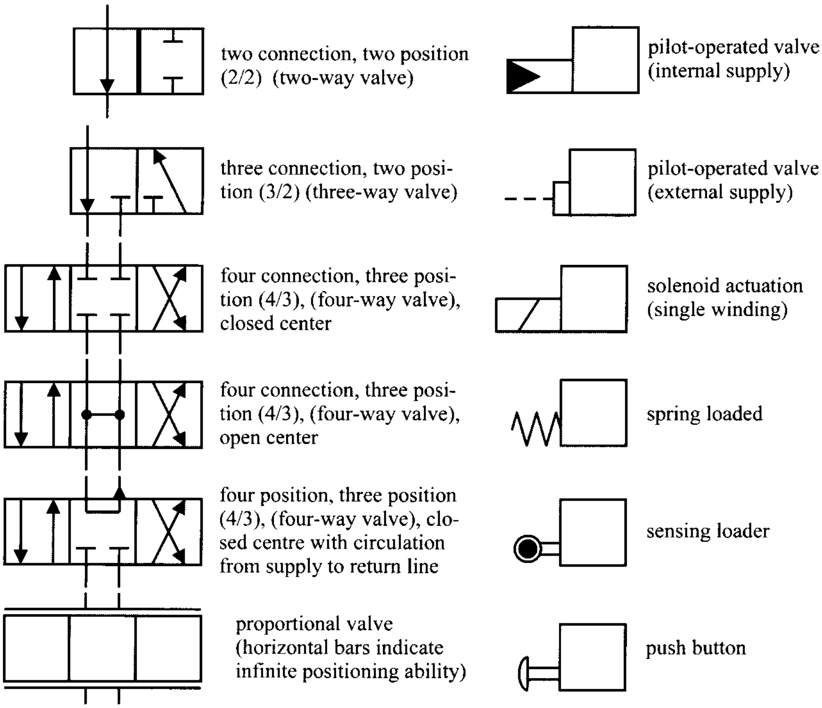
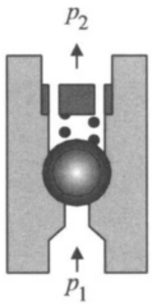
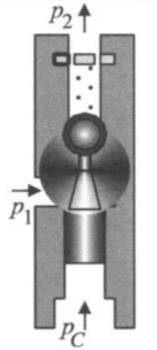
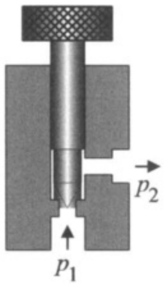
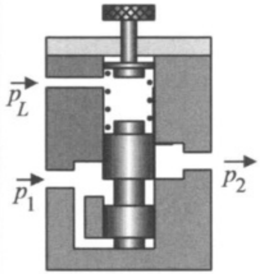
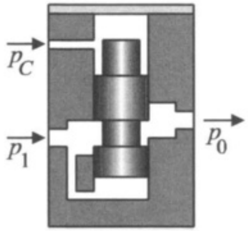
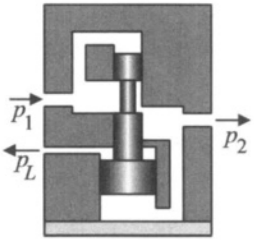
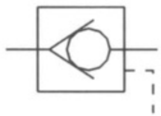
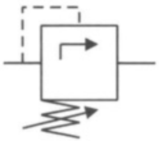
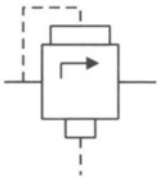
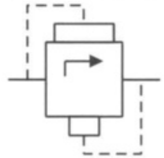


Figure 10.15. Valve types and symbols: a segment is used for each switching position

Table 10.21. Construction and symbols of different hydraulic valves

class	locking valves		throttle valves	check valves		
type	simple non-return valve	pilot-operated non-return valve	needle valve	pressure relief valve/sequence valve	externally controlled pressure control valve	counterbalance valve with fixed pressure ratio
constructional principle						
graphical representation						
function	blocks flow if $p_2 > p_1$. Because of the spring load, the valve only opens if the pressure difference $p_1 - p_2$ is above 0.5 ... 4 bar	same as simple check valve, but has a hydraulic control port to open the valve even if $p_2 > p_1$	flow-dependent pressure reduction, energy dissipation	opens if pressure p_1 exceeds a certain pressure level as controlled by the spring load	reduces the pressure such that p_1 is equal to or less than the control pressure p_C	keeps the pressure ratio of p_1 / p_2 constant

The dynamics of the internal mechanical control loops are typically much faster than those of the other hydraulic components used in the hydraulic system. Thus, one does not need to model the individual components of the internal control loop, such as masses, springs, pistons, *etc.* Instead, these valves are typically represented by a static model that describes the pressure-flow characteristic of the valve.

c) Hydraulic transmission lines

Depending on the length of the pipe, different effects must be modeled. Very short pipes are usually not modeled, especially if they have a relatively large cross-sectional area. For longer pipes, the compressibility must be taken into account. The inertia of the fluid has to be considered if the fluid is subject to an oscillatory excitation or if segments of the line have a small cross-sectional area. The flow resistance plays an important role for long lines or those with small cross-sectional areas.

A section of a prismatic tube of length l and cross-sectional area A has the mass

$$m_l = A l \rho \tag{10.4.30}$$

It is assumed that the compressibility of the fluid requires a force

$$F_c = c_l (z_1 - z_2) \tag{10.4.31}$$

with c_l being the oil-spring constant and z_1 and z_2 being the displacement of the two ends, see Figure 10.16. The damping force through liquid motion is

$$F_d = d v \tag{10.4.32}$$

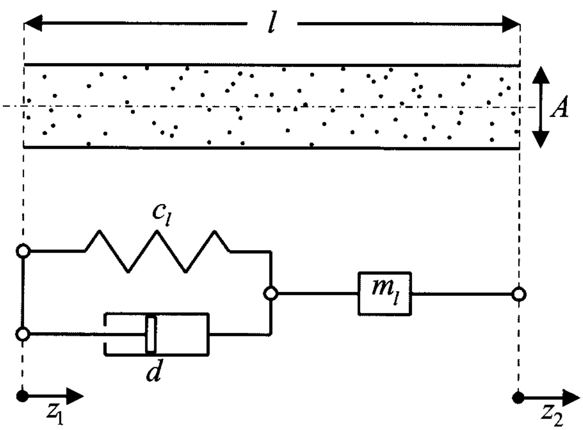


Figure 10.16. Mechanical equivalent model of a hydraulic transmission tube section

Applying the law of momentum yields

$$m_l \ddot{z}_2(t) + d \dot{z}_2(t) + c_l z_2(t) = c_l z_1(t) \tag{10.4.33}$$

or in terms of volume V and pressure p

$$m_l \ddot{V}_2(t) + d\dot{V}_2(t) + c_l V_2(t) = A^2 \Delta p_1(t) \tag{10.4.34}$$

Hence, the second order oscillatory model of an oil column shows the following characteristic values, compare (4.5.5)

$$\omega_0 = \sqrt{\frac{c_l}{m}} = \frac{1}{l}\sqrt{\frac{\beta}{\rho}} \quad \text{(undamped natural frequency)} \tag{10.4.35}$$

$$\zeta = \frac{d}{2\sqrt{c_l m}} = \frac{d}{2A\sqrt{\beta\,\rho}} \quad \text{(damping factor)} \tag{10.4.36}$$

d) Hydraulic pressure storages (hydraulic accumulators)

The task of hydraulic accumulators is to ingest a certain quantity of pressurized hydraulic fluid, which can be fed back into the hydraulic circuit. These storage devices can supply hydraulic fluid to satisfy the peak demand, compensate for temperature-related variations, supply energy during emergency conditions (power loss) or dampen pump-induced pressure oscillations.

There are three basic constructional principles, *piston- type accumulators*, *bladder-type accumulators* and *diaphragm-type accumulators*, see Figure 10.17. For the piston type accumulator, two chambers are separated by a piston assembly. One chamber is filled with a gas, typically nitrogen (N_2). The other chamber is connected to the hydraulic system and thus filled with oil. This accumulator type is primarily used if a high storage capacity at a high system pressure is required. Due to the comparably large mass of the piston, the piston-type accumulator is much slower in its dynamics than the bladder and diaphragm type accumulators. The latter use elastomeric elements to separate gas and liquid. Piston-type accumulators seal the two chambers in an almost hermetical way, whereas the other two designs are impaired in that the nitrogen may diffuse through the divider element over time.

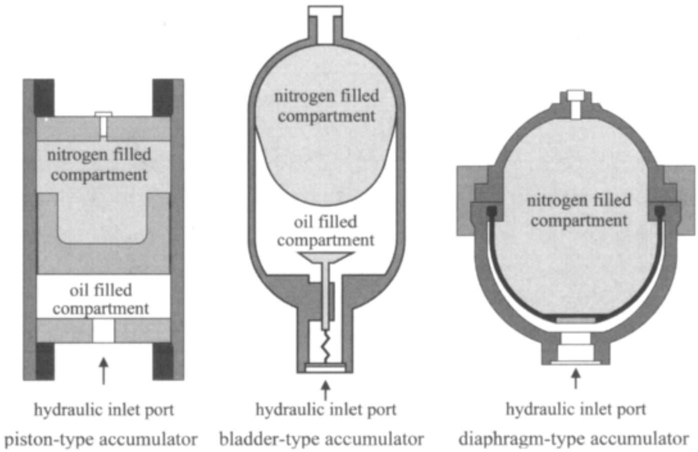


Figure 10.17. Hydraulic accumulators

The behavior of these fluid storage devices is mainly determined by the behavior of the charge gas. For simple calculations, nitrogen can be treated as an ideal gas, thus obeying the equation of state of the ideal gas, given as

$$pV = mRT \tag{10.4.37}$$

This equation models the behavior of the charge gas sufficiently precisely for system pressures up to 10 bar. For higher pressures, one must use experimentally derived equations, such as the Beattie and Bridgman equation, *e.g.*, Korkmaz (1982), see Section 10.5,

$$p = \frac{RT(1-\varepsilon)}{V^2}(V+B) - \frac{A}{V^2}$$
$$A = A_0\left(1 - \frac{\alpha}{V}\right),\ B = B_0\left(1 - \frac{b}{V}\right),\ \varepsilon = \frac{C}{VT^3} \tag{10.4.38}$$

for system pressures up to 250 bar.

Upon compression and expansion of the charge gas, the internal energy of the filling gas will change and heat will be exchanged with other components. The simplest model of this process assumes polytropic expansion of the gas with a constant polytropic coefficient,

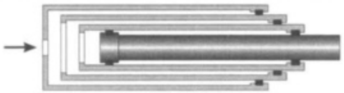
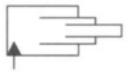
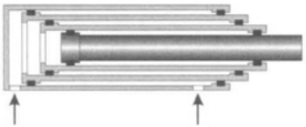
$$p_0 V_0^n = p V^n \tag{10.4.39}$$

The choice of the polytropic coefficient n depends on the general nature of volume changes. For a slow charging process, one can assume an isothermal change of state ($n = 1$). Theoretically, the polytropic coefficient is limited by $n \leq 1.4$, where the limiting case is the adiabatic expansion. This would be the case for a fast discharge, for example, to cover the peak demand. However, in reality, the heat exchange is much more complex and cannot be governed by (10.4.39). For the simulation of hydraulic accumulators, it is thus necessary to model the temperature exchange with the surroundings and to introduce the temperature as a new state.

e) Hydraulic translatory motors

Depending on the motion to be generated, one differentiates between rotary motors (hydraulic motors) and translatory motors (hydraulic cylinders). Translatory motion is generated by hydraulic cylinders, which – based on the type of piston support – can be classified as cylinders with low-friction sealing and cylinders with hydrostatic bearings. They can be constructed as single-acting and double-acting cylinders, see Table 10.22. For single-acting cylinders (plunger cylinders, telescopic cylinders), which have only one hydraulic connector, the rod is extended by means of a hydraulic force while it is retracted by means of some external force (gravity, return spring, counter cylinder). Double-acting cylinders have two hydraulic connectors, thus the piston can be both extended and retracted by a hydraulic force.

Table 10.22. Construction principles and key features of hydraulic translatory motors

	hydraulic cylinder type	construction principle	graphical symbol	maximum stroke	minimum stroke	features
single-acting	single rod cylinder					- lower weight than plunger cylinder; - more friction than plunger cylinder; - more sophisticated design.
	plunger cylinder					- simple design; - low initial cost; - low friction; - good efficiency; - stroke limited due to weight of plunger.
	telescopic cylinder			8 m	130 kN	- large stroke at small installation volume; - complicated design; - without compensation methods, the velocity and force generated by the cylinder are position-dependent.
double-acting	single rod cylinder					- unequal active piston areas; - generated force and velocity are direction-dependent; - hydraulic retraction possible.
	double rod cylinder					- equal active piston areas; - force and velocity are not direction-dependent; - transfer function parameters are identical for both chambers provided piston is centered hydraulic retraction possible.
	telescopic cylinder			2.5 m		- complicated design; - large stroke at small installation volume; - hydraulic retraction possible.

The double-acting cylinder can be employed as a pull-type or push-type cylinder. The single-rod piston is the most widespread type, available with many different piston-area ratios, different types of mounting and different types of connections. Compared to double rod cylinders, they need less installation space for the same piston stroke. Due to the different piston areas, the differential cylinder has a direction-dependent piston speed and actuation force in the case of constant supply pressure and volume flow rate.

For most applications, servo-cylinders with special contact sealing can meet the imposed requirements. However, this puts demanding constraints on the surface quality of the cylinder tube, piston rod and piston support. With this special contact sealing, undesired friction and stick-slip effects can be reduced noticeably, Backé (1986b).

For modeling a double-acting cylinder, a scheme like in Figure 10.18 is considered. The displacement of the piston is z. Both chambers are parameterized by the initial volume V_0 for $z = 0$ and the effective cross-sectional area of the piston A. The volume flow rate into the chamber is designated as $\dot{V}$ and the pressure inside the cylinder chamber as p.

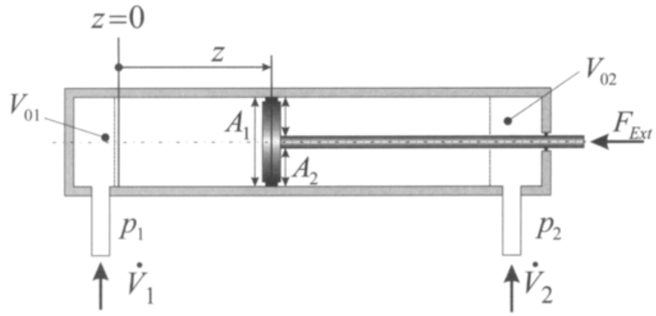


Figure 10.18. Schematic view of a hydraulic cylinder

Applying the mass balance equation (2.3.7) to the left cylinder chamber leads to

$$\begin{aligned}\dot{m}_1(t) &= \frac{d}{dt}\big(V_1(t)\rho_1(t)\big) = \dot{V}_1(t)\rho_1(t) + V_1(t)\dot{\rho}_1(t) \\ &= A_1\rho_1(t)\dot{z}(t) + \big(V_{01} + A_1 z(t)\big)\dot{\rho}_1(t)\end{aligned} \quad (10.4.40a)$$

and, for the right chamber, one obtains correspondingly

$$\begin{aligned}\dot{m}_2(t) &= \frac{d}{dt}\big(V_2(t)\rho_2(t)\big) = \dot{V}_2(t)\rho_2(t) + V_2(t)\dot{\rho}_2(t) \\ &= -A_2\rho_2(t)\dot{z}(t) + \big(V_{02} - A_2 z(t)\big)\dot{\rho}_2(t)\end{aligned} \quad (10.4.40b)$$

In order to eliminate the density from the above equations, a relation between the pressure and the density must be established. Applying $V = m/\rho$ and $\partial V/\partial\rho = -V/\rho$ leads to

$$\frac{\partial p}{\partial \rho} = \left(\frac{\partial p}{\partial V}\right)\left(\frac{\partial V}{\partial \rho}\right) = \left(-\frac{V}{\rho}\right)\left(\frac{\partial p}{\partial V}\right) \tag{10.4.41}$$

which can also be written in the form

$$\partial p = \left[\left(-\frac{1}{V}\right)\left(\frac{\partial V}{\partial \rho}\right)\right]\rho\, \partial p \tag{10.4.42}$$

Inserting the bulk modulus β (10.4.19) yields

$$\dot{\rho} = \frac{\rho}{\beta}\dot{p} \tag{10.4.43}$$

This equation can be used to rewrite the equations in (10.4.40) independent of the density as

$$\begin{aligned} \dot{p}_1(t) &= \frac{V_{01} + A_1 z(t)}{\beta} + A_1 \dot{z}(t) = \dot{V}_1(t) \\ \dot{p}_2(t) &= \frac{V_{02} - A_2 z(t)}{\beta} - A_2 \dot{z}(t) = \dot{V}_2(t) \end{aligned} \tag{10.4.44}$$

For double-acting cylinders, one must also model the cross-port leakage flow between the two chambers, which is assumed to be laminar. The pressure-flow relation is

$$\dot{V}_{1\to 2}(t) = C_{1\to 2}\big(p_1(t) - p_2(t)\big) \tag{10.4.45}$$

for a flow from chamber 1 to chamber 2. The cross-port leakage coefficient $C_{1\to 2}$ is found to be

$$C_{1\to 2} = \frac{\pi}{96\nu\rho}\,\frac{(D_o + D_i)}{2}\,\frac{(D_o - D_i)^3}{l} \tag{10.4.46}$$

where the inner diameter of the housing is assigned to D_o and the diameter of the piston to D_i. Once the pressure build-up in the chambers is known, the dynamic behavior of the piston rod can be calculated. Other forces acting on the piston besides the hydraulic forces are friction forces, inertia forces due to the mass of the piston and external forces. The balance of forces for the piston rod is finally given by

$$m_P \ddot{z}(t) + d_P \dot{z}(t) + c_P z(t) + f_C \operatorname{sign}\big(\dot{z}(t)\big) = p_1(t) A_1 - p_2(t) A_2 - F_{Ext}(t) \tag{10.4.47}$$

where F_{Ext} is the sum of all external forces, m_P is the mass of rod and piston, d_P and c_P are the damping coefficient and spring stiffness respectively and f_C is the friction coefficient for dry friction. An overall cylinder model is derived in Section 10.4.3.

f) Hydraulic rotary motors

In contrast to their pneumatic counterparts, hydraulic motors play a more important role since they offer high torque at small size (the power density of a hydraulic motor is 20–25 times greater than that of an electric motor) and have a small moment of inertia. This results in very small time constants and permits a highly dynamic manipulation of the rotational speed.

Hydraulic rotary motors are based on the positive displacement principle. They must have some mechanical element on which the pressure acts and which recurrently increases and decreases the size of an enclosed volume. Based on the shape and kinematics of this element, one can classify hydraulic motors as axial- and radial-piston motors, external and internal gear types, and sliding vane and rolling vane types. An overview of the different construction principles is given in Table 10.23. Each principle has its distinct features, which makes it suited to certain applications.

Despite this wide variety of engineering designs for hydrostatic motors, all motors work on the same positive displacement principle and can thus be modeled using the same equations. Furthermore, most functional principles can be utilized as both pumps and motors. Thus, the term "hydraulic machine" is used in the following.

First, an ideal hydraulic machine will be modeled. This model is based on the assumption that there are no losses and that the fluid is incompressible. Under these presumptions, the volume flow rate is proportional to the active volume per revolution of the machine chamber V_M and the rotational speed ω as

$$\dot{V}_{theo} = \frac{\omega}{2\pi} V_M \tag{10.4.48}$$

For both modes of operation, *i.e.*, pump and motor, the power consumed and then emitted by the ideal machine is given by

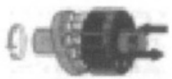
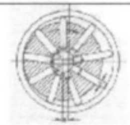
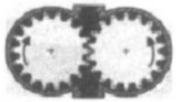
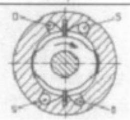
$$P = \Delta p \dot{V} \tag{10.4.49}$$

The indicator diagram for the ideal hydraulic machine is shown in Figure 10.19. Between points A and B, the chamber is connected to the intake. For an ideal pump, the chamber is filled with fluid as the volume grows from V_{min} to V_{max}. During this phase, the pressure in the chamber is identical to the pressure at the inlet p_1 .Then, the connection to the intake is closed and the connection to the outlet is opened. For an ideal pump, the pressure in the chamber rises to the pressure at the outlet p_2. Between points C and D, the fluid is pushed out of the chamber through the outlet as the volume shrinks from V_{max} to V_{min}. Finally, the connection to the outlet is closed and the chamber is coupled again to the inlet. For an ideal motor, this work cycle is executed in the reverse order.

Two effects influence the shape of the indicator diagram. The first effect is the *compressibility* of the fluid. Upon pressurization, the volume will shrink. It will expand again as the pressure level decreases. This effect is considered in the indicator diagram shown in Figure 10.19 and is governed according to (10.4.19) by

$$\frac{\Delta V}{(p_2 - p_1)} = -\frac{V}{\beta'} \tag{10.4.50}$$

Table 10.23. Characteristics of hydraulic rotary motors (hydraulic motors), Nordmann, Isermann (1999), Matthies (1995), Bauer (1998)

hydraulic rotary type	construction principle (example)	operation pressure/bar	rotational velocity / $\min^{-1}$	displacement per rev./cm^3	overall efficiency	advantages	disadvantages
axial piston motor		100–500	5–8000	2–4000	0.85–0.9	- high operating pressure; - high power density; - adjustable displacement; - good efficiency_ - low operational cost.	- complicated manufacturing; - high initial cost; - possibly long construction length.
radial piston motor		120–750	5–3000	2–35000	0.85–0.9	- high operating pressure; - high power density; - good efficiency; - generation of high torque and rotational velocities	- complicated manufacturing; - less compactly built than axial piston pumps.
external gear motor (spur gear)		80–300	200–8000	1–1000	0.6–0.9	- small unit volume; - high power density; - simple design; - rugged.	- constant displacement; - change of flow only by means of a throttle valve, thus heating-up of the oil; - high operational cost.
internal gear motor (gerotor)		< 260	10–2000	10–900	0.6–0.8	- small unit volume; - high power density; - simple design; - high torque at low rotational velocity.	- constant displacement; - bad efficiency; - high operational cost.
sliding vane motor		50–200	10–4000	2–2000	0.7–0.8	- small unit volume; - low noise level; - adjustable displacement; - low flow pulsation.	- sensitive to pressure spikes; - bad efficiency.
rolling vane and swinging vane motors		< 280	1–3000	8–1600	0.7–0.9	- small unit volume; - low noise level; - hydraulically balanced.	- sensitive to pressure spikes; - bad efficiency; - constant displacement for rolling vane motors.

This effect requires that the connections to the intake and outlet are not changed instantaneously. The shape of the indicator diagram is also affected by the flow-induced pressure losses. The indicator diagram of an ideal pump, considering the aforementioned effects, is shown in Figure 10.19. It can be seen that during the suction phase, the pressure in the pump chamber falls below the pressure at the intake. This is a critical point in the design of hydraulic pumps. If the pressure in the chambers drops below the vapor pressure, part of the fluid will vaporize. Thus, the chamber will only be partially filled with hydraulic fluid. This effect may also lead to "cavitation". On the other hand, while the fluid is pushed out of the chamber, the pressure will rise above the pressure at the outlet port. For the hydraulic motor, a similar indicator diagram is plotted in Figure 10.19. For a hydraulic motor, the pressure in the chamber will never drop below the pressure at the low pressure port. Therefore, hydraulic motors are not susceptible to cavitation.

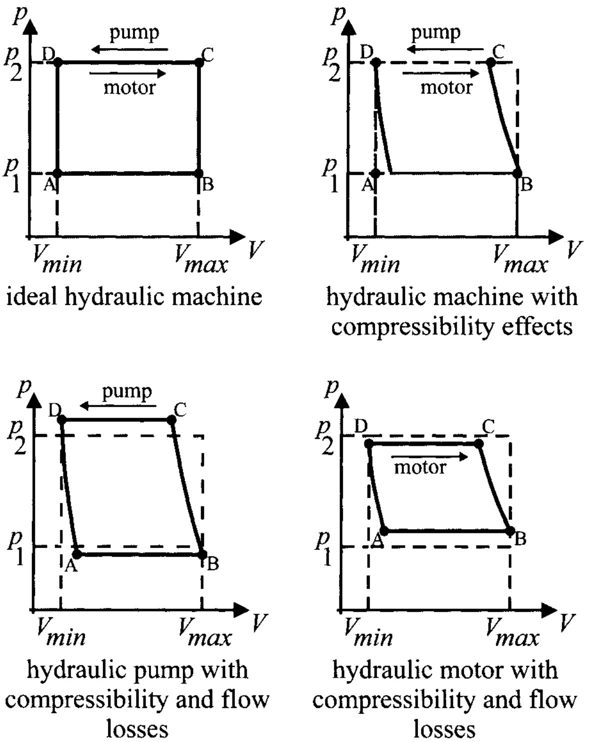


Figure 10.19. Indicator diagrams for hydraulic machines

In contrast to an ideal hydraulic machine, the performance of a real hydraulic machine is impaired by a couple of losses, which are largely divided into volumetric and hydraulic-mechanical losses. The former cause a difference between the volume flow rate of an ideal machine and a real machine. Several factors affect the volume flow rate. The first effect that induces *volumetric losses* is the compressibility of the fluid, because the volume will shrink upon pressurization in the hydraulic pump by

$$\dot{V}_{vol} = \frac{\omega}{2\pi} \frac{V_M}{\beta'} \Delta p \tag{10.4.51}$$

It may be noted that as far as the energy flow is concerned, this effect cannot be called a loss, since the energy consumed during the compression of the fluid in the pump is stored in the fluid itself and can thus be regained in the hydraulic motor. Another effect is the incomplete charge of the pump chamber if the pressure in the intake drops below the vapor pressure (cavitation), which is characteristic of hydraulic pumps. Cavitation typically occurs at higher rotational velocities, since the volume flow rate increases with the rotational velocity as does the flow-induced pressure loss. Both of these effects, compressibility of the oil and cavitation, are usually not modeled, since they are negligible compared to the internal and external leakage flows, the biggest source of volumetric losses. Since most machines are constructed such that fluid leaking from the machine is directed to the port with the lower pressure level, external leakage flows can also be treated as being internal. The exact nature of the leakage flows cannot be determined in most cases, therefore this flow is modeled as being partially laminar and partially turbulent

$$\dot{V}_{Leak} = k_{Lam} \Delta p + k_{Turb} \sqrt{\frac{\Delta p}{\rho}} \tag{10.4.52}$$

where k_{Lam} and k_{Turb} are the corresponding coefficients. This model is well suited to lower pressure levels. For higher pressure, the losses increase noticeably as the gaps expand and as the energy dissipation and thus the temperature of the fluid increases.

Hydro-mechanical losses stem from friction within the fluid, on friction surfaces and in bearings. The change in the momentum of the enclosed fluid also uses up part of the torque.

Taking all of these effects into account, the volume flow rate is given by

$$\begin{aligned} \dot{V} &= \dot{V}_{theo} - \dot{V}_{Vol} - \dot{V}_{Leak} \\ &= -\dot{V}_{theo}\left(1 - \frac{\dot{V}_{Vol}}{\dot{V}_{theo}} - \frac{\dot{V}_{Leak}}{\dot{V}_{theo}}\right) = \dot{V}_{theo}\eta_{Vol} \end{aligned} \tag{10.4.53}$$

the output power of a hydraulic pump is therefore

$$P_1 = \dot{V}_1 \Delta p_1 = \dot{V}_{1theo} \eta_{1Vol} \Delta p_1 = \frac{\omega_1}{2\pi} V_{1M} \Delta p_1 \eta_{1Vol} \tag{10.4.54}$$

and its input power

$$P_1 = \frac{\omega_1}{2\pi} V_{1M} \Delta p_1 \frac{1}{\eta_{1m}} \tag{10.4.55}$$

where the mathematical efficiency takes into account mechanical losses. The driving torque then becomes

$$T_1 = \frac{P_1}{\omega_1} = \frac{1}{2\pi} V_{1M} \Delta p_1 \frac{1}{\eta_{1m}} \tag{10.4.56}$$

A hydraulic rotary motor receives volume flow $\dot{V}_1$, which results in an effective motor volume flow

$$\dot{V}_2 = \dot{V}_1 \eta_{2Vol} \tag{10.4.57}$$

and generates the output power of the motor

$$P_2 = \dot{V}_1 \eta_{2Vol} \eta_{2m} \Delta p_2 = \frac{\omega_2}{2\pi} V_{2M} \Delta p_2 \eta_{2Vol} \eta_{2m} \tag{10.4.58}$$

and the output torque

$$T_2 = \frac{P_2}{\omega_2} = \frac{1}{2\pi} V_{2M} \Delta p_2 \eta_{2Vol} \eta_{2m} \tag{10.4.59}$$

Figure 10.20 shows the torque-volume flow characteristics of a hydraulic motor.

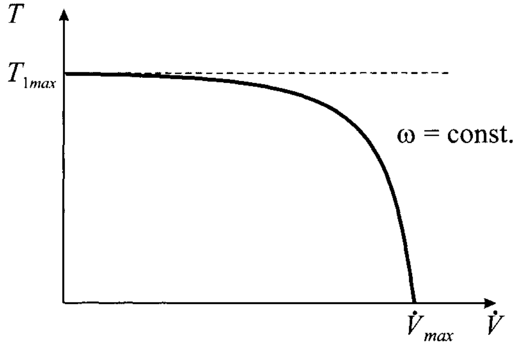


Figure 10.20. Characteristic torque-curve for a hydraulic motor

Almost all positive displacement machines have a finite number of displacing bodies. Therefore, they are not able to provide a continuous supply of pressurized hydraulic fluid. Rather, the flow is pulsating, which is a major issue in the design of hydraulic power systems as this causes pressure oscillations that might excite oscillations in subsequent hydraulic elements and can be a major source of noise emitted by the hydraulic system.

For a *variable displacement machine*, the volume flow rate is multiplied with a factor κ,

$$\dot{V} = \kappa n V_M \tag{10.4.60}$$

which lies between 0 and 1 for a normal machine or −1 to 1 for a reversing machine. This factor captures the change in output due to the variation of the machine geometry. Most variable displacement machines, in particular wash plate and bent-axis machines, are controlled by the angular displacement α of some part of the set-up. For these machines, κ is given by

$$\kappa = \tan \alpha = f(U) \tag{10.4.61}$$

A schematic cross-sectional view of a wash plate machine is shown in Figure 10.21. The figure is meant to illustrate the basic principle. In real designs, typically an odd number of pistons is chosen since this will reduce pressure oscillations in the high-pressure line connected to the pump outlet.

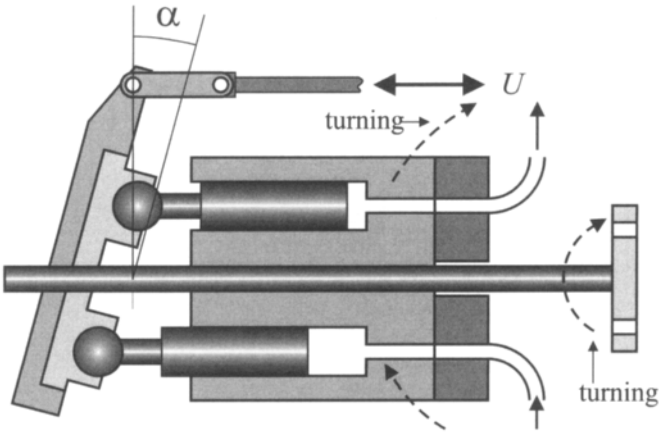


Figure 10.21. Schematic view of a wash plate machine

10.4.3 Models of a Hydraulic Servo-axis: An Example

An example for modeling a hydrodynamic overall system, the hydraulic servo-axis of Figure 10.10 is considered. The system consists of an axial piston pump, a proportional valve and a hydraulic cylinder, which will work against a mass m_{Load}. This set-up represents the work-table drive of a CNC drilling machine. For machine tools, linear hydraulic actuators can be used to replace the traditional ball bearing spindle-servomotor combination. It is assumed that the hydraulic actuator is supplied by a piston pump that is operated at a constant speed. The input of the actuator model will be the voltage V applied at the proportional valve and the output will be the position z of the cylinder and the work-table respectively.

a) Non-linear model

It will be assumed that the supply pressure p_s is constant and that the pump can always deliver the required amount of hydraulic fluid to the cylinder. Thus, the behavior of the pump does not need to be modeled. The connection pipes will not be modeled either. It is presumed that their contribution to the dynamics of the overall system is negligible.

The first subsystem under consideration is the proportional valve. The degree of valve opening H will be defined as the ratio between the actual and the maximum spool displacement,

$$H = \frac{z_s}{z_{s,\max}} \tag{10.4.62}$$

For the control of the double-acting cylinder, a four-way (4/3) valve will be used. There are four orifices, whose geometry can be controlled. For each orifice, the pressure-flow relation will be modeled using the equation

$$\Delta p = \frac{R_{\min}}{H^2} \dot{V}^2 \operatorname{sign}(\dot{V}) = \frac{R_{\min}}{H^2} \dot{V} |\dot{V}| \tag{10.4.63}$$

where R_{min} denotes the minimum hydraulic resistance, *i.e.*, the hydraulic resistance of the fully open valve. It is assumed that the valve is zero-lapped. For $z_s = 0$, all four orifices will be closed. Two orifices will open if $z_s > 0$. For $z_s < 0$, the other two orifices will conduct the hydraulic flow. The pressure at the outlet will be set to zero. Under normal operating conditions, it is assumed that the flow into one cylinder chamber is equal to the flow out of the opposite chamber (this assumption is referred to as "matched and symmetric orifices"), thus

$$\dot{V}_{Load} = \dot{V}_1 = -\dot{V}_2 \tag{10.4.64}$$

If the volume flow rate through two identical orifices is the same, the resulting pressure loss across the apertures will also be the same. Thus, the pressure loss across one orifice, denoted by Δp, can be expressed as

$$\Delta p = \frac{1}{2}(p_s - p_{Load}) \tag{10.4.65}$$

where p_{Load} is the pressure difference between the two cylinder chambers. These equations can be combined into

$$\dot{V}_{Load} = \frac{|H|}{\sqrt{2R_{\min}}} \sqrt{|p_s \operatorname{sign}(H) - p_{Load}|} \operatorname{sign}(p_s \operatorname{sign}(H) - p_{Load}) \tag{10.4.66}$$

which results in the block diagram shown in Figure 10.22. Usually, proportional valves contain an inner control loop, which controls the spool positioning. The control loop is modeled as a first order system.

The valve is connected to a hydraulic cylinder, which will be supplied with the load flows $\dot{V}_1$ and $\dot{V}_2$. According to the equation of continuity, the flow into the cylinder chamber will lead to a change in the density of the fluid contained in the cylinder or to a change in the displacement of the piston, which effectively changes the volume of the cylinder chamber. For the following derivation, the zero reference point of the piston displacement will be the cylinder mid-point. The volume of the two cylinder chambers is then given by

$$\begin{aligned} V_1 &= V_{01} + A_1 z_p \\ V_2 &= V_{02} - A_2 z_p \end{aligned} \tag{10.4.67}$$

In these equations, V_{01} and V_{02} denote the initial volume of the two chambers. The increase in pressure due to compression of the fluid is governed by

$$\Delta p = \frac{\beta}{V} \Delta V \tag{10.4.68}$$

These lead to the block diagram shown in Figure 10.22. For the mechanical subsystem, the force acting on the piston rod is the difference between the two hydraulic forces less any frictional forces. Friction will be modeled as a combination of Coulomb and viscous friction, see Figure 10.22.

Two simulations running with parameters according to Table 10.24 show the behavior of the hydraulic servo-axes. The first simulation run, shown in Figure 10.23, depicts the system response due to a step input applied at the control input of the valve, *i.e.*, at the terminals of the electromagnet. One can see the time lag of the step response due to the pressure build-up in the chambers and the inertia effects of the connected mass. Later, the system comes to a point where the acceleration gets more and more determined by the flow resistance in the valve.

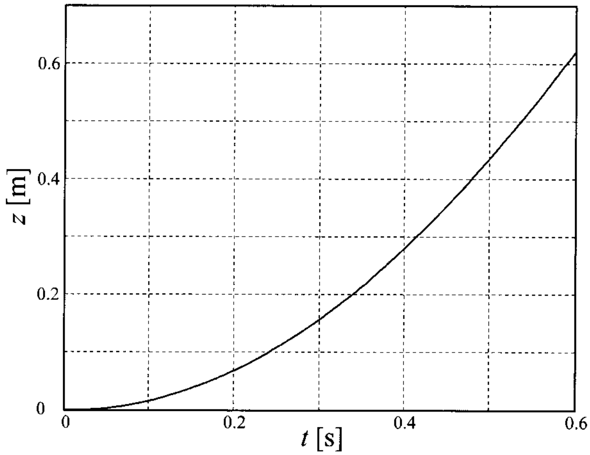


Figure 10.23. Simulated transient function of the position

The second simulation run, Figure 10.24, describes the system behavior if an external force is applied at the piston rod at $t = 0.1$ s. As the piston moves in the negative x direction, the volume enclosed by one chamber shrinks, whereas the volume enclosed by the other chamber grows. The chamber whose volume decreases is tightly sealed.

Therefore, the encompassed fluid acts like a spring, the so-termed oil spring. The oscillations of the piston can be seen in the diagram. The dotted horizontal line denotes the theoretically expected displacement of the piston rod as calculated using (10.4.19). The difference of the theoretically derived and the effectively reached volume can be explained by Coulomb friction effects.

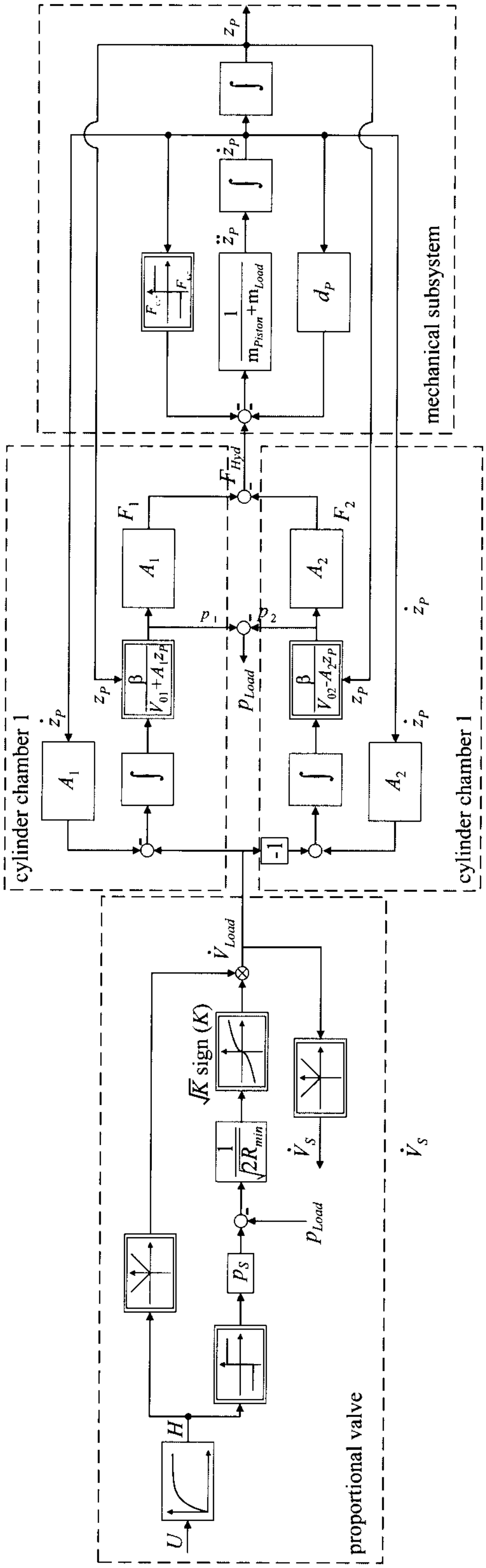


Figure 10.22. Block diagram of the hydraulic servo-axis; $K = |p_S \,\mathrm{sign}\,(H) - p_{Load}|$, see (10.4.66)

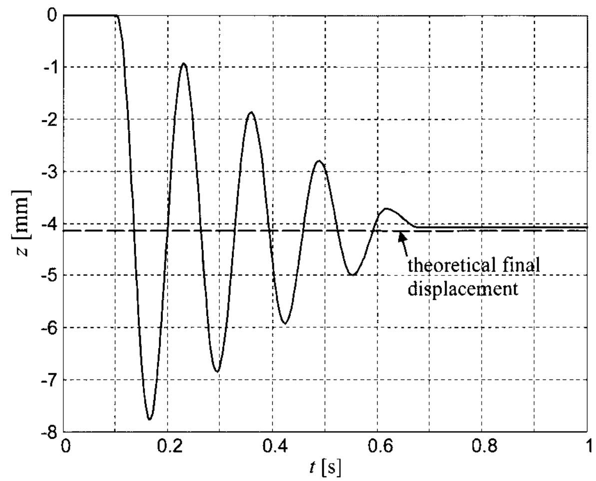


Figure 10.24. Simulated transient function of the position for a stepwise external force input

The parameters for the simulation runs are listed in Table 10.24. The value for Coulomb friction for simulation 1 resembles a low friction (servo) cylinder, whereas the value for simulation 2 was chosen rather high to illustrate the problems associated with Coulomb friction.

Table 10.24. Parameters for the simulation of an hydraulic servo-axis

	simulation 1 valve-controlled		simulation 2 external force	
time constant of the servo-valve (critically-damped system)	$T_{Valve,I} = T_{Valve,I} = 0.01$	s	-	
minimum hydraulic resistance	$R_{min} = 250000$	$\frac{Pas^2}{m^6}$	-	
supply pressure	$p_S = 50 \cdot 10$	Pa	-	
piston area left and right	$d = 765.76 \cdot 10^{-6}$	m^2	$d = 765.76 \cdot 10^{-6}$	m^2
maximum displacement of cylinder	$l = 0.6$	m	$l = 0.6$	m
dead volume left	$V_{01} = 3.2182 \cdot 10^{-4}$	m^3	$V_{01} = 3.2182 \cdot 10^{-4}$	m^3
dead volume right	$V_{02} = 7.047 \cdot 10^{-4}$	m^3	$V_{02} = 7.047 \cdot 10^{-4}$	m^3
total weight of mechanical system and moving parts of hydraulic cylinder	$m_{Load} = 1000$	kg	$m_{Load} = 1000$	kg
Coulomb friction	$F_C = 100$	N	$F_C = 500$	N

b) Linearized model

For some tasks, it is possible to work with a linear system model. This is especially true for problems such as stability analysis and controller design. Therefore, linear models of valves and hydraulic cylinders are presented.

In order to derive a linear valve model, the flow through the valve is expressed as a Taylor-series expansion around the operating point,

$$\dot{V}_{Load} = \dot{V}_{Load,0} + \frac{\partial \dot{V}_{Load}}{\partial z_S}\Big|_{O.P.} \Delta z_S + \frac{\partial \dot{V}_{Load}}{\partial p_{Load}}\Big|_{O.P.} \Delta p_{Load} + \ldots \qquad (10.4.69)$$

The quantity

$$K_{\dot{V}} = \frac{\partial \dot{V}_{Load}}{\partial z_S} \qquad (10.4.70)$$

is called the flow gain. The flow-pressure coefficient is defined as

$$K_C = \frac{\partial \dot{V}_{Load}}{\partial p_{Load}} \qquad (10.4.71)$$

One can also define the pressure sensitivity

$$K_P = \frac{\partial p_{Load}}{\partial z_S} = \frac{K_C}{K_{\dot{V}}} \qquad (10.4.72)$$

These three parameters are termed valve coefficients. With these coefficients, (10.4.64) can be written as

$$\dot{V}_{Load} = K_{\dot{V}} z_s - K_C p_{Load} \qquad (10.4.73)$$

which is the valve equation linearized around a certain operating point. The resulting block diagram is shown in Figure 10.25.

The most important point is $\dot{V}_{Load} = p_{Load} = z_S = 0$, because all normal operating points will be in the close vicinity of this point. In addition, this operating point is of special importance for the investigation of stability. Here, the flow gain is largest, resulting in a large overall system gain. Furthermore, the flow-pressure coefficient is the smallest, yielding a small damping ratio. Therefore, this is usually the most critical point as far as stability of the system is concerned. The valve coefficients evaluated at this point are called the null valve coefficients.

In most servo-hydraulic applications, zero-lapped valves, also referred to as critical center valves, are used. For them, the values of the valve coefficients will now be determined analytically by differentiating the pressure-flow relation (10.4.17), which for the case at hand can be written as

$$\dot{V}_{Load} = \alpha_D A' z_S \sqrt{\left| \frac{1}{\rho}\left(p_S - \mathrm{sign}(z_S) p_{Load}\right)\right|}\, \mathrm{sign}\left(p_S - sign(z_S) p_{Load}\right) \mathrm{sign}(z_S)$$

$$(10.4.74)$$

where the direction of the flow is determined by both the spool stroke and the pressure gradient across the orifice. For the following analysis, a positive pressure gradient and a positive spool displacement is assumed. Due to the symmetry assumptions, this can be done without a loss of generality. Taking the partial derivatives results in

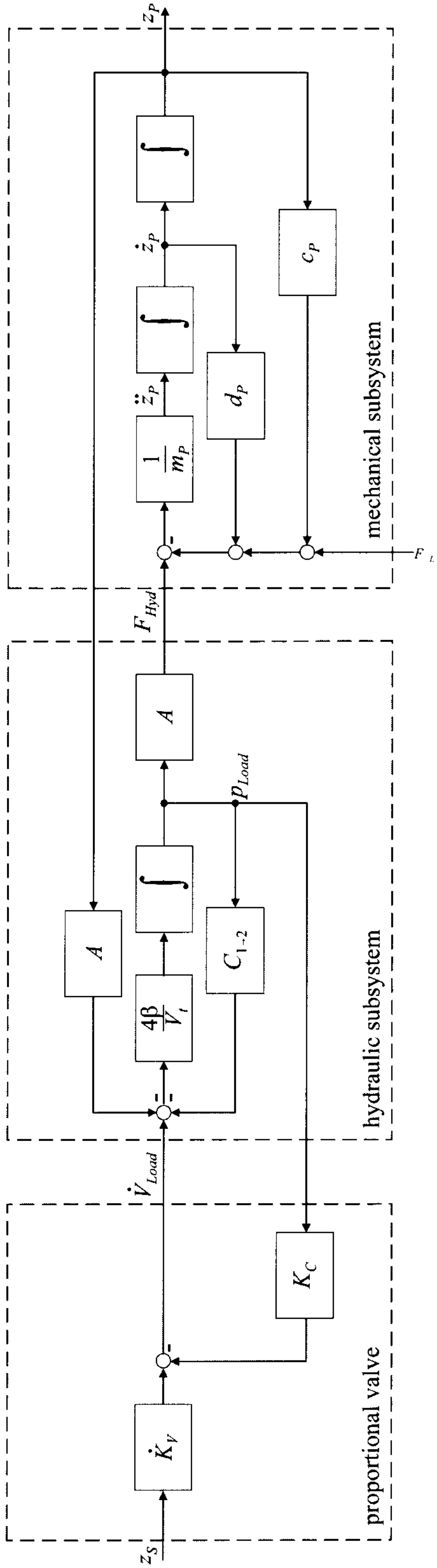


Figure 10.25. Block diagram of the total linearized system of a hydraulic actuator

$$K_{\dot V} = \frac{\partial \dot V_{Load}}{\partial z_S} = \alpha_D A' \sqrt{\frac{1}{\rho}} \sqrt{p_S - p_{Load}}$$

$$K_C = -\frac{\partial \dot V_{Load}}{\partial p_{Load}} = \alpha_D A' z_S \frac{\sqrt{\frac{1}{\rho}} \sqrt{p_S - p_{Load}}}{2\left(p_S - p_{Load}\right)}$$

$$K_p = \frac{K_{\dot V}}{K_C} = \frac{2\left(p_S - p_{Load}\right)}{z_S} \tag{10.4.75}$$

Evaluating these coefficients at the point $\dot V_{Load} = p_{Load} = z_S = 0$ results in

$$K_{\dot V,0} = \alpha_D A' \sqrt{\frac{1}{\rho}} \sqrt{p_S}$$

$$K_{C,0} = 0$$

$$K_{p,0} \rightarrow \infty \tag{10.4.76}$$

For the flow gain, the value matches well with experimental results, Merrit (1967), whereas the other two coefficients should be determined using a different method that determines these coefficients by looking at the center flow curve. The curve is a recording of the leakage flow for a centered spool and blocked load ports. For this experimental configuration, a leakage flow inside the valve can be observed upon pressurization. The slope of the curve at the particular operating pressure can be used as the null flow-pressure coefficient.

Now, the piston dynamics will be linearized. The volume flow into the two chambers can be calculated by evaluating the equation of continuity (10.4.2), which evaluates to

$$\dot V_1 - C_{1\rightarrow 2}\left(p_1 - p_2\right) = \frac{\partial V_1}{\partial t} + \frac{V_1}{\beta} \frac{\partial p_1}{\partial t}$$

$$\dot V_1 - C_{1\rightarrow 2}\left(p_1 - p_2\right) = \frac{\partial V_2}{\partial t} + \frac{V_2}{\beta} \frac{\partial p_2}{\partial t} \tag{10.4.77}$$

The volume of the two chambers can be calculated as shown in (10.4.67). Here, it is assumed that both piston areas are identical, hence

$$V_1 = V_{01} + A z_P$$

$$V_2 = V_{02} - A z_P \tag{10.4.78}$$

The total volume of the cylinder will be denoted V_{tot}. It is called the *total compressed volume* (or also total contained volume) and is given by

$$V_{tot} = V_{01} + V_{02} = V_1 + V_2 \tag{10.4.79}$$

Equations (10.4.78) and (10.4.79) will now be combined into one equation. Substituting the total compressed volume and the volume flow rate yields

$$\dot V_{Load} - 2 C_{1\rightarrow 2} p_{Load} = 2 A \dot z_P + \frac{A z_P}{\beta}\left(\dot p_1 + \dot p_2\right) + \frac{V_{01}}{\beta} \dot p_1 - \frac{V_{02}}{\beta} \dot p_2 \tag{10.4.80}$$

It will be assumed that the initial volume of both chambers is identical, $V_{01} = V_{02} = V_0$. Then, the equation simplifies to

$$\dot{V}_{Load} - C_{1-2}p_{Load} + A\dot{z}_P + \frac{V_{tot}}{4\beta}p_{Load} \tag{10.4.81}$$

Since the supply pressure is assumed to be constant, the term

$$\frac{\partial p_1}{\partial t} + \frac{\partial p_2}{\partial t} = \frac{\partial}{\partial t}\left(\frac{p_S - p_V}{2}\right) + \frac{\partial}{\partial t}\left(\frac{p_S + p_V}{2}\right) = \frac{\partial}{\partial t}p_S = 0 \tag{10.4.82}$$

vanishes. The block diagram based on the linearized equations for the pressure build-up in the cylinder chambers is shown in Figure 10.25.

This equation is supplemented by Newton's second equation, which governs the dynamics of the piston

$$Ap_{Load} - F_L = m_P\ddot{z}_P + d_P\dot{z}_P + c_Pz_P \tag{10.4.83}$$

where F_L denotes an arbitrary load force on the piston. The block diagram of the mechanical subsystem is illustrated in Figure 10.25.

Equations (10.4.73), (10.4.80) and (10.4.83) are linear and can now be Laplace-transformed to derive the transfer functions of the overall system. The complete transfer function is given by

$$z_p(s) = \frac{\frac{K_{\dot{V}}}{A}z_S(s) - \frac{K^*}{A^2}\left(1 + \frac{V_t}{4\beta K^*}s\right)F_L(s)}{\frac{m_PV_t}{4\beta A^2}s^3 + \left(\frac{V_td_P}{4\beta A^2} + \frac{m_PK^*}{A^2}\right)s^2 + \left(\frac{d_PK^*}{A^2} + 1 + \frac{c_PV_t}{4\beta A^2}\right)s + \frac{c_PK^*}{A^2}} \tag{10.4.84}$$

Here, m_p denotes the driven mass, *i.e.*, piston and load. V_t is the total compressed volume. A special case can now be considered. Usually, the stiffness of the entire set-up can be neglected. For many systems, the term $\frac{d_PK^*}{A^2}$ is negligible compared to 1. Then, the transfer function is given by

$$z_p(s) = \frac{\frac{K_{\dot{V}}}{A}z_S(s) - \frac{K^*}{A^2}\left(1 + \frac{V_t}{4\beta K^*}s\right)F_L(s)}{s\left(\frac{m_PV_t}{4\beta A^2}s^2 + \left(\frac{V_td_P}{4\beta A^2} + \frac{m_PK^*}{A^2}\right)s + 1\right)} \tag{10.4.85}$$

with

$$K^* = K_C + C_{1-2} \tag{10.4.86}$$

Comparing the denominator with a second order system yields a natural frequency of

$$\omega_0 = \sqrt{\frac{4\beta A^2}{m_P V_t}} \tag{10.4.87}$$

and a damping ratio of

$$\zeta = \frac{d_P}{4A}\sqrt{\frac{V_t}{\beta m_P}} + \frac{K^*}{A}\sqrt{\frac{m_P \beta}{V_t}} \tag{10.4.88}$$

These two equations also give hints to the designer on how to reach good system performance. First, (10.4.88) will be discussed. The damping ratio is, in general, influenced by viscous friction as well as flow losses evoked by the leakage flow between the two cylinder chambers. Many hydraulic cylinders are designed as low-friction cylinders. For these, the latter term in (10.4.88) is predominantly influencing the damping ratio. It will now be investigated in detail how different parameters change the system performance.

The parameter K^* can be increased by increasing the leakage flow between the two chambers. However, this goes along with a decrease in stiffness of the servo-axis. The total mass also influences the damping ratio. Unfortunately, this quantity is most of the time beyond the control of the engineer designing the hydraulic system. It is worthwhile mentioning that for hydraulic systems – as opposed to mechanical spring-mass systems – an increase in mass also increases the damping ratio.

The other variable quantities of the cylinder, namely the active cross-sectional area A and the total compressed volume in the cylinder, are determined by the required maximum force and displacement and can therefore not be varied. The total compressed volume is given as the sum of the compressed volume of the cylinder and that of the connection lines. One way to reduce the total compressed volume is to shorten the connection lines. It would be best to eliminate the connection lines totally, thus the valve should be mounted right on top of the hydraulic cylinder. The natural frequency (10.4.87) can also be varied by changing the driven mass m_P or the total compressed volume V_t. This fact emphasizes the importance of having short connection lines as a means of reducing the total compressed volume.

There are some non-linearities in hydraulic circuits that have not been captured by the linearized model. In the development of the transfer function, a zero-lapped valve was assumed. An underlapped valve increases the leakage flow between the chambers for small or zero spool displacement. In contrast, an overlapped valve introduces a dead band into the control circuit and thus tends to destabilize the entire system. If the spool is displaced, then the valve coefficients will also vary drastically. Upon opening the valve, the load pressure will increase due to the dynamic forces. The inter-chamber leakage will also increase.

During operation of the hydraulic cylinder, the displacement of the piston will vary. This means that the amount of fluid contained in the two cylinder chambers will vary. This causes a change in the stiffness of the so-called oil springs and will thus alter the natural frequency of the system.

10.5 PNEUMATIC ACTUATORS

10.5.1 Pneumatic Actuating Systems

Pneumatic actuators exploit the physical characteristics of compressed air. The high compressibility of air, along with the capability of storing a larger amount of energy and the low viscosity of the transmission medium, permit the design of efficient and fast drives. Offering a rugged and simple design (only one supply line necessary), these pneumatic drives are well suited to applications where typical forces of a few kN must be supplied by the actuators. They can move at high velocities and over long ranges. Besides these features, they are characterized by very safe operation even under extreme ambient conditions (temperature resistance, contamination resistance, overload capability, explosion-proof construction). The system is immune against interference caused by electric and magnetic fields, as well as radiation, Heinbach *et al.* (1977), Backé (1986a), Schriek, Sonemann (1988), Table 10.25.

Table 10.25. Characteristics of pneumatic actuators

advantages	disadvantages
• good working capacity; • good thermal operating range; • good power–weight ratio; • high reliability and operating safety; • good price–performance ratio; • one supply line.	• conditioning of compressed air necessary; • to some extent: large dimensions; • friction and compressibility complicate control; • limited positioning accuracy.
Scope of application:	• medium to high actuating force; • medium to high displacement range; • hazardous-duty and high-temperature applications; • high movement speed; • low positioning accuracy.

Pneumatic actuators basically consist of a valve and an actuating device, which transforms the pneumatic energy into mechanical energy, Figure 10.26. The valve is connected to the pneumatic pressure line, which is supplied by an air compressor and controls the pressurized air flow to the actuating device. Different to hydraulic systems, pneumatic systems require only one forward line and no closed fluid circuit, as the air flows into the environment after expansion.

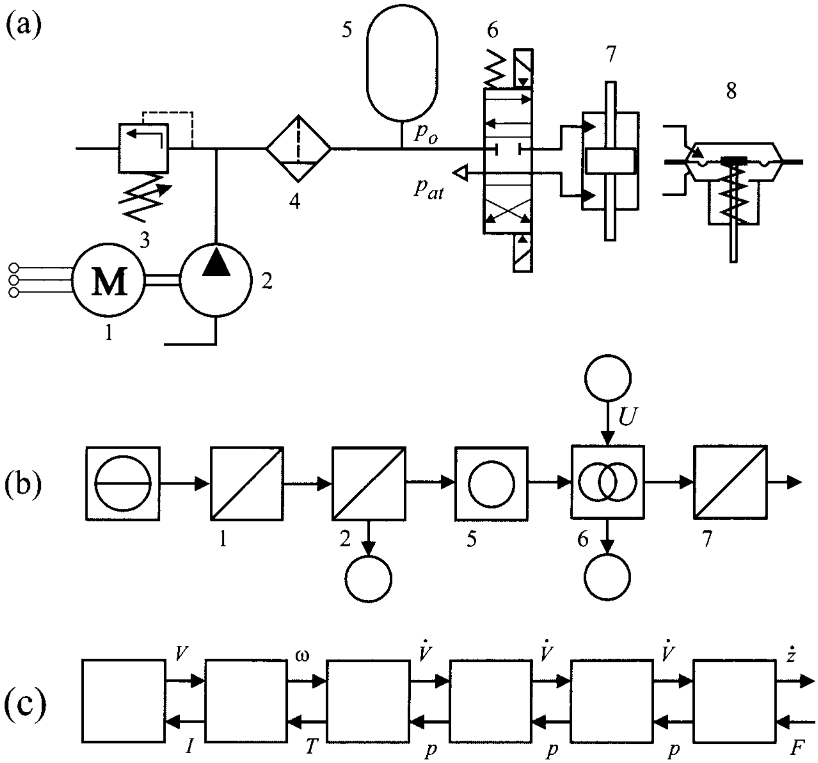


Figure 10.26. Pneumatic actuator for linear motion with power supply: (a) schematic; (b) energy flow scheme; (c) two-port representation 1: AC motor; 2: air compressor; 3: pressure relief valve; 4: air filter with water trap; 5: air storage (accumulator); 6: 4/3 proportional valve, solenoid actuation, spring return; 7: double rod cylinder; 8: diaphragm drive

The valves are either switching valves or proportional acting valves. In the case of switching valves, they have usually either two positions for both movement directions of the actuator or three positions for both directions and a holding position. The used symbols for the valves are the same as for hydraulic systems, see Figure 10.15. These switching valves are usually moved by on/off electromagnets.

Proportional acting valves allow a continuous manipulation of the air flow and need a proportional acting electromagnet, frequently with position feedback control.

Pneumatic actuating devices can largely be divided into pneumatic cylinders or diaphragms generating a translatory motion, and air motors generating a rotary motion. Figure 10.27 shows some schemes of pneumatic cylinders. The cylinders consists of a hollow cylindrical tube along which a piston and a ram can slide. The piston with seals or piston rings separates the two chambers. Single-acting cylinders are used when the pressure is applied on one side of the piston and the external force or a return spring is used to provide the opposition of the motion. In the case of double-acting cylinders, the controlled pressures are applied to each side of the piston and the pressure difference then results in a force and motion of the piston being able to move in either direction. Magnetic coupling between the piston and an outer ring allow a rodless construction. Larger travelways are possible by telescope arrangements.

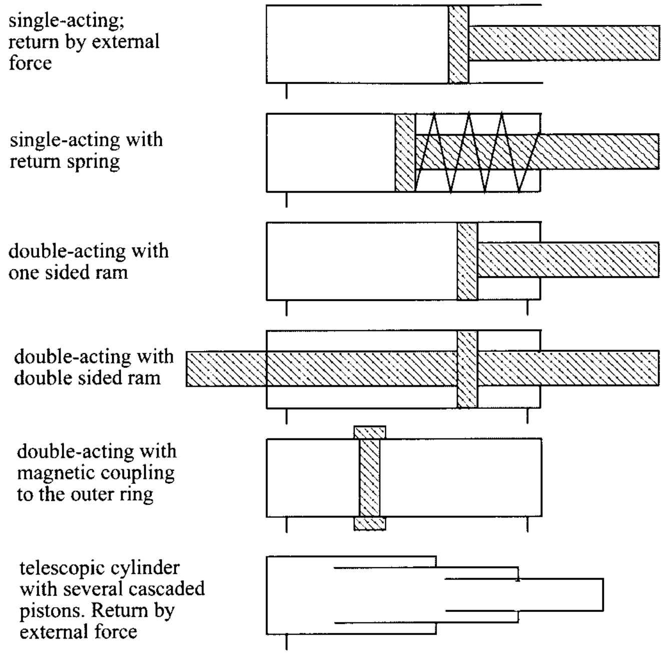


Figure 10.27. Schemes of pneumatic cylinders

Pneumatic diaphragm actuators are mainly used for flow control valves in the process industries, especially in explosive surroundings or as position control valves as, *e.g.*, for exhaust gas turbochargers or brake force amplifiers in automobiles. Figure 10.28a shows a schematic diagram. It consists of a diaphragm with the controlled pressure on one side and another pressure, most frequently atmospheric pressure, on the other side. The diaphragm is made of rubber, which is centered between two steel discs such that the control part of the diaphragm results in the motion of a shaft. The shaft then acts on a valve or lever construction. Usually, the diaphragm acts against a return spring. In the case of pressure loss, the return spring then closes or opens, *e.g.*, the control valve, dependent on which side of the diaphragm the working pressure is applied and an intended fail-safe position of the valve. This is also referred to as a "closing valve" or "opening valve".

For both arrangements, the cylinder and the diaphragm, changes in the external force on the ram or shaft result in displacement of the actuator. In addition, the relatively large friction forces of the piston or the shaft have an effect on the precision of positioning. Therefore, a position controller is frequently used to compensate for these negative effects.

For relatively light loads and small displacements, pneumatic *bellows* can be used, Figure 10.28b. Most commonly they are used as pressure-sensing devices, like in pneumatic controllers.

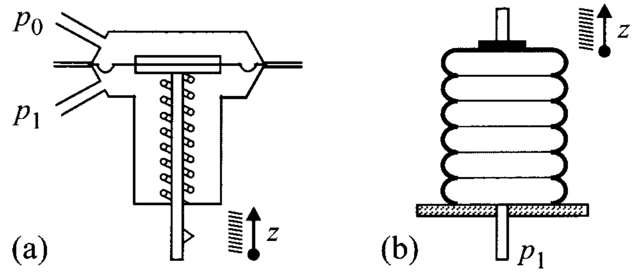


Figure 10.28. (a) pneumatic diaphragm actuator; (b) pneumatic bellow actuator

To generate a continuous rotation, pneumatic motors are available, *e.g.*, as vane types, see Figure 10.29, or piston types. They are intrinsically safe (explosive areas) and robust, but have relatively low efficiency and are noisy. Pneumatic rotary motors are used for tools in manufacturing or on ships or as starters for combustion engines, see Atlas-Copco (1977).

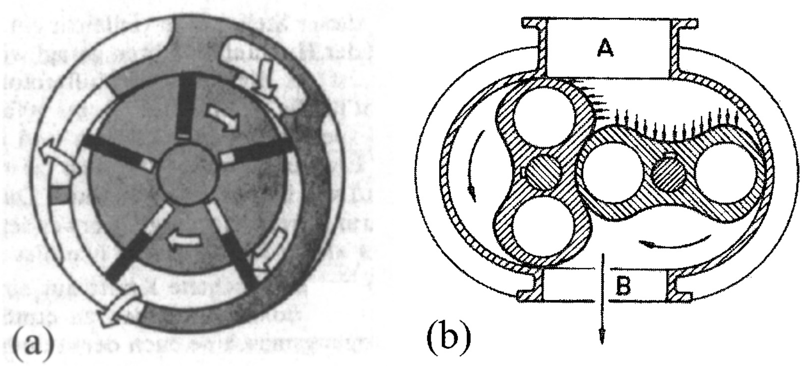


Figure 10.29. Pneumatic rotary motors: (a) vane motor; (b) roots motor

10.5.2 Pneumatic Components and their Models

a) Some gas dynamic properties

Like hydraulic systems, pneumatic systems use a fluid to transmit power between different components. However, the fluid here is the air and is therefore a compressible gas. As the compressibility of fluids is not the dominating property of a hydraulic system and can be neglected in many cases, it is dominating for pneumatic systems. On the other hand, the effect of accelerated masses can be neglected in many cases of pneumatic system behavior.

The principle of linear momentum for a pneumatic line is like for general fluids described by (10.4.1). Also, the flow dynamic equations like Bernouilli's equation (10.4.8) to (10.4.16) dependent on the Reynolds number hold for pneumatic pipes.

The physical states of air, consisting of 79.09% N_2, 20.25% O_2, 0.92% Ar, 0.03% CO_2, 0.002% Ne, 0.0005% He, the pressure, volume, density and temperature follow the constitutive equation of gas, the *gas law* for an ideal gas

$$pV = nR_m T \tag{10.5.1a}$$

or

$$pV = mRT \tag{10.5.1b}$$

or

$$pv = RT \tag{10.5.2}$$

with the specific volume $v = 1/\rho$, absolute temperature T and the gas constant R. If in (10.5.1a) n is the quantity in mol, then R_m is the universal gas constant

$$R_m = 8.314510 \quad \mathrm{J/mol{\cdot}K}$$

and is valid for any ideal gas. However, with m as the mass of the gas in (10.5.1b), R is called the specific gas constant and becomes dependent on the type of gas, Table 10.26. The state of the gas is defined if three of the four variables in (10.5.1) or two of the three variables in (10.5.2) are given.

Table 10.26. Specific gas constant R of some gases

Gas		N_2	O_2	CO_2	H_2O	CO	H_2	air
R	$\frac{\mathrm{J}}{\mathrm{kgK}}$	296.8	259.8	188.9	461.5	296.8	4124.4	286.9

Using (10.5.2), the gas constant R can be interpreted as the required energy to change the volume of m = 1 kg gas for constant pressure p through heating by $\Delta T = 1$ K.

The gas law is valid for ideal gases, which are all gases that are far away from condensation. This holds for all real gases if $p \leq 1$ bar. For $p \approx 20$ bar, errors of the gas law are smaller than 1%.

Caloric equations of state describe the relation between one caloric state variable and two thermal states. Thus, the specific inner energy of an ideal gas is

$$u = c_v T \tag{10.5.3}$$

with c_v the specific heat capacity for constant volume. For the enthalpy, it is valid that

$$h = c_p T \tag{10.5.4}$$

where c_p is the specific heat capacity for constant pressure. Further, it holds that

$$R = c_p - c_v \tag{10.5.5}$$

The isentropic exponent (adiabatic coefficient) is defined as

$$\kappa = c_p / c_v \tag{10.5.6}$$

and is for one atomic gas $\kappa = 1.66$, for two atomic gases $\kappa = 1.4$ and for three atomic gases $\kappa = 1.3$. Some characteristic properties of air are given in Table 10.27.

Changes of state of gases are usually represented in p-v diagrams or T-s diagrams (s: entropy). According to the gas law, the state of one variable depends for a certain mass of gas on two other variables as, *e.g.*, $p = f(v, T)$. Therefore, generally three-dimensional trajectories result.

Table 10.27. Characteristic properties of air

variable	density	specific volume	specific heat p = const.	specific heat v = const.	gas constant		heat conduction coefficient
symbol	ρ	v	c_p	c_v	R	$\kappa = \frac{c_p}{c_v}$	λ
dimension	$\frac{\text{kg}}{\text{m}^3}$	$\frac{\text{m}^3}{\text{kg}}$	$\frac{\text{J}}{\text{kgK}}$	$\frac{\text{J}}{\text{kgK}}$	$\frac{\text{J}}{\text{kgK}}$	-	$\frac{\text{W}}{\text{mK}}$
defined at	273 K 1.013 bar	273 K 1.013 bar			ideal gas		293 K 1.013 bar
value	1.293	0.773	1005	718	287	1.4	0.026

To simplify the graphic representation and calculation, special assumptions are made to obtain two-dimensional state equations or family of curves. They are called isobaric if p = const., isochoric if v = const., isothermal if T = const. and isentropic if s = const. (or no heat loss). Real changes of gas state are polytropic and are characterized by the polytropic state equation between two states

$$p_1 v_1^n = p_2 v_2^n \qquad (1 \le n \le \kappa) \tag{10.5.7}$$

Then, the special changes of state follow with $n = 0$ for isobaric, $n = \infty$ for isochoric, $n = 1$ for isothermal and $n = \kappa$ for isentropic state changes. Applying the specific gas law (10.5.2) yields

$$\frac{v_1}{v_2} = \left(\frac{p_2}{p_1}\right)^{\frac{1}{n}} = \left(\frac{T_2}{T_1}\right)^{\frac{1}{n-1}} \tag{10.5.8}$$

and the external work becomes

$$W_{12} = m \int_1^2 p\, dv = m(T_1 - T_2) R/(n-1)$$
$$= m p_1 v_1 \left[1 - (p_2/p_1)^{(n-1)/n}\right]/(n-1) \tag{10.5.9}$$

and the technical work

$$W_t = m\int_1^2 v\,dp = n\,W_{12} \qquad (10.5.10)$$

Now, a certain closed air volume V is considered (like on one side of a cylinder with piston). If the compressibility module or elasticity module is defined as for hydraulic oil, (10.4.19)

$$E_{gas} = -V\left(\frac{\partial p}{\partial V}\right) \qquad (10.5.11)$$

then it follows from the polytropic state equation

$$\begin{aligned} p V^n &= K_{pol} \\ \frac{\partial p}{\partial V} &= -\frac{K_{pol} n}{V^{n+1}} \\ E_{gas} &= \frac{K_{pol} n}{V^n} = np \end{aligned} \qquad (10.5.12)$$

Hence, the compressibility module only depends on the pressure p and the polytropic exponent n.

The stiffness of the closed air volume in a cylinder with area A results in

$$c_{gas} = \frac{dF}{dz} = A\frac{dp}{dz} = A\frac{dp}{dV}\frac{dV}{dz}$$

The volume is

$$V = V_0 - A_Z$$

and with $dV/dz = -A$, it holds that

$$c_{gas} = -A^2\frac{dp}{dV}$$

Introducing (10.5.11) leads to

$$c_{gas} = -A^2\frac{E_{gas}}{V} = A^2\frac{np}{V} = \frac{A}{z}np \qquad (10.5.13)$$

Therefore the stiffness is inverse proportional to the deflection z and proportional to the pressure p at the operating point.

b) Pneumatic control valves

For manipulating the air flow, pneumatic control valves are used. The various types are the same or at least similar to those of Figures 10.13 to 10.15 and 10.17 for hydraulic systems. However, the flow characteristic through a valve is different, due to the compressibility of the air and sound velocity.

A gas flow from a container with pressure p_1 and temperature T_1 through a nozzle with rounded shape into a container with pressure p_2 and temperature T_2 is considered, see Figure 10.30. It is assumed that the flow is without friction and without heat exchange with the environment, hence an isentropic flow. Then it follows from the energy balance that

$$h_1 + \frac{v_1^2}{2} = h_2 + \frac{v_2^2}{2} \tag{10.5.14}$$

With the assumption for the velocities $v_1 << v_2$, it holds that

$$\frac{v_2^2}{2} = h_1 - h_2 = c_p(T_1 - T_2)$$

Introducing the gas equation, specific heat capacity and isentropic state equation

$$T_1 = \frac{p_1}{R\rho_1}; \qquad \frac{c_p}{R} = \frac{\kappa}{\kappa - 1}; \qquad \left(\frac{T_2}{T_1}\right) = \left(\frac{p_2}{p_1}\right)^{\frac{\kappa-1}{\kappa}}$$

leads to the outflow velocity

$$v_2 = \sqrt{2\frac{\kappa}{\kappa-1}\frac{p_1}{\rho_1}\left[1 - \left(\frac{p_2}{p_1}\right)^{\frac{\kappa-1}{\kappa}}\right]} \tag{10.5.15}$$

The mass flow follows with $\dot{m} = A\rho_2 v_2$ (10.5.8)

$$\dot{m} = A\psi\sqrt{2p_1\rho_1} = A\psi p_1\sqrt{\frac{2}{RT_1}} \tag{10.5.16}$$

where the outflow function is

$$\psi = \sqrt{\frac{k}{\kappa-1}\left[\left(\frac{p_2}{p_1}\right)^{\frac{2}{\kappa}} - \left(\frac{p_2}{p_1}\right)^{\frac{\kappa+1}{\kappa}}\right]} \tag{10.5.17}$$

ψ depends on p_2/p_1 and has a maximum at

$$d\psi / d(p_2/p_1) = 0$$

The pressure ratio p_2/p_1 at the maximum outflow function is called the critical pressure ratio and is

$$\left(\frac{p_2}{p_1}\right)_{crit} = \left(\frac{2}{\kappa+1}\right)^{\frac{k}{\kappa-1}} = 0.53 \tag{10.5.18}$$

for air. Then, the maximum of the outflow function becomes

$$\psi_{\max} = 0.484$$

Figure 10.31 shows the outflow function ψ (p_2/p_1). For constant p_1 and decreasing p_2, the outflow function increases until ψ_{max} and holds this value also for smaller p_2 because the sound velocity

$$v_{2crit} = a = \sqrt{\frac{2\kappa}{\kappa+1}\frac{p_1}{\rho_1}} = \sqrt{\frac{2\kappa}{\kappa+1}RT_1} \tag{10.5.19}$$

is reached. Hence, for overcritical $p_2/p_1 > 0.53$, the mass flow $\dot{m}$ only depends on p_1 and T_1, see Backé (1986a).

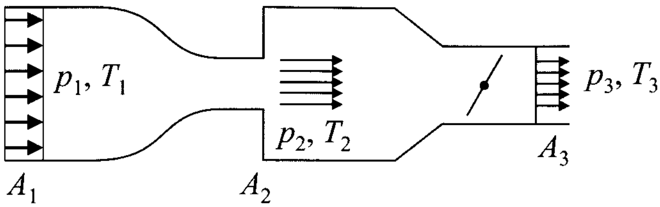


Figure 10.30. Gas flow through a contraction (nozzle)

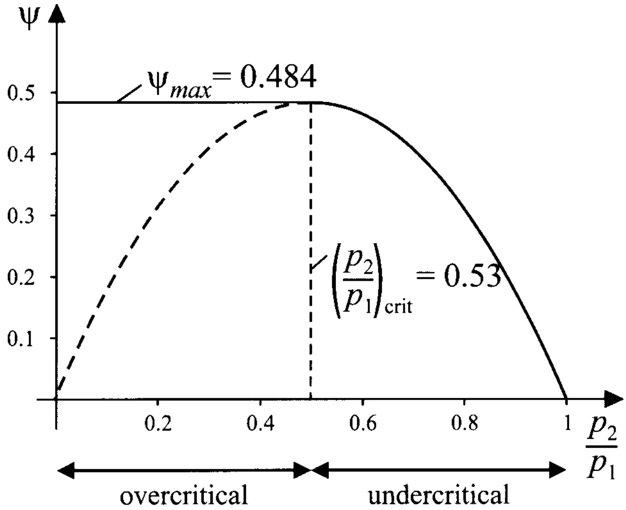


Figure 10.31. Outflow function for air with $\kappa = 1.4$

For practical use, an approximation of (10.5.16) can be used, ISO/DIN 6358

$$\dot{m} = \left\{ \begin{array}{ll} Acp_1 & \frac{p_2}{p_1} < b \\ Acp_1\sqrt{1-\left(\frac{p_2/p_1-b}{1-b}\right)^2} & b \le \frac{p_2}{p_1} \le 1 \end{array} \right\} \tag{10.5.20}$$

With this equation, pneumatic resistances can be described, where the parameters c and b are determined experimentally, Minxue *et al.* (1986). Figure 10.32 shows the resulting characteristics for a flow valve. The cross-sectional area A of the valve depends on the manipulated variable U.

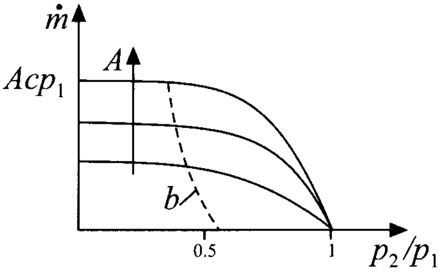


Figure 10.32. Flow characteristic of a pneumatic valve

c) Pneumatic pressure storages (accumulators)

A gas storage with volume V_S according to Figure 10.33 with a gas mass flow $\dot{m}_1(t)$ at the inlet and $\dot{m}_2(t)$ at the outlet is considered. It is assumed that no flow resistance occurs, *i.e.*, $p_1(t) = p_2(t) = p_s(t)$. The mass flow balance then yields

$$\dot{m}_1(t) - \dot{m}_2(t) = \frac{d}{dt} m_S(t) = \frac{d}{dt} V_S \rho(t) = V_S \frac{d\rho(t)}{dt} \tag{10.5.21}$$

If a polytropic gas equation holds

$$p v^n = p \left(\frac{1}{\rho} \right)^n = k_{pol}$$

then

$$\rho_S = \left(\frac{1}{k_{pol}} p_S \right)^{\frac{1}{n}}$$

and

$$\frac{d\rho_S}{dt} = \frac{1}{n k_{pol}^{\frac{1}{n}}} p_S(t)^{\frac{1}{n}-1} \frac{dp_S}{dt}$$

Thus, it leads to

$$\dot{m}_1(t) - \dot{m}_2(t) = \frac{V_S}{n k_{pol}^{\frac{1}{n}}} p_S(t)^{\frac{1}{n}-1} \frac{dp_S}{dt}$$

and with

$$m_S(t) = V_S \rho_S(t) = V_S \left(\frac{1}{k_{pol}} p_S(t) \right)^{\frac{1}{n}}$$

it becomes

$$\dot{m}_1(t) - \dot{m}_2(t) = \frac{m_S(p)}{n} \frac{1}{p_S(t)} \frac{dp_S(t)}{dt} \tag{10.5.22}$$

The gas volume therefore has non-linear integral behavior with a pressure-dependent integration time

$$T_I(p_S) = \frac{m_s(p_S)}{np_S} = \frac{V_S}{n k_{pol}^{\frac{1}{n}}} p_S^{\frac{1}{n}-1} \tag{10.5.23}$$

Only for isothermal state with $n = 1$ the mass balance becomes linear with $T_I = V_S k_{pol}^{-1}$

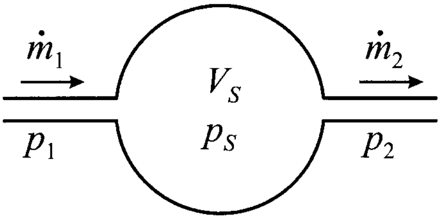


Figure 10.33. Schematic of a gas storage (accumulator)

d) Pneumatic valve-accumulator elements and transmission lines

A valve and a storage are now connected to a *valve-accumulator element* as shown in Figure 10.34a. The cross-sectional area A of the valve can be changed by the manipulated variable *U*. Figure 10.34b represents a two-port representation, assuming that the valve is supplied by a pressurized tank source. (Here, the volume flows $\dot{V} = \dot{m}/\rho$ are used according to the power variable definitions of Table 2.3.) Assuming $\dot{m}_2 = \dot{m}_1$ and $p_2 = p_S = p_3$, it follows that with (10.5.16) and (10.5.22)

$$\dot{m}_1 = A(U)\,\psi\left(\frac{p_3}{p_1}\right) p_1 \sqrt{\frac{2}{RT_1}} \tag{10.5.24}$$

$$\dot{m}_1(t) - \dot{m}_3(t) = \frac{1}{T_I(p_3)} \frac{dp_3(t)}{dt} \tag{10.5.25}$$

This leads to the block diagram shown in Figure 10.35 showing several non-linearities through multiplications and characteristics.

A simplification of the non-linear relation can be made for small changes around a certain operation point. This leads to the valve equation

$$\begin{aligned} \Delta\dot{m}_1 &= \frac{\partial \dot{m}_1}{\partial p_1}\Delta p_1 + \frac{\partial \dot{m}}{\partial p_2}\Delta p_2 \\ &= A\left[c_1 \Delta p_1 + c_2 \Delta p_2\right] \end{aligned} \tag{10.5.26}$$

Another, simpler approach is to use the pressure loss equation (10.4.18) for the turbulent flow of orifices in the case of undercritical pressure ratio.

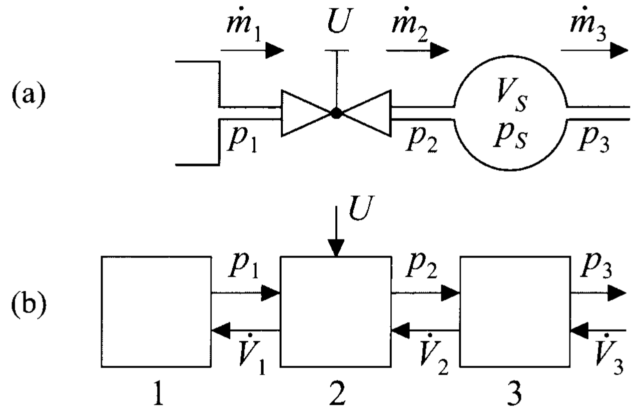


Figure 10.34. Pneumatic valve-accumulator element: (a) schematic; (b) two-port representation. 1: source; 2: valve; 3: accumulator

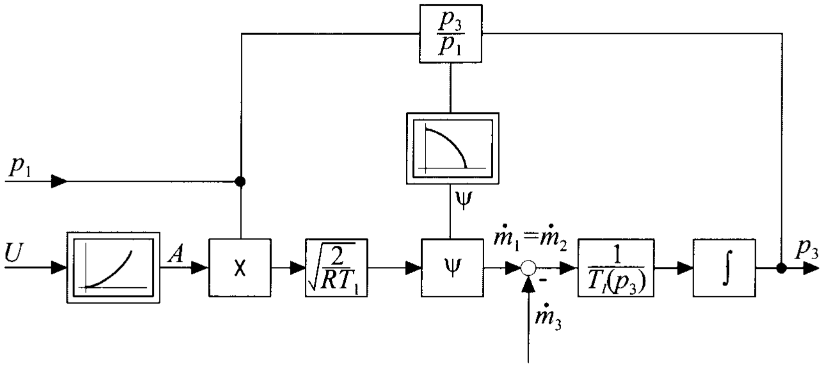


Figure 10.35. Signal flow diagram of the valve-accumulator element of Figure 10.34

Then, it holds that

$$\dot{m}_1 = A\sqrt{\frac{2\rho}{\zeta}}\sqrt{p_1 - p_2} \tag{10.5.27}$$

Linearization of this equation around a certain operation point $(A, p_1, p_2, \dot{m}_1)$ leads to

$$\Delta\dot{m}_1 = \frac{\partial\dot{m}}{\partial p_1}\Delta p_1 + \frac{\partial\dot{m}_1}{\partial p_2} - \Delta p_2 + \frac{\partial\dot{m}_1}{\partial A}\Delta A \tag{10.5.28}$$

with

$$\frac{\partial\dot{m}_1}{\partial p_1} = -\frac{\partial\dot{m}_2}{\partial p_2} = Ac_3 = A\sqrt{\frac{\rho}{2\xi}}(p_1 - p_2)^{\frac{1}{2}}$$

$$\frac{\partial \dot{m}_1}{\partial A} = c_4 = \sqrt{\frac{2\rho}{\xi}}\sqrt{p_1 - p_2}$$

Finally, with linearization of the valve characteristics

$$\frac{\partial A}{\partial U} = c_5$$

and $\Delta p_2 = \Delta p_3$, a linearized signal flow diagram as given in Figure 10.36 results. After Laplace transformation, one obtains

$$\Delta p_3 = \frac{1}{Ts+1}\left[\frac{c_4 c_5}{A c_3}\Delta U(s) + \Delta p_1(s) - \frac{1}{A c_3}\Delta \dot{m}_3(s)\right] \tag{10.5.29}$$

Hence, first order transfer functions with time constant

$$T = \frac{1}{A c_3} T_I = \frac{m_s}{p_3 A c_3} \tag{10.5.30}$$

result, assuming an isothermal state change with $n = 1$. Therefore, a valve-accumulator can be roughly described by a first order lag element if only small deviations are considered and undercritical pressure ratio across the valve applies.

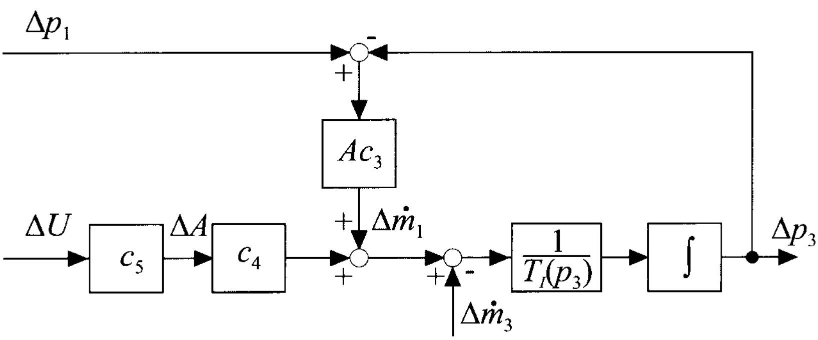


Figure 10.36. Signal flow diagram of the linearized behavior of the valve-accumulator element of Figure 10.34 for an undercritical pressure ratio

For an overcritical pressure ratio, the mass flow through the valve does not depend on the pressure p_3 in the storage and the transfer behavior becomes

$$\Delta p_3(s) = \frac{1}{T_I s}\left[c_4 c_5 \Delta U(s) + A c_3 \Delta p_1(s) - \Delta \dot{m}_3(s)\right] \tag{10.5.31}$$

which means integral behavior.

Pneumatic transmission lines can be considered as distributed resistance-volume elements. For not-too-long lines, a lumped parameter approach is applicable as shown in Figure 10.35 and the linearized model (10.5.29) with $\Delta U = 0$ can be used.

e) Pneumatic translatory motors

Pneumatic translatory motors have the same construction principle as shown in Figure 10.18 for hydraulic cylinders. Also, the same equations (10.4.45) for the volume flows hold if the compressibility module β of oil is replaced by the compressibility module of gas $E_{gas} = np$, (10.5.12)

$$\dot{p}_1(t)\frac{V_{01}+A_1 z(t)}{np_1(t)} + A_1\dot{z}(t) = \dot{V}_1(t)$$

$$\dot{p}_2(t)\frac{V_{02}-A_2 z(t)}{np_1(t)} - A_2\dot{z}(t) = \dot{V}_2(t) \tag{10.5.32}$$

If the mass flow $\dot{m}(t)$ is used instead of the volume flow $\dot{V}$, one obtains with $m = V\rho$ and the gas equation (10.5.1b)

$$\dot{p}_1(t)\left[V_{01}+A_1 z(t)\right] + np_1(t)A_1\dot{z}(t) = nRT_0\dot{m}_1(t)$$

$$\dot{p}_2(t)\left[V_{02}-A_2 z(t)\right] - np_2(t)A_2\dot{z}(t) = nRT_0\dot{m}_2(t) \tag{10.5.33}$$

where T_0 is a reference temperature, *e.g.*, the supply air temperature. For isothermic state changes, $n = 1$ can be assumed, and the dynamic behavior of the pressure in chamber 1 becomes

$$\dot{p}_1(t) + \frac{A_1}{V_{01}+A_1 z(t)}\dot{z}(t)p_1 = \frac{RT_0}{V_{01}+A_1 z(t)}\dot{m}_1(t) \tag{10.5.34}$$

Hence, the parameters of this pressure differential equation are time-variant and depend on the mechanical motion of the piston. The resulting signal flow for this pneumatic part is shown in Figure 10.37, left cylinder chamber.

The pressures in both cylinder chambers generate the force on the piston. The balance of forces on the piston rod is the same as for hydraulic cylinders and follows (10.4.48). Finally, a complete signal flow diagram can be drawn as in Figure 10.37, Keller (1994). The servo-valve with position controller can be approximated by a first order lag with time constant T_{valve}.

10.5.3 Model-based Control of a Pneumatic Servo-axis

Pneumatic cylinder-piston actuators are mainly used for transportation and assembling tasks. Due to their non-linear behavior, they operate predominantly with limit stops or brakes. The non-linear static and dynamic behavior results from the compressibility of the air, the friction of the piston and the characteristics of the electromechanical proportional control valves. Therefore, precise motion control is rather involved and does not result in sufficient control performance with linear controllers only.

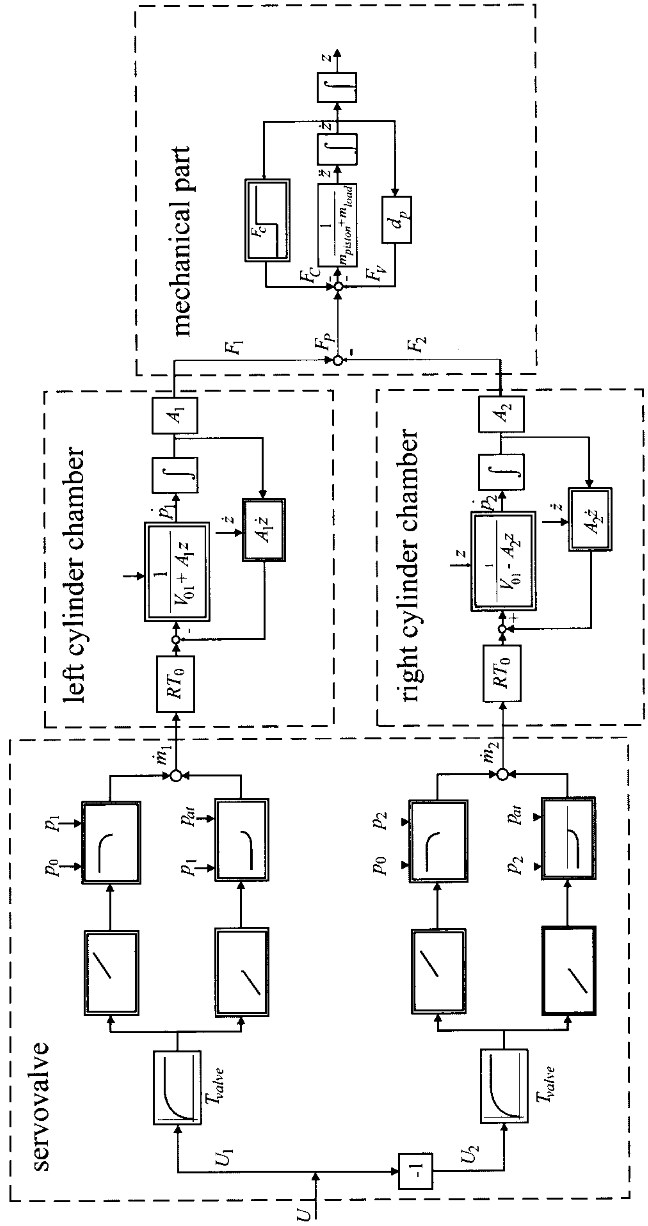


Figure 10.37. Signal flow diagram of a pneumatic servo-axis, Keller (1994). p_0: supply pressure; p_{at}: atmospheric pressure

In the following, it will be shown how through a non-linear model-based adaptive control system it becomes possible to obtain a relatively high position accuracy, Keller (1994).

The investigated pneumatic actuator is a standard pneumatic cylinder as shown in Figure 10.38. The motion transmission of the outer roller is obtained by magnetic coupling. The positioning range is 200 mm and the piston diameter is 25 mm, which leads to a force of 213 N by applying a pressure of 6 bar.

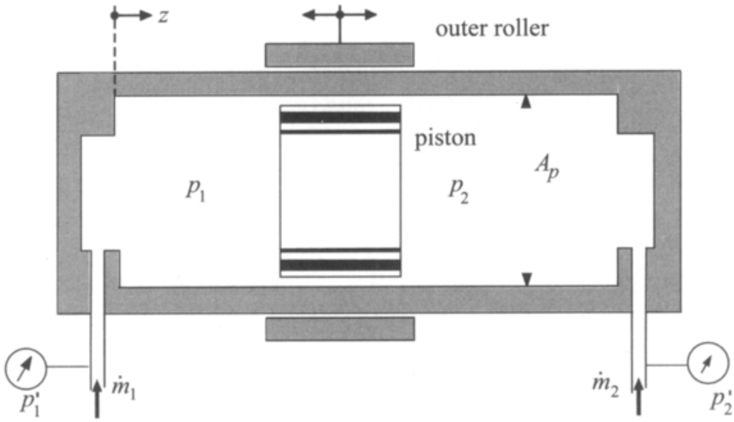


Figure 10.38. Scheme of the investigated pneumatic cylinder

Process input is a voltage V that manipulates the mass flows, either $\dot{m}_1$ or $\dot{m}_2$ into the cylinder chambers through a control valve. Then, either chamber 2 or 1 is connected to atmospheric air. The position z can be measured by a linear potentiometer. In addition, the pressures p_1' and p_2' at the cylinder tube connections are available.

Theoretical modeling leads to a model with several dynamic non-linearities as is shown in Figure 10.37. Here, $A_1 = A_2 = A_p$. Therefore, a static non-linear correction only will not work. However, the implementation of an underlying differential pressure control loop with a simple P-controller is able to generate an adjustable input force F_p to the piston. This control then requires the knowledge of the difference pressure $\Delta p = p_1 - p_2$.

A further significant non-linearity of this actuator is the friction. The typical stiction force is roughly 45 N (21% of the nominal input force) and values of about 30 N have been estimated for Coulomb friction via system identification. Therefore, friction compensation is required. Further, because the friction force changes with time, the friction compensation should be adaptive.

Figure 10.39 depicts the implemented overall control strategy with state controller, underlying differential pressure control and adaptive feedforward friction compensation. The pressure difference control of $\Delta p = p_1 - p_2$ is based on the model-based reconstruction of the chamber pressure p_1 and p_2 based on the measurements of p_1' and p_2'. Furthermore, the position of z *is* measured and the speed $\dot{z}$ and acceleration $\ddot{z}$ are determined by numerical differentiation.

Feedforward friction compensation, as is shown in Figure 10.9, yields continuous oscillations with an amplitude of at least 0.5 mm. This occurs due to the non-negligible dynamics of the underlying difference pressure loop. Therefore, the compensation has to be switched off if the control variable z lies within a tolerance band (± 0.05 mm) of the set point r.

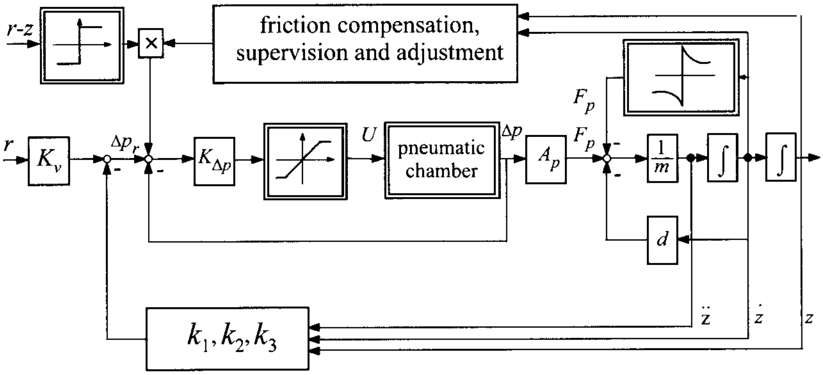


Figure 10.39. Overall structure for the adaptive non-linear positioning control of a pneumatic actuator with friction compensation

As several investigations have shown, *e.g.*, Rusterholz (1985), Chen, Leufgen (1987), friction forces of pneumatic cylinders are highly position-dependent and vary with the applied pressure as well as with the time period of standstill. Therefore, under- and overcompensation is very likely and supervision of the steady state control behavior must take place in order to adapt the amplitude of the compensation values. The amplitude will be reduced if oscillations are detected and increased if the control value does not lie in the set point tolerance band. This is done by "friction compensation supervision and adjustment" (FCSA), Keller (1994). Figure 10.40 shows the improvement by friction compensation.

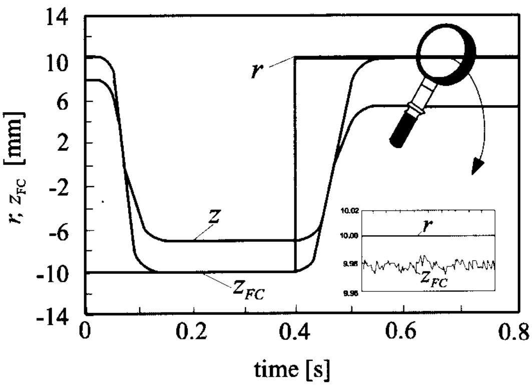


Figure 10.40. Comparison of the position control performance without and with (index$_{FC}$) adaptive friction compensation, $T_0 = 1$ ms, $m = 3.5$ kg

The state controller values (k_1, k_2, k_3) are obtained by numerical optimization of a quadratic loss function using a non-linear system model. Without friction compensation, an offset of about 3 mm is obtained. Applying friction compensation, the positioning accuracy is significantly improved to about 0.05 mm.

This is an example where the negative properties of the pneumatic actuator are compensated for a considerable degree by model-based control. Hence, integration by information processing takes place to result in an appropriate overall behavior of the pneumatic mechatronic actuator.

10.5.4 Models of a Pneumatic Valve

Fluid flow is frequently controlled by means of pneumatically driven valves. These pneumatic valves are composed of two units: a pneumatic and a mechanical subsystem, Figure 10.41. The pneumatic subsystem comprises a chamber sealed by a diaphragm that is acting on the valve stem. At the tip of the valve stem, a body is mounted which, in conjunction with its counterpart, the valve seat, controls the hydraulic flow. Depending on the precision accuracy and the kind of fluid, different geometries are used. For very precise control tasks, usually needle-shaped bodies are used, whereas disc- or ball-shaped bodies are commonly chosen for fully opening and closing valves. The stem passes through a gland in order to seal the hydraulic system. Figure 10.41 shows a cross-sectional view of such a valve. A position controller is mounted directly on the valve.

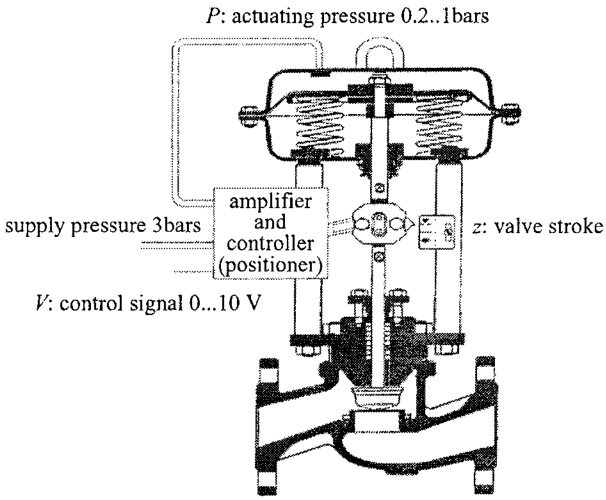


Figure 10.41. Cross-sectional view of a pneumatic valve

The valve contains an inner pneumatic control loop, which varies the displacement of the valve stem in accordance with the input voltage. By means of a nozzle-flapper arrangement, the control error between the valve stem position and its reference value is sensed and the pressure supplied to the diaphragm chamber is varied accordingly. Since the nozzle-flapper arrangement would not be able to supply a sufficient air flow, an air amplifier is connected in between the nozzle-flapper arrangement and the working chamber.

For modeling of the valve, the same equation as for the pneumatic cylinder can be used if A_D is the area of the diaphragm, z the position of the valve stem and c_s the constant of the return spring. Figure 10.42 shows the resulting signal flow diagram. For more details see Deibert (1997).

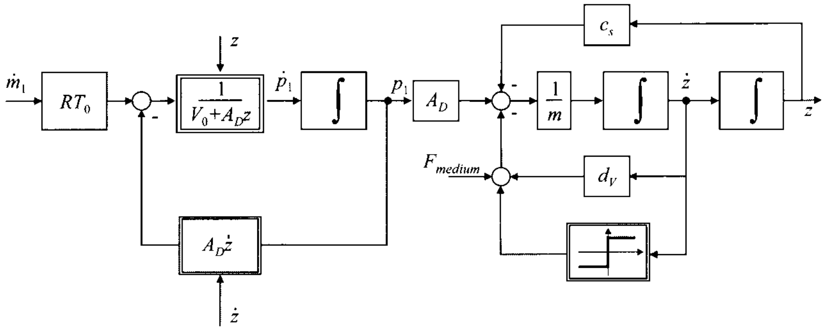


Figure 10.42. Signal flow diagram of a pneumatic flow valve

Usually, the hysteresis effect of the friction of the gland is compensated for by a position controller. However, if no position controller can be applied, limit cycles can appear in the closed-loop flow control. In this case, an adaptive friction compensation as shown in the last Section 10.5.4 can improve the control behavior considerably. This was demonstrated by Schaffnit (2002) for the position control of a pneumatic actuator for a variable geometry turbocharger of a diesel engine.

10.6 UNCONVENTIONAL ACTUATORS

Over the past few years, a number of new concepts for "unconventional actuators" have been developed and existing designs have been improved, which can be attributed to both the on-going research in the area of materials science and the application of modern manufacturing technologies. The commonality of these actuators is the fact that they use certain physical phenomena. However, the different technical realizations exhibit a large degree of specialization, which leads to a limited scope of application. The high cost of these materials further limits the spread of these actuators. Presently, especially the development of piezoelectric actuators or traveling-wave motors show an interesting prospect for the future.

A survey of unconventional actuating principles is summarized in Table 10.28. The first main group consists of the so-called direct energy converters, such as piezoelectric, electroviscous and magnetostrictive actuators. They generate a force by changes in the atomic/molecular structure upon energization by an electric input signal. Their main area of application is for fast and highly precise positioning tasks with a very small stroke.

Table 10.28. Overview of unconventional actuators (normal: translatory motion; *italic: rotary motion*)

generation of force	actuator	technical realization
molecular forces (direct energy converters)	piezoelectric actuator	stack design *bending actuator* *traveling- wave motor* *inchworm motor*
	magnetostrictive actuator	linear actuator *inchworm motor*
	electroviscous fluid	adjustable dampers *clutches*
shape memory effect	shape memory alloys	*bending actuator* *torsional actuator*
thermal expansion	thermo-bimetal	*bending actuator*
	thermal expansion elements	membrane actuator elastomer actuator
chemical reaction forces	electromechanical actuators	pyrotechnic actuator
	electrochemical actuator	membrane actuator

Shape memory alloys and thermal expansion actuators do not necessarily need an auxiliary power supply. The actuation energy can be gathered from the surroundings (*e.g.*, ambient heat for the control of a radiator).

In the following, some concepts for unconventional actuators will be examined in more detail. The interested reader is also referred to survey articles by Janocha (1992), Lenz *et al.* (1990), Tautzenberger (1989). Introductions to smart materials are given by Srinivasan, McFarland (2000) and Culshaw (1996). The technology field of smart materials refers to active materials. Upon one or more input stimuli, these materials change their shape or exchange energy with the surroundings. This characterizes these materials as energy converters. In the literature, these materials are referred to as smart materials, active materials, as well as adaptive materials interchangeably. Typically, the group of active materials encompasses piezoelectric actuators, magnetostrictive actuators, electrostrictive actuators, shape memory alloys and thermo-bimetals. Since the area of unconventional actuators is fairly new, the aforementioned definitions are still evolving.

10.6.1 Thermo-bimetals

Thermal expansion describes the effect of materials changing their length if subject to a temperature change. This effect can be exploited for thermo-bimetals. If two materials with different coefficients of thermal expansion are joined together and the temperature changes, one metallic strip will contract or expand more than the other and thus the beam starts to bend. If the beam is encumbered in its bending motion, it can be used as a "force actuator". Instead of the motion, tension inside the thermo-bimetal starts to build up. Thus, the thermo-bimetal can store energy, comparable to a spring. The functional principle of thermo-bimetal actuators is shown in Figure 10.43.

The change in temperature of the thermo-bimetal may not only be caused by a change in the ambient temperature, but can also be evoked by means of heat radiation or convection as well as electrical heating. For the latter, one differentiates indirect heating by an electrical heating element, mounted close by the thermo-bimetal, and direct heating, where the current is conducted by the thermo-bimetal itself (*e.g.*, thermal fuses for all kinds of electrical machinery). A summary of the material properties, the areas of application as well as the major features of thermo-bimetals is given in Table 10.29.

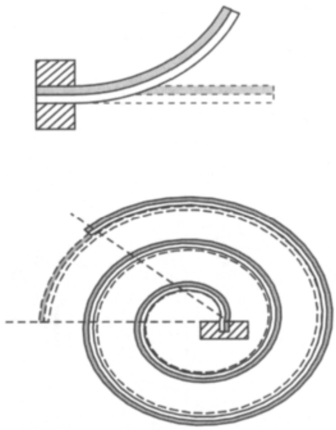


Figure 10.43. Thermo-bimetal actuators

Table 10.29. Main features of thermo-bimetals, Rau (1974)

design principle	change of shape due to different thermal expansion of two metals
advantages	disadvantages
• available in different sizes and shapes; • inexpensive; • linear temperature-displacement relationship over a wide range of temperatures.	• small actuation forces; • only bending motion can be realized directly; • small energy density.
areas of application	• electric toaster; • fuse.
physical properties (according to DIN 1715, *e.g.*, for TB 1425)	applicable up to 450°C specific curvature $26.1 \cdot 10^{-6}$ 1/K Young's modulus $170 \cdot 10^{3}$ N/mm^2 maximum bending stress 200 N/mm^2 specific resistance $0.26 \cdot 10^{-6}$ Ω m density $8.3 \cdot 10^{3}$ kg/m^3 linear range -20°C to 200°C

10.6.2 Shape Memory Alloys

Shape memory alloys are materials that can "remember" shapes, *i.e.*, they can return to some previously defined shape when subject to a certain thermal procedure. This behavior is known as the *shape memory effect* and arises from a change in phases. The change is between two solid phases and embraces a rearrangement of the atoms in the crystal lattice. Depending on the ambient temperature or applied stress and strain, the shaped memory alloy takes up one of two different crystalline configurations. *Martensite* is the low-temperature phase and shows a highly twinned crystalline structure, whereas *austenite* is the phase that exists at higher temperatures and is based upon a body-centered cubic structure. By applying external forces in the martensitic phase, the highly twinned structure can be de-twinned. The entire process is depicted in Figure 10.44.

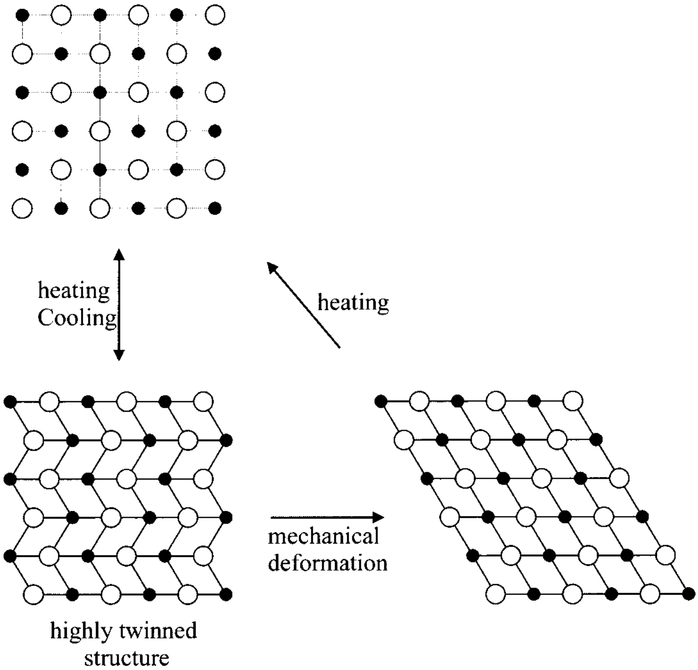


Figure 10.44. Crystal lattice of a shape memory alloy (one-way effect)

The shape memory alloy (SMA) can react to its cooling down in two different ways. With the so-termed one-way shape memory effect, elements just remain in the shape they had while being in the austenitic configuration. In contrast, the two-way shape memory elements assume their old shape again, thus they have been taught two different shapes, which can be recalled unlimitedly just by changing the temperature of the shape memory alloy. While both the one-way and the two-way elements can generate rather large forces during heating up, the force exerted by the two-way element during cooling down is rather limited.

The applicability of SMAs will be determined mainly by their transformation time and characteristic temperature. There are a magnitude of different shape memory alloys available today with transition temperatures between −100°C and +100°C. Most alloys are nickel-titanium or copper-zinc-aluminum alloys. A summary of the features of shape memory alloys is given in Table 10.30. This table also lists material properties of nickel-titanium alloys. The temperature change of the shape memory alloy can be evoked by different means. First of all, an electric heater can be mounted on the elements. The element may also be heated up by a current flowing through the element itself. Also, the element can dissipate heat from the surroundings, such as, *e.g.*, from a fluid the shape memory element is immersed into. More problematic, however, is the cooling down of the material. Here, one has two competing design goals. During static phases, the energy consumption should be minimized, which mandates good insulation of the SMA. During the cooling down, however, the heat energy should be purged as fast as possible, which is hindered by the insulation. Another possibility would be the introduction of an active cooling element. This, however, makes the actuator more bulky.

Table 10.30. Main features of shape memory alloys

actuating principle	change of shape during transitions from and to martensitic/austenitic crystal configuration in certain alloys
advantages	disadvantages
• good temperature response–full transformation over temperature range as small as 10°C; • can exert high forces during heating up; • shape can change in many ways (expansion, contraction, bending motion, ...); • large number of different SMAs with transition temperatures between –100C and +100C available.	• very sensitive to composition variations and fabrication; • Ni-Ti alloys are expensive (Cu-based alloys are much cheaper); • two-way elements can only exert small forces during cooling down; • slow dynamics; • strong hysteresis.
areas of application	actuation of safety-related devices valve and flap actuation clamping and locking
physical properties (given for NiTi alloys) Hodgson (1989)	melting point 1300°C density 6.4 kg/m^3 Young's modulus austenite 82×10^9 N/m^2 martensite 31×10^9 N/m^2 yield strength austenite 690×10^6 N/m^2 martensite 150×10^6 N/m^2 ultimate strength 900×10^6 N/m^2 transformation temperature –200...100 °C maximum shape memory strain one-way 8.4% maximum shape memory strain two-way 3%

Due to their behavior, SMAs are well suited to switching applications. Here, their capability of creating large actuation forces and their temperature response characteristics come into favor. On the other hand, SMAs are rather unsuitable for applications where a continuous motion or fast response times are required. Although the heat-up cycle can be expedited by using a larger heating element, the cool-down cycle remains unchanged, leading to long manoeuver times. Also, the strong hysteresis contradicts the use as a continuously variable actuator.

The first practical applications of SMAs as switching type actuators date back to the 1970s. Nowadays, areas of application include actuators for vanes that control the flow of air through jet engines and window latches that open and close automatically. Another interesting application is as a blood clot filter. This opens up when inserted into the blood vein and heated up to body temperature.

The most important characteristics of SMAs are outlined in Table 10.30. More detailed information can be found in Waram (1993).

10.6.3 Thermal Expansion Elements

These actuators, just like thermo-bimetals, are based on thermal expansion. While the latter exploit the change in length due to a change in temperature, thermal expansion elements rely on the change in volume associated with high thermal coefficients of certain solid and fluid materials. Upon an increase in temperature, the volume of an enclosed material amount will grow. The material is usually contained in some sort of cylinder and gives rise to the movement of a piston upon expansion. The functional principle is depicted in Figure 10.45. Depending on the material used, these thermal expansion elements have different temperature displacement characteristics. In general, fluids show a better linearity in the temperature-displacement characteristics than solids. Expansion elements are typically controlled by the ambient temperature only. However, there also exist versions with an attached electric heating element. A typical area of application is the actuation of valves (*e.g.*, radiators). The major advantages and disadvantages are listed in Table 10.31.

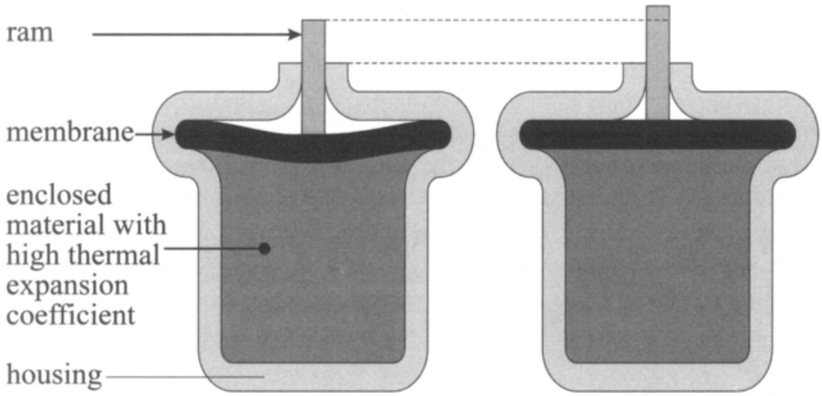


Figure 10.45. Cross-sectional view of thermal expansion elements (membrane actuator)

Table 10.31. Main features of thermal expansion elements, Janocha (1992)

actuating principle	movement of a piston due to thermal expansion of an enclosed material
advantages	disadvantages
• mechanically robust; • inexpensive; • large displacement and actuation force.	• mediocre dynamic behavior; • limited thermal range of application (-20°C to +150°C).
areas of application	actuation of valves
physical properties	displacement 5...15 mm maximum actuation force 250...1500 N

10.6.4 Electrochemical Actuators

The electrochemical actuator exploits certain electrochemical reactions, which, evoked by a direct current, lead to the development of gas. This gas is contained within an enclosed volume and exerts a force on a membrane. By either reversing the current flow or shortening the cathode and the anode, the chemical reaction can be reversed and the gas will be dissolved again, resulting in a reduction of the pressure inside the sealed volume. This actuator is still at an experimental stage, nevertheless it has been included to show the variety of possible actuation principles that can be employed for mechatronic systems. Due to this multifariousness, the designer of a mechatronic system will experience a great latitude of different actuators to choose from. Some expected features of these newly proposed electrochemical actuators are listed in Table 10.32.

Table 10.32. Main features of electrochemical actuators, Gevatter (2000)

actuating principle	gas generated by electro-chemical reaction displaces a membrane
advantages	disadvantages
• no energy consumption during static phases; • no moving parts; • retraction possible without external power (fail-safe); • noiseless.	• slow; • strong variations in the time-displacement behavior necessitate closed-loop control; • not much experience with this actuation principle available so far.
areas of application	still at experimental status
physical properties	displacement 5...16 mm actuation force 300...3000 N response time ...150 s rated voltage 12...36 V rated current 0.3 ... 1 A operating temperature -5...60 °C

10.6.5 Electro-rheological and Magneto-rheological Fluids

The fluids considered in this section change their rheological properties in response to an applied electric or magnetic field. The main focus of this section will be on electro-rheological (ER) fluids, but, due to their close resemblance, magneto-rheological (MR) fluids will also be considered. From their chemical composition, ER fluids are suspensions of non-metallic hydrophilic solid particles (around 1 to 10 μm in size) along with adsorbed water in an inert carrier liquid. Typically, certain additives are also added as to improve the stability of the structure or to adjust the physical properties. From an engineering point of view, however, these additives need not be considered.

In the case of zero-field, ER and MR fluids are modeled as ordinary Newtonian fluids, with the shear stress increasing proportionally to the strain rate. In reality, this assumption may not be true. The deviation from the ideal Newtonian model can be attributed to the heavy particles suspended in the fluid. More interesting than their behavior in the absence of electrical fields, however, is their behavior if subject to an electric or magnetic field. Then, the shear strain-velocity characteristics change remarkably. The underlying physical principle is still not thoroughly understood. It is believed that the electric field induces the almost instantaneous formation of chains – so-called *fibrils* – in the fluid, see Figure 10.46. Upon relative movement between the fluid and the walls of the enclosing vessel or upon a relative movement within the fluid, the fibrils will break and reform continuously, thus resisting such movements. This resistance gives rise to an offset of the shear-strain-velocity curve, resulting in a non-zero flow limit, also referred to as the *yield stress*. The physical model of ER and MR fluids, is for most applications, based on the Bingham plastic model. The shear stress is often modeled as increasing linearly with the field strength. Measurements indicate that the growth is normally proportional to the field strength raised to a power between 1 and 2. The upper limit of the electric field is typically given by a value of 4 kV/mm, above which the insulation of the fluid breaks down. For MR fluids, the yield stress is limited by saturation, typically occurring at a field strength of about 250 A/mm.

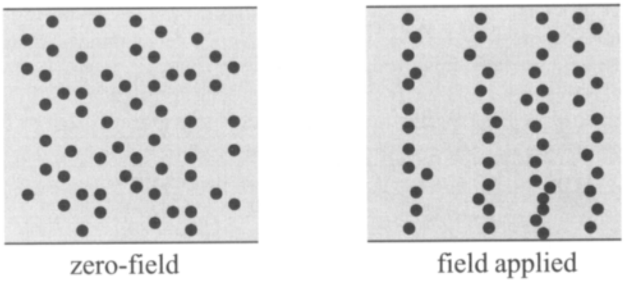


Figure 10.46. Formation of fibrils

For the technical realization, one can differentiate between two types of ER applications. Either the electrodes move relative to each other as is the case for ER clutches and some ER shock absorbers, or the fluid flows through the rigidly connected pair of electrodes. This set-up can also be used for shock absorbers, where the pair of electrodes, shaped as an orifice, represent a bypass valve. Since the electro-rheological effect is independent of the polarity of the external electrical field, electro-rheological fluids can be controlled by either a DC or an AC excitation. For AC excitation, however, the viscosity also becomes a function of the frequency of the applied electric field. For MR fluids, similar areas of application emerge, *e.g.*, magneto-clutch. Table 10.33 summarizes the features of ER and MR fluids. Information about electro-rheology can be found in Block, Kelly (1998) and Duclos *et al.* (1992). Applications of the magneto-rheological effect are described in Kpordonsky (1993).

Table 10.33. Main features of electro-rheological and magneto-rheological fluids

actuating principle	change of viscosity of certain fluids if subjected to an electric or magnetic field
advantages	disadvantages
• viscosity easily controllable; • fast response time.	• ER fluids in particular sensitive to water enclosures; • temperature-dependent properties; • sedimentation can become problematic; • not highly available; • can only evoke reaction forces and cannot generate primary actuation forces → semi-active actuators.
areas of application	adjustable shock absorber clutch
physical properties	electro-rheological fluids zero-field viscosity 100...1000 mPa/s maximum yield stress 2...5 kPa density $1\ldots2\cdot10^3$ kg/m^3 magneto-rheological fluids zero-field viscosity 100...1000 mPa/s maximum yield stress 50...100 kPa density $3\ldots4\cdot10^3$ kg/m^3

10.6.6 Piezoelectric Actuators

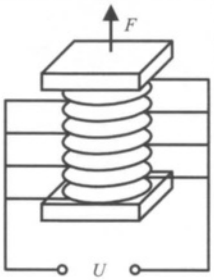
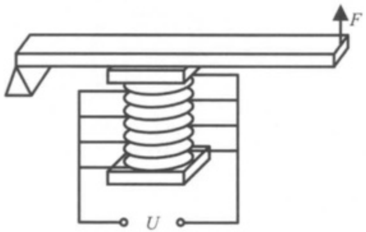
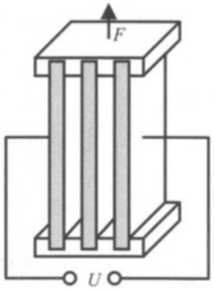
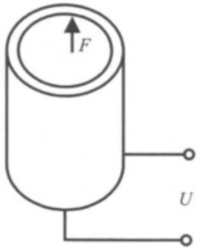
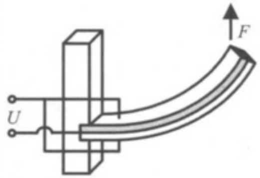
Certain crystals, *e.g.*, quartz, show a physical relationship between the mechanical stress and their electric charge. If the ions of the crystal lattice are displaced due to an externally applied mechanical stress, this displacement manifests itself as an electrical polarization of the crystal. The polarization can be measured by electrodes mounted on the crystal surface. The effect is termed the *direct piezoelectric effect* and is used, *e.g.*, for pressure and force transducers. The piezoelectric effect can also be reversed. Upon applying an electric voltage to a piezoelectric crystal, the crystal will change its thickness. This is termed the *reciprocal piezoelectric effect* and allows for the design of piezoelectric actuators. The best-known material that exhibits this effect is lead-zirconate-titanium (PZT), in fact, the term "PZT" is commonly used to refer to piezoelectric materials regardless of their actual chemical composition.

When first manufactured, the piezoelectric material has a random arrangement of the electric dipoles. It is not until the *poling* that the material exhibits the piezoelectric effect. In the process termed *poling*, the dipoles are permanently aligned. For this to happen, the material must first be heated above its Curie temperature, which typically lies between 120°C and 350°C. Above this temperature level, the dipoles can change their orientation in the solid phase material. If, in this state, the dipoles are subjected to a strong electric field, they will align themselves with this external field. If this field is maintained during the cooling down of the material, this alignment will be retained permanently. In the subsequent operation, special attention must be paid to the operating conditions, in particular to the ambient temperature, the compressible stress and the field strength of the electric field. If heated above the Curie temperature, the dipoles can again change their alignment and thus the crystal might lose its piezoelectric effect. Typically, the operating temperature is limited to 50 – 75% of the Curie temperature in Kelvin. Furthermore, the polarization can also vanish due to a strong electric field if this field is applied opposite to the direction of polarization. As a general rule, the electric field strength may not exceed 500 V/mm. Finally, polarization can also fade due to excessive compressible strain. Here, a limit is given as 100...150 N/mm^2.

There exists a variety of different shapes of piezoelectric actuators, which makes them well suited to different applications. Different PZT actuators, along with their respective specifications, are shown in Table 10.34. A few of them will be described in the following. A basic PZT actuator can only cover a very small range of realizable displacements. In order to make PZT-based actuators more appealing, the realizable displacement must be increased, which is accomplished by different means, such as stacking PZT actuators or using a lever arrangement.

The most commonly used type is the *stack design actuator*. Here, stacks of small ceramic discs (0.3 – 1 mm high) are stacked on top of each other and glued together. From a mechanical point of view, the individual actuators are thus connected in series. Their electrodes are connected in parallel. In order to avoid a current flow, which could be evoked from the applied electric field, the entire column is covered with highly insulating material. To achieve an even larger range of realizable displacements, the motion of the PZT can be amplified by means of a lever arrangement. Another constructional principle is the *laminar design*. For this kind of actuator, small stripes of PZT material are stacked. These actuators exploit the transversal piezoelectric effect, which means that the electric field is applied orthogonal to the desired change in length. These actuators are usually very flat, since the piezoelectric effect is more prominent with a larger ratio of length to thickness. Upon application of an electric voltage, this class of actuators actually shortens. There is also another constructional principle exploiting the longitudinal effect, the *bender-type actuator*.

Table 10.34. Technical designs of piezoelectric actuators, Jendritza (1998)

	transversal		longitudinal		
	stack design	stack design with integrated lever motion amplifier	laminar design	tube design	bender-type design
design					
displacement	20...200 μm	...1000 μm	...50 μm	...50 μm	...1000 μm
actuating force	...30000 N	...3500 N	...1000 N	...1000 N	...5 N
actuating force	60...200 V 200...500 V 500...1000 V	60...200 V 200...500 V 500...1000 V	60...500 V	120...10000 V	10...400 V

Somewhat similar to the thermo-bimetals, the design consists of two small beams that are joined together such that they cannot move independently of each other. If the two beams expand or contract differently, the entire structure will start to bend. For the unimorph design, a piezoceramic strip and a normal metallic strip are joined together, whereas for the bimorph design, two PZT strips are linked together. The PZT effect then induces a change in length of the ceramic strips.

For most applications, PZTs will be bonded to or embedded in a passive base structure. The piezoelectric effect can also be used for the design of linear motors, such as the inchworm motor.

The inchworm motor consists of three piezoelectric actuators. Two of these are clamping elements that can hold an axle that runs through these two clamping elements. A third PZT is used to displace these two clamping elements. It is rigidly supported in the middle. The motion of this motor is coordinated by an electronic control circuit and is shown in Figure 10.47. There, the behavior of the inch worm motor is detailed for movement of the rod to the right. At the beginning of one motion cycle (1), the middle element is relaxed and the left clamping element holds the beam. Then (2), an electric field is applied to the middle element, which contracts due to this input stimulus. The rod, which was tightly clamped by the left element, moves in accordance with the left clamping element. After the right clamping element has got hold of the rod, the left clamp is disengaged (3). Next, the electric field applied to the middle element is shut off (4). While the middle element expands to its original length, the clamping elements move back to their home positions. Since the right clamp moves to the right, the tightly attached rod will also move further to the right. Now, the whole cycle can be repeated (5). Due to the frictionally engaged connection of the clamping elements and the rod, the entire set-up must typically be operated in closed-loop control. The characteristics of piezoelectric actuators are listed into Table 10.35.

A detailed description of piezoelectric materials can be found in Cady (1964) and Jaffe *et al.* (1971). An outline is also given in Takuro (1996).

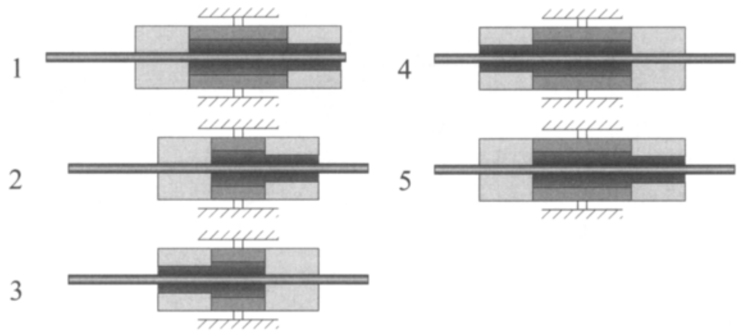


Figure 10.47. Operating principle of the inchworm motor, see description in text (not drawn to scale)

Table 10.35. Main features of piezoelectric actuators, Raab (1993)

actuating principle	displacement of ions in certain crystals upon application of an electric field
advantages	disadvantages
• large actuation forces; • high power–weight ratio; • fast response time; • negligible energy consumption during static phases; • linear field-strain characteristic; • ceramics can be configured in many shapes and are highly available; • almost no wear-out.	• characteristics vary with temperature changes and aging; • piezoeffect can be lost if crystal is subject to high temperatures, strong electric fields or mechanical shock; • high voltage supply necessary which must be able to drive capacitive loads; • open D-E loop implies high hysteresis and high losses, material-heat up.
areas of application	fuel injection valves
physical properties	maximum displacement 20...1000 μm static large-signal stiffness 75...1800 N/μm natural frequency 3.5...60 kHz maximum compressive strain ...800 N/mm² maximum tensile strain ...55 N/mm² rated voltage ...1500 V capacity ...6500 nF

10.6.7 Electrostrictive and Magnetostrictive Actuators

Similar to piezoelectric materials, *electrostrictive materials* are also predominantly made of ceramics, typically lead manganese niobate: lead titanate (PMN:PT) and lead lanthanum zirconate titanate (PLZT). In contrast to PZTs, no poling is necessary, since electrostriction occurs in virtually any material – although the effect is not strongly developed in some materials. Upon application of an electric field, the electric charges in the material attract each other, resulting in compression along the axis of the electric field but independent of the polarization of the applied field.

The strain is typically modeled as being proportional to the square of the applied field. Electrostrictive materials outperform PZTs in the smaller hysteresis loop of the strain-electric field relation. They exhibit less losses and can be operated at higher frequencies. Their major drawback is the quadratic dependency between strain and electric field. Their area of application will be the same as that of piezoelectric actuators.

Similar to the electrostrictive effect, *magnetostrictive materials* shrink in the presence of a magnetic field. This is caused by rotation of the magnetic domains and a shift of the Bloch walls. Magnetostrictives are made of alloys of iron, nickel and cobalt doped with rare earths. The complicated production process restricts the available sizes and shapes of magnetostrictives and also makes them very expensive. The main advantage of magnetostrictive actuators is their high energy density, which allows for higher actuation forces. However, only a very limited number of applications justify the use of these expensive materials. Typically, these are defence-related, *e.g.*, sonar “pingers”. Typical material properties of magnetostrictive and electrostrictive actuators can be found in Table 10.36.

Table 10.36. Main features of electrostrictive and magnetostrictive actuators, Raab (1993)

actuating principle	change in length of certain materials if subject to an electric or magnetic field
advantages	disadvantages
• high actuation forces; • electrostrictives: high efficiency due to closed D-E loop, can be operated at higher frequencies than PZTs; • high energy density; • fast response times; • actuator can be made of single piece of material; • electrostrictives: ceramics can be produced in many forms/shapes; • almost no wear.	• magnetostrictives: expensive and restricted availability (military applications); • energy consumption during static phases; • material properties are temperature-dependent; • non-linear strain-field characteristics.
areas of application	• sonar; • other applications still at experimental stage.
physical properties (given for magnetostrictive material TERFENOL-D)	maximum elongation ...$1200 \cdot 10^{-6}$ m/m Young's modulus 25...$30 \cdot 10^3$ N/mm^2 specific electric resistance $0.6 \cdot 10^{-6}\,\Omega$ m maximum compressive strain 700 N/mm^2 maximum tensile strain 28 N/mm^2 density $9.25 \cdot 10^3$ kg/m^3

10.6.8 Micro-actuators

Micro-actuators are those actuators whose functional components are manufactured employing production processes used in the area of micro-technology, such as etching and lithography. The functional principle of micro-actuators can be based on any of the physical effects that have been described in this chapter so far, such as electromagnetism, thermal expansion, piezoelectricity, electrostriction or magnetostriction. One physical effect that has been found to be extremely well suited to micro-actuators is electrostatic force. At the small scale of micro-actuators, the distance between the electrodes is in the range of micrometers, and so even normal transistor comparative voltages of less than 5 V can generate a field strength in the area of a few kV/mm. Secondly, the disruptive strength of isolators grows with a decrease in thickness, thus the maximum allowable electric field strength increases.

Most micro-actuators are manufactured from silicon, which has already been used extensively for the production of integrated circuits and is thus well researched. Furthermore, since silicon is also used for the production of integrated circuits, the actuator and the control circuitry can be combined and manufactured together, leading to so-called MEMS (micro-electrical-mechanical systems). Detailed information about MEMS can also be found in Gad-El-Hak (2002).

Micro-actuators are still at an experimental stage, so typical properties and areas of application have not yet been determined clearly, but a few advantages and disadvantages are given in Table 10.37.

Table 10.37. Main features of micro-actuators

actuating principle	very small-scale actuators using different actuation principles (predominantly electrostatic forces)
advantages	disadvantages
• inexpensive production using techniques well known from the manufacture of integrated circuits; • microelectronic-compatible voltage levels allow integration of actuator and controller on one wafer; • reliable; • cheap.	• due to the size of the actuator: only very small displacement and actuation forces realizable.
areas of application	still at experimental stage
physical properties	still at experimental stage

10.7. COMPARISON OF APPLICATION AREAS

In this section, diagrams are presented that allow the graphical comparison of the different actuators introduced in the preceding sections. These diagrams will focus on the following actuators:

- electric motors and stepper motors with matching gear;
- electromagnets;
- pneumatic and hydraulic cylinders;
- piezoelectric stack actuators.

The specifications have been taken from company data sheets and publications, Raab (1990,1993). They refer to translatory actuators for small to medium actuation power. The diagrams compare typical properties. The actuation force has been chosen as the common ordinate for all diagrams.

Figure 10.48 illustrates the maximum speed of the different actuators, which also allows for a comparison of the power output (product of speed and force). The maximum displacement is plotted in Figure 10.49. The lower limit is determined either by the smallest possible displacement or by the positioning accuracy. The upper limit represents the upper bound for the maximum displacement that can be reached by a certain kind of actuator. The product of force and maximum displacement determines the working capacity of the actuator. One can see that electric motors cover a large range of realizable displacements. Piezoelectric actuators offer the best positioning accuracy.

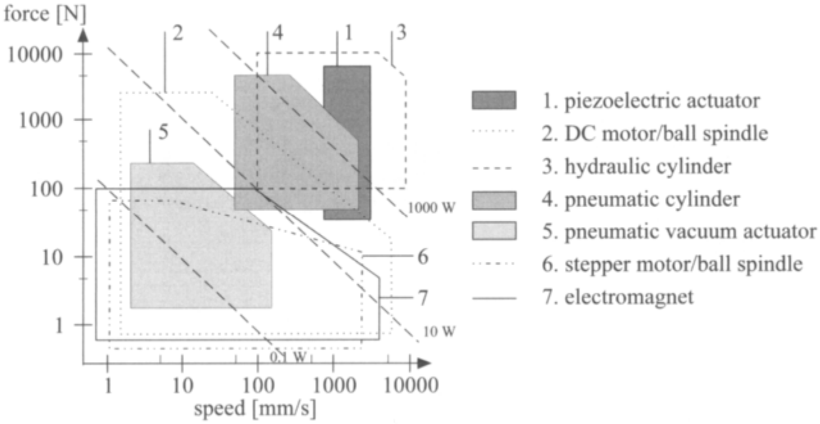


Figure 10.48. Actuating force versus speed range of selected actuators

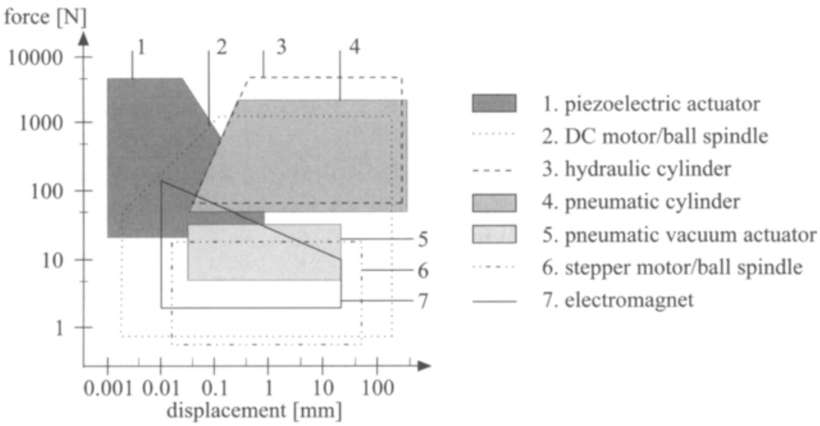


Figure 10.49. Actuating force versus displacement range of selected actuators

Figure 10.50 plots the positioning time for feedback-controlled operation. In this diagram, the smaller times represent small stroke manoeuvres whereas the longer times are representative of a full stroke, *e.g.*, traversing the total control range. Small positioning times can be reached by employing electric actuators. However, piezoelectric actuators (and magnetostrictive actuators, which have not been considered in this comparison), stepper motors and electromagnets also allow for fast positioning.

The power–weight ratio [W/kg] is plotted in Figure 10.51. Hydraulic actuators surpass other actuators in their power–weight ratio provided that the generation of the hydraulic (auxiliary) energy is not considered in the computation of the ratio.

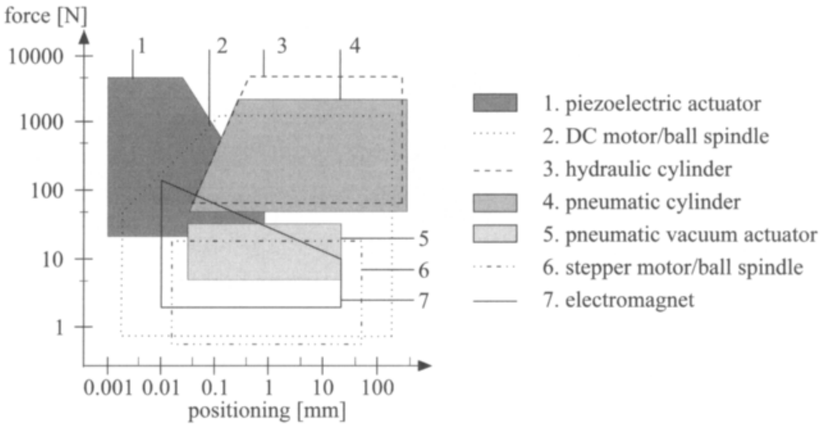


Figure 10.50. Actuating force versus positioning time of selected actuators (feedback-controlled operation)

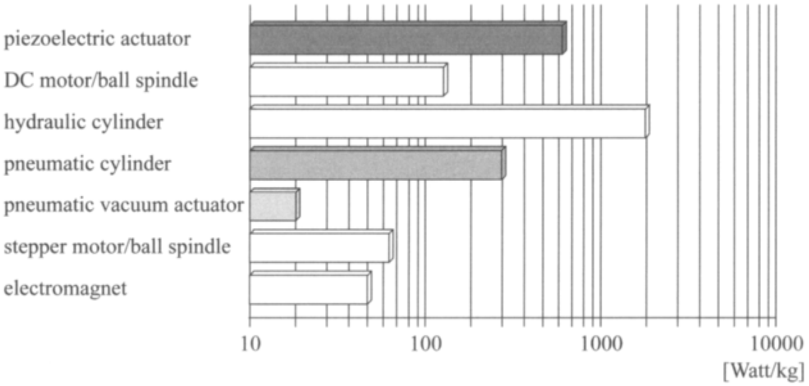


Figure 10.51. Power density of selected actuators (without consideration of auxiliary energy generation)

The demand for high positioning accuracy and good dynamic behavior necessitates feedback-controlled operation of the actuators. Effects such as:

- friction and backlash in mechanical gears and guidance elements;
- hysteresis and saturation effects of certain materials;
- static non-linearities;
- deviation of plant parameters due to operating point dependencies of the parameters or external influences such as aging, wear-and-tear, temperature, auxiliary energy fluctuation

limit the maximum positioning accuracy, Raab, Isermann (1990).

An assessment of the input/output behavior of selected actuators has been summarized in Table 10.38. This table proves the good control characteristics of electric motors, which are to some extent degraded by the subsequent matching gears and feeding drives. Electromagnets exhibit distinct friction and hysteresis effects and have static non-linearities (magnetic force curve). Piezoelectric stack actuators are impaired by hysteresis and backlash. Pneumatic and hydraulic actuators suffer from distinctive friction and non-linearities in the control valves. Their parameters are both direction- and position-dependent and vary with temperature.

To summarize, it can be noted that all actuators are hampered by similar, undesired effects. They play a more dominant role for simple actuator designs. Control approaches that record and compensate these unwanted effects can be used to increase the performance of all actuator systems. Consequently, the mechanism and the microelectronic control circuit must be developed as an "integrated mechatronic actuator system". Fluidic and electromagnetic actuators show a high potential for improvement, since one can integrate both a model-based control system and an automatic fault detection system, Isermann, Raab (1993), Isermann, Keller (1993).

Table 10.38. Assessment of the input/output behavior of selected actuators

properties / actuator type	linearity of force-torque generation	special non-linearities			deviation of plant parameters	
		friction	backlash	electric hysteresis	internal	external
electric motor with gear	+	o	o			o
stepper motor with gear	o	o	o			o
electromagnet	-	-	+	-	o	-
pneumatic cylinder	-	-			-	o
hydraulic cylinder	-	o	+		-	o
PZT-stacked actuator	o	-	-	-		o

symbols: + good, negligible; o mediocre, exisiting; - bad, distinctly noticeable

10.8 ACTUATORS AS SYSTEM COMPONENTS

In the face of the application of actuators as a part of mechatronic systems, the interfacing becomes more important. The term "interfacing" refers to all the properties that allow or obstruct the integration of an actuator in a total system. These are:

- type of auxiliary power;
- input-output behavior;
- ports;
- integration of actuator and process;
- implemented functionality, degree of "intelligence", and smartness;
- measures to increase the reliability.

The first two characteristics have been treated in the preceding chapters.

10.8.1 Ports

In the case of analog signals, the transmitted signals should conform to the standardized current ranges 0 ... 20 mA, 4 ... 20 mA or to the standardized voltage ranges –10 V ... 0 ... +10 V. For connection to a digital bus system, interfaces for serial or parallel data transmission must be used as described in Chapter 11.

10.8.2 Integration of Actuator and Process

There exist a vast variety of ways to connect an actuator to a process. For a small number of units, it is best to resort to standardized flanges, terminals, screw joints and scaled families of actuators (*e.g.*, solenoid valves). For mass production, proprietary constructions will be employed (*e.g.*, fuel-injection pumps, throttle-valve actuators). Actuators are typically available as standardized products.

10.8.3 Implemented Functionality

If control of the actuator, Figure 10.2, is delegated to an integrated microcomputer, then it is possible to implement model-based non-linear adaptive control algorithms, which significantly increase the control performance. Integration of advanced control algorithms can also lead to a reduction of the unit cost. The electromechanical design can be simplified, leading to a simplified and thus cheaper manufacturing process. Then, the induced loss in performance is compensated by the more sophisticated control concepts. Since actuators will increasingly turn into mechatronic components, one can apply the ideas developed in Chapter 1. This will result in "intelligent actuators", so-called *smart* actuators, which can offer the following functions:

- model-based, non-linear adaptive control;
- model-based fault detection (parameter estimation, parity methods, state observer);
- fault diagnosis which informs about the type of fault and the required maintenance;
- energy-optimal and reduced-wear control strategies.

The underlying methods are described in Isermann, Raab (1993), Isermann, Keller (1993).

10.9 FAULT-TOLERANT COMPONENTS

The improvement of reliability can be increased by two different approaches, *perfectness* or *tolerance*, Lauber (1988). Perfectness refers to the idea of avoiding faults and failures by means of an improved mechanical design. This includes the continued technical advancement of actuator components that increase the service life. During operation of the component, the intactness of the component must be maintained by regular maintenance and replacement of wearing parts. Methods that facilitate fault detection at an early stage allow for replacing the regular maintenance schedule with a maintenance-on-demand scheme.

Tolerance describes the notion of trying to contain the consequences of faults and failures thus that the components remain functional. This can be reached by the principle of *fault-tolerance*. Herewith, faults are compensated in such a way that they do not lead to system failures. The most obvious way to reach this goal is *redundancy* in components, units or subsystems. However, the overall systems then become more complex and costly. In the following, various types of fault-tolerant methods are reviewed briefly, see Isermann *et al.* (2000)

10.9.1 Fault-tolerance for Components

Fault-tolerance methods generally use *redundancy*. This means that in addition to the considered module, one or more modules are connected, usually in parallel. These redundant modules are either *identical* or *diverse*. Such redundant schemes can be designed for hardware, software, information processing, and mechanical and electrical components like sensors, actuators, microcomputers, buses, power supplies, *etc.*

Basic redundant structures

There exist mainly two basic approaches for fault-tolerance, static redundancy and dynamic redundancy. The corresponding configurations are first considered for *electronic hardware* and then for other components. Figure 10.52a shows a scheme for *static redundancy*. It uses three or more parallel modules that have the same input signal and are all active. Their outputs are connected to a voter, who compares these signals and decides by majority which signal value is the correct one. If a triple modular-redundant system is applied, and the fault in one of the modules generates a wrong output, this faulty module is masked (*i.e.*, not taken into account) by the two-out-of-three voting. Hence, a single faulty module is tolerated without any effort for specific fault detection, n redundant modules can tolerate $(n - 1)/2$ faults (n odd).

Dynamic redundancy needs less modules at the cost of more information processing. A minimal configuration consists of two modules, Figure 10.52b and c. One module is usually in operation and, if it fails, the stand-by or back-up unit takes over. This requires fault detection to observe if the operation modules become faulty. Simple fault-detection methods only use the output signal for, *e.g.*, consistency checking (range of the signal), comparison with redundant modules or use of information redundancy in computers like parity checking or watchdog timers. After fault detection, it is the task of the reconfiguration to switch to the standby module and to remove the faulty one.

In the arrangement of Figure 10.52b, the standby module is continuously operating, called "*hot standby*". Then, the transfer time is small at the cost of operational aging (wear-out) of the standby module.

Dynamic redundancy, where the standby system is out of function and does not wear, is shown in Figure 10.52c, called "*cold standby*". This arrangement needs two more switches at the input and more transfer time due to a start-up procedure. For both schemes, the performance of the fault detection is essential.

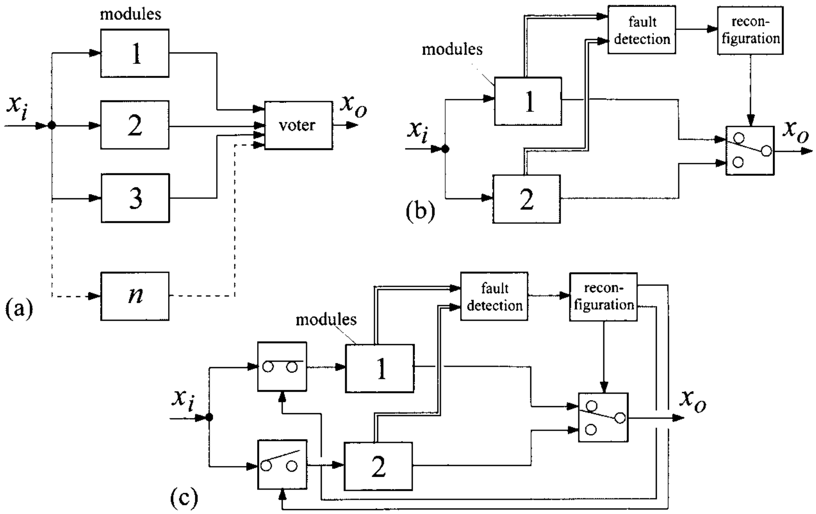


Figure 10.52. Fault-tolerant schemes for electronic hardware: (a) static redundancy: multiple-redundant modules with majority voting and fault masking, m out of n systems (all modules are active);(b) dynamic redundancy: standby module that is continuously active, "hot standby"; (c) dynamic redundancy: standby module that is inactive, "cold standby"

Similar redundant schemes as for electronic hardware exist for *software fault-tolerance, i.e.*, tolerance against mistakes in coding or errors of calculations. The simplest form of static redundancy is repeated running ($n \geq 3$) of the same software and majority voting for the result. However, this only helps for some transient faults. As software faults in general are systematic and not random, a duplication of the same software does not help. Therefore, the redundancy must include diversity of software, like other programming teams, other languages, or other compilers. With $n \geq 3$ diverse programs, a multiple-redundant system can be established followed by majority voting as in Figure 10.52a. However, if only one processor is used, calculation time is increased and using n processors may be too costly.

Dynamic redundancy by using standby software with diverse programs can be realized by using recovering blocks. This means that in addition to the main software module, other diverse software modules exist, Storey (1996), Leveson (1995).

Fault-tolerance can also be designed for purely mechanical and electrical systems. Static redundancy is very often used in all kinds of homogeneous and inhomogeneous materials (*e.g.*, metals and fibers) and in special mechanical constructions like lattice-structures, spoke-wheels, dual tires or in electrical components with multiple wiring, multiple coil windings, multiple brushes for DC motors and multiple contacts for potentiometers. This quite natural built-in fault-tolerance is generally characterized by a parallel configuration, However, the inputs and outputs are not signals but, *e.g.*, forces, electrical currents or energy flows, and a voter does not exist. All elements operate in parallel and if one element fails (*e.g.*, by breakage) the others take over a higher force or current, following the physical laws of compatibility or continuity. Hence, this is a kind of "stressful degradation". Mechanical and electrical systems with dynamic redundancy as depicted in Figure 10.52b, c can also be built. Mostly, only cold standby is meaningful.

Fault tolerance with dynamic redundancy and cold standby is especially attractive for mechatronic systems where more measured signals and embedded computers are already available and therefore fault detection can be improved considerably by applying process model-based approaches. Table 10.39 summarizes the appropriate fault-tolerance methods for the case of electronic hardware.

Table 10.39. Fail behavior of electronic hardware for different redundant structures. FO: fail-operational; F: fail; (FS: fail-safe not considered)

		static redundancy		dynamic redundancy		
structures	number of elements	tolerated faults	fail behavior	tolerated failures	fault behavior	discrepency detection
duplex	2	0	F	0	F	two comparators
				1	FO-F	fault detection
triplex	3	1	FO-F	2	FO-FO-F	fault detection
quadruplex	4	1	FO-F	3	FO-FO-FO-F	fault detection
duo-duplex	4	1	FO-F	-	-	-

Redundant structures for mechatronic systems

Mainly because of costs, space and weight, a suitable compromise between the degree of fault tolerance and the number of redundant components has to be found for mechatronic systems. In contrast to fly-by-wire systems, only one single or two failures can be tolerated for hazardous cases, mainly because a safe state can be reached easier and faster. This means that not all components need very stringent fault-tolerance requirements. The following steps of degradation are distinguished:

- *fail-operational* (FO): one failure is tolerated, *i.e.*, the component stays operational after one failure. This is required if no safe state exists immediately after the component fails;
- *fail-safe* (FS): after one (or several) failure(s), the component directly possesses a safe state (passive fail-safe, without external power) or is brought to a safe state by a special action (active fail-safe, with external power);
- *fail-silent* (FSIL): after one (or several) failure(s), the component is quiet externally, *i.e.*, stays passive by switching off and therefore does not influence other components in a wrong way.

For vehicles, it is proposed to subdivide FO into "long time" and "short time". Considering these degradation steps for various components, one has to check first if a safe state exists. For automobiles, (usually) a safe state is stand still (or low speed) at a non-hazardous place. For components of automobiles, a fail-safe status is (usually) a mechanical back-up (*i.e.*, a mechanical or hydraulic linkage) for direct manipulation by the driver. Passive fail-safe is then reached, *e.g.*, after failure of electronics if the vehicle comes to a stop independently of the electronics, *e.g.*, by a closing spring in the throttle or by actions of the driver via mechanical backup. However, if no mechanical back-up exists after failure of electronics, only an action by other electronics (switch to a still operating module) can bring the vehicle (in motion) to a safe state, *i.e.*, to reach a stop through active fail-safe. This requires the availability of electric power.

Generally, a *graceful degradation* is envisaged, where less critical functions are dropped to maintain the more critical functions available, using priorities, IEC 61508 (1997). Table 10.39 shows degradation steps to fail-operational for different redundant structures of electronic hardware. As the fail-safe status depends on the controlled system and the kind of components, it is not considered here.

For flight-control computers, usually a triplex structure with dynamic redundancy (hot standby) is used, which leads to FO-FO-FS, such that two failures are tolerated and a third one allows the pilot to operate manually. If the fault tolerance has to cover only one fault to stay fail-operational (FO-F), a triplex system with static redundancy or a duplex system with dynamic redundancy is appropriate. If fail-safe can be reached after one failure (FS), a duplex system with two comparators is sufficient. However, if one fault has to be tolerated to continue fail-operational and after a next fault it is possible to switch to a fail-safe (FO-FS), either a triplex system with static redundancy or a duo-duplex system may be used. The duo-duplex system has the advantages of simpler failure detection and modularity.

10.9.2 Fault-tolerance for Control Systems

For automatically controlled systems, the appearance of faults and failures in the actuators, the process and the sensors will usually affect the operating behavior. With feedforward control, generally all small or large faults influence the output variables and therefore more or less the operation.

If the system operates with feedback control, small additive or multiplicative faults in the actuator or process are in general covered by the controller, because of the usual robustness properties. This property is therefore *passive controller fault-tolerance*. However, additive and gain sensor faults will immediately lead to deviations from the reference values. For large changes in actuators, process and sensors, the dynamic control behavior becomes either too sluggish or too less damped or even unstable. Then either a very robust control system or an *active fault-tolerant* control system is required to save the operation. In the last case, it consists of fault-detection methods and reconfiguration mechanisms, which modify the controller. Depending on the kind of faults, the reconfiguration may change the structure and/or parameters or the controller. This can also include the change to other manipulated variables or actuators or sensors, if available.

Examples are fault-tolerant flight control with reconfiguration to other control surfaces after failure of actuators or ailerons, elevators and rudders, see, *e.g.*, Rauch (1995), Chandler (1997), Patton (1997), Chen *et al.* (1999). For failures in the satellite altitude control system, see Blanke *et al.* (1997). Failures in heat exchangers are treated in Ballé *et al.* (1998) and fault-tolerant control for lateral vehicle control in Suryanaryanan, Tomizuka (2000).

10.9.3 Fault Detection for Sensors, Actuators and Mechatronic Servo-systems

Fault-detection methods based on measured signals can be classified as:

- *limit value checking* (thresholds) and *plausibility checks* (ranges) of single signals;
- *signal model-based methods* for single periodic or stochastic signals;
- *process model-based methods* for two or more related signals.

Figure 10.53 shows a scheme for these methods.

For a description of the various method, refer to the literature, *e.g.*, the special section in IFAC Journal Control Engineering Practice (1996) or the books Chen, Patton (1999), Gertler (1999), Isermann (1994a).

In order to obtain specific symptoms it is necessary to have more than one input and one output signal for parity equations or output observers. For parameter estimation, one input and one output may be sufficient. Because of the various properties, it is recommended to combine different methods in order to have a large fault detection coverage, Isermann (1994b), Pfeufer (1997).

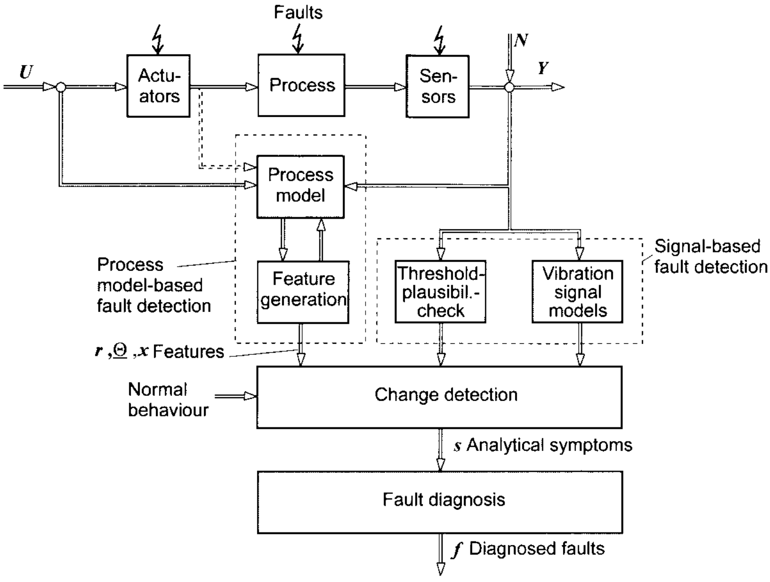


Figure 10.53. General scheme of process model-based and signal-based fault detection

10.9.4 Fault-tolerant Components for Mechatronic Systems

High-integrity systems require a comprehensive overall fault-tolerance by fault-tolerant components and corresponding control. This means the design of fault-tolerant sensors, actuators, process parts, computers, communication (bus systems), and control algorithms. Examples of components with multiple redundancy are know for aircraft, space and nuclear power systems. However, lower cost components with built-in fault tolerance have to be developed. In the following, some examples are given for sensors and actuators.

10.9.5 Fault-tolerant Sensors

A fault-tolerant sensor configuration should be at least fail-operational (FO) for one sensor fault. This can be obtained by applying hardware redundancy with the same type of sensors or by analytical redundancy with different sensors and process models.

Hardware sensor redundancy
Sensor systems with static redundancy are realized, for example, with a triplex system and a voter, Figure 10.54a. A configuration with dynamic redundancy needs at least two sensors and fault detection for each sensor, Figure 10.54b. Usually, only hot standby is feasible. Another less powerful possibility is plausibility checks for two sensors, also by using signal

models (*e.g.*, variance) to select the more plausible one, Figure 10.54c.

The fault detection can be performed by *self-tests*, *e.g.*, by applying a known measurement value to the sensor. Another way uses *self-validating sensors*, Henry, Clarke (1993), Clarke (1995), where the sensor, transducer and a microprocessor form an integrated, decentralized unit with self-diagnostic capability. The self-diagnosis takes place within the sensor or transducer and uses several internal measurements. The output consists of the sensor's best estimate of the measurement and a validity status, like good, suspect, impaired, bad and critical.

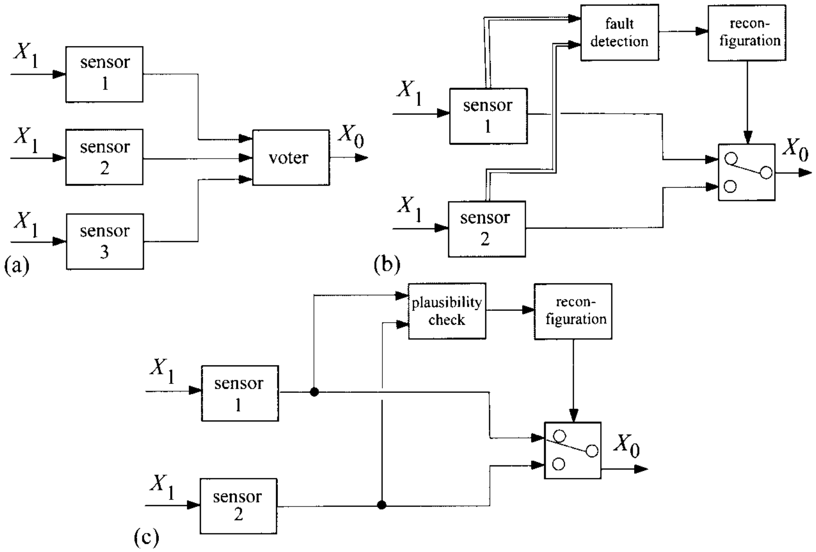


Figure 10.54. Fault-tolerant sensors with hardware redundancy: (a) triplex system with static redundancy and hot standby; (b) duplex system with dynamic redundancy, hot standby; (c) duplex system with dynamic redundancy, hot standby and plausibility checks

Analytical sensor redundancy

As a simple example, a process with one input and one main output y_1 and an auxiliary output y_2 is considered, see Figure 10.55a. Assuming the process input signal u is not available but two output signals y_1 and y_2, which both depend on u, one of the signals, *e.g.*, $\hat{y}_1$ can be reconstructed and used as a redundant signal if process models G_{M1} and G_{M2} are known and considerable disturbances do not appear (ideal cases).

For a process with only one output sensor y_1 and one input sensor u, the output $\hat{y}_1$ can be reconstructed if the process model G_{M1} is known, Figure 10.55b. In both cases, the relationship between the signals of the process are used and expressed in the form of analytical models.

To obtain one usable fault-tolerant measurement value y_{1FT}, at least three different values for y, *e.g.*, the measured one and two reconstructed ones, must be available. This can be obtained by combining the schemes of Figure 10.55a and b as shown in Figure 10.56a. A sensor fault y_1 is then detected and masked by a majority voter and either $\hat{y}_1$ or $\hat{y}_{1u}$ is used as a replacement depending on a further decision. (Also, single sensor faults in y_2 or u are tolerated with this scheme.)

One example for this combined analytical redundancy is the yaw rate sensor for the ESP (electronic stability program) of vehicles, where additionally the steering wheel angle as input is used to reconstruct the yaw rate through a vehicle model as in Figure 10.55b, and the lateral acceleration and the wheel speed difference of the right and left wheel (no slip) are used to reconstruct the yaw rate according to Figure 10.55a.

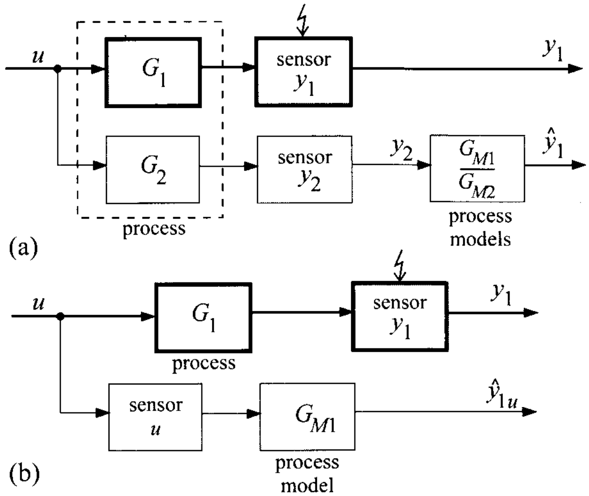
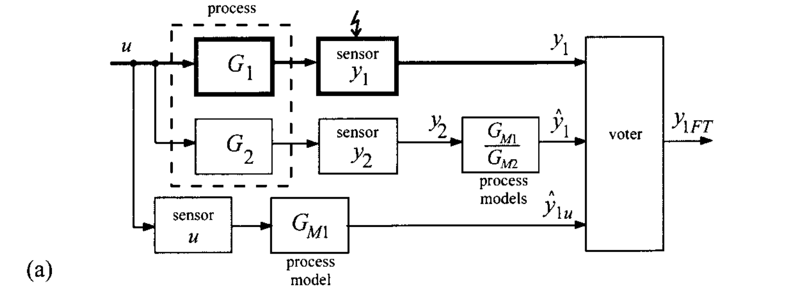


Figure 10.55. Sensor fault-tolerance for one output signal y_1 (main sensor) through analytical redundancy by process models (basic schemes): (a) two measured outputs, no measured input; (b) one measured input and one measured output

A more general sensor fault-tolerant system can be designed if two output sensors and one input sensor yield measurements of the same quality. Then, by a scheme as shown in Figure 10.56b, three residuals can be generated and by a decision logic, fault-tolerant outputs can be obtained in the case of single faults of any of the three sensors. The residuals are generated based on parity equations. In this case, state observers can also be used for residual generation, compare, *e.g.*, the dedicated observers by Clark (1989). (Note that all schemes assume ideal cases. For the realizibility, constraints and additional filters have to be considered.)

If possible, a faulty sensor should be fail-silent, *i.e.*, should be switched off. However, this needs additional switches that lower the reliability. For both hardware and analytical sensor redundancy without fault detection for individual sensors, at least three measurements must be available to make one sensor fail-operational. However, if the sensor (system) has in-built fault detection (integrated self-test or self-validating), two measurements are enough and a scheme like Figure 10.54b can be applied. (This means that by methods of fault detection, one element can be saved).



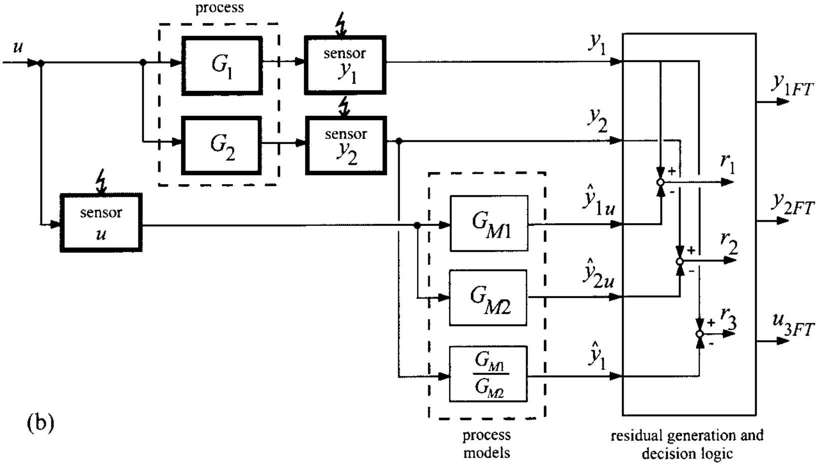


Figure 10.56. Fault-tolerant sensors with combined analytical redundancy for two measured outputs and one measured input through (analytical) process models: (a) y_1 is main measurement, y_2, u are auxiliary measurements (combination of Figure 10.55a and b); (b) y_1, y_2 and u are measurements of same quality (parity equation approach)

10.9.6 Fault-tolerant Actuators

Actuators generally consist of different parts: input transformer, actuation converter, actuation transformer and actuation element (*e.g.*, a set of DC amplifier, DC motor, gear and valve, as shown in Figure 10.57a. The actuation converter converts one form of energy (*e.g.*, electrical or pneumatic) into another form (*e.g.*, mechanical or hydraulic). Available measurements are frequently the input signal U_i, the manipulated variable U_0 and an intermediate signal U_3.

Fault-tolerant actuators can be designed by using *multiple complete actuators* in parallel, either with static redundancy or dynamic redundancy with cold or hot standby (Figure 10.52). One example of static redundancy are hydraulic actuators for fly-by-wire aircraft where at least two independent actuators operate with two independent hydraulic energy circuits.

Another possibility is to limit the redundancy to parts of the actuator that have the lowest reliability. Figure 10.57b shows a scheme where the actuation converter (motor) is split into separate parallel parts. Examples with static redundancy are two servo-valves for hydraulic actuators, Oehler *et al.* (1997) or three windings of an electrical motor (including power electronics), Krautstrunk, Mutschler (1999). Within electromotor-driven throttles for SI engines, only the slider is doubled to make the potentiometer position sensor static-redundant.

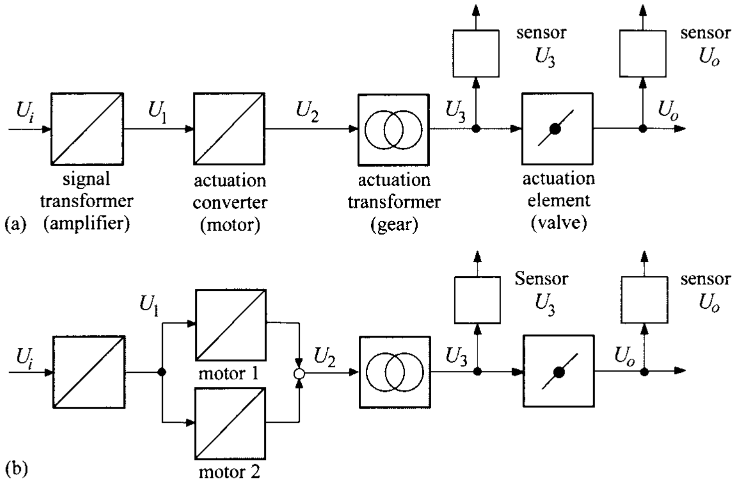


Figure 10.57. Fault-tolerant actuator: (a) common actuator; (b) actuator with duplex drive

One example for dynamic redundancy with cold standby is the cabin pressure flap actuator in aircraft, where two independent DC motors exist and act on one planetary gear, Moseler *et al.* (1999).

As cost and weight generally are higher than for sensors, actuators with fail-operational duplex configuration are to be preferred. Then, either static-redundant structures, where both parts operate continuously, Figure 10.52a, or dynamic redundant structures with hot standby, Figure 10.52b, or cold standby, Figure 10.52c, can be chosen. For dynamic redundancy fault-detection methods of the actuator parts are required, Isermann, Raab (1993). One goal should always be that the faulty part of the actuator fails silent, *i.e.*, has no influence on the redundant parts.

10.10 PROBLEMS

10.10 .1 Describe the different components of an AC motor-driven flow valve by drawings according to Figure 10.2. Include a position controller for the valve shaft.

10.10.2 In which cases are linear or non-linear characteristics of flow valves selected?

10.10.3 State the advantages and disadvantages of electromotors, pneumatic membrane drives and hydraulic cylinder drives for:

a) flow control of steam flow for a 500 MW steam turbine;
b) position control for a machine tool feed drive;
c) ailerons of an aircraft.

10.10.4 Compare the properties of DC electrical brush and brushless motor.

10.10.5 Which actuator drives and subsequent gears need or do not need power for position holding under load?

10.10.6 How can the non-linear current-position behavior of electromagnets be improved for position control of the armature by constructive means or by algorithmic ways?

10.10.7 Which actuator drives should be preferred for the following requirements for

a) displacement 1 m and very large speed and force;
b) displacement 10 mm, force 5 N and cheap mass production;
c) displacement 0.01 mm and force 100 N;
d) displacement 0.1 m, force 1000 N and explosive environment;
e) displacement 0.2 m, force 1000 N and high power-weight ratio.

10.10.8 How large are approximately the smallest time constants of pneumatic, hydraulic, electromotoric, electro magnets and piezoelectric actuator drives?

10.10.9 Determine the time constant of a pneumatic diaphragm valve with diameter D = 0.1 m, air volume V = $1 \cdot 10^{-3}$ m³, inflow $\dot{V}_{max} = 0.5\, l/s$, $p_i = 1$ bar. It is assumed that friction an be neglected.

10.10.10 What types of linear controllers can be used for precise and fast position control of an electromagnet, pneumatic diaphragm and hydraulic cylinder drive?

10.10.11 Describe the possibilities for the construction of fault-tolerant drives with two electrical motors.

10.10.12 How can a fault-tolerant temperature measurement system be built with two sensors (thermocouple and resistance thermometer) and three sensors of the same type?