Chapter 1
Introduction

Control means a specific action to reach the desired behavior of a system. In the control of industrial processes generally technological processes, are considered, but control is highly required to keep any physical, chemical, biological, communication, economic, or social process functioning in a desired manner.

Control methods should be used whenever some quantity must be kept at a desired value. For example, control is used to maintain the temperature of our flat at a comfortable specific value both in winter and summer. Controlling an aircraft, the pilot (or the robot pilot) has to execute extremely diverse control tasks to keep the speed, the direction, and the altitude of the aircraft at desired values. Control systems are all around us, in the household (e.g., setting the program of a washing machine, ironing by on-off temperature control, air conditioning, etc.), in transportation, space research, communication, industrial manufacturing, economics, medicine, etc. A lot of control systems do operate in living organisms as well.

Control systems are everywhere in our surroundings. A control system is realized e.g., when taking a shower, where the temperature of the shower is to be kept at a comfortable value (Fig. 1.1). If the temperature sensed by our body differs from its desired value, we intervene by opening the cold tap or the warm tap. After being mixed, the water goes through the shower pipe. The effect of the change takes place after a delay. The effect of the delay has to be considered when deciding on a possible newer execution. The control process taking place is symbolized by the block-diagram shown in Fig. 1.2.

Figure 1.3 shows schematically a control system for room temperature control.

Figure 1.4 illustrates some processes which require control to ensure appropriate performance. The speed or angular position of the motor, as well as the level of the tank, is to be kept at a constant value. The temperature of the liquid flowing through the heat exchanger has to be maintained. In the chemical reactor, the quality and quantity of the materials being created during the chemical reaction have to be maintained. In the distillation column the individual components of the crude oil are

L. Keviczky et al., *Control Engineering*, Advanced Textbooks in Control and Signal Processing, https://doi.org/10.1007/978-981-10-8297-9_1

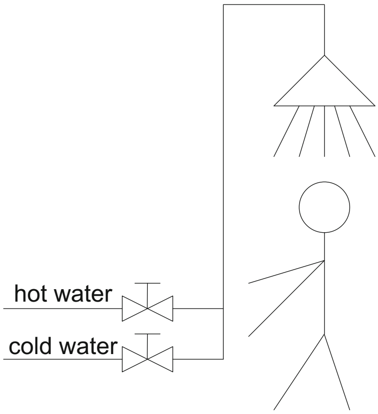


Fig. 1.1 Shower-bath as a control task

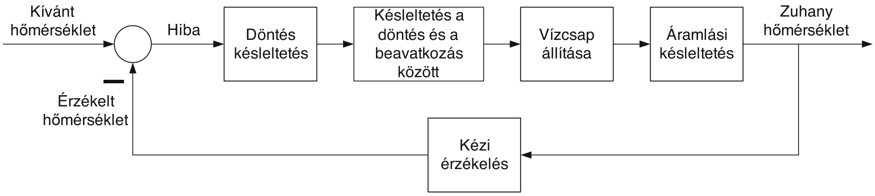


Fig. 1.2 Control block-scheme of the shower-bath

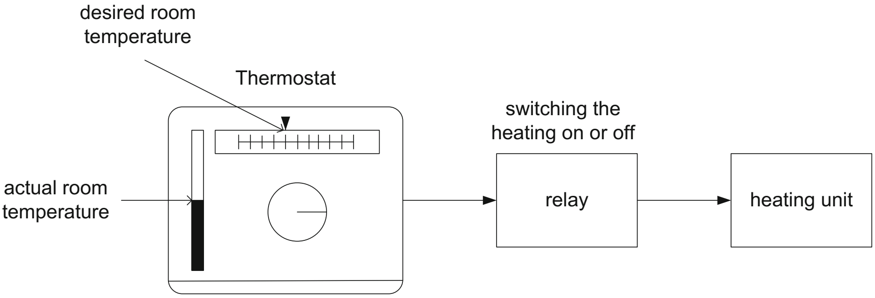


Fig. 1.3 Room temperature control

to be separated. For this purpose, the temperatures of the plates in the column have to be appropriately controlled relative to each other. Furthermore, in everyday practice in the household and in a variety of production processes different control tasks have to be solved.

In what follows, the control processes of technological systems will be discussed. The control of industrial processes plays a significant role in ensuring better product quality, minimizing energy consumption, increasing safety and decreasing environmental pollution.

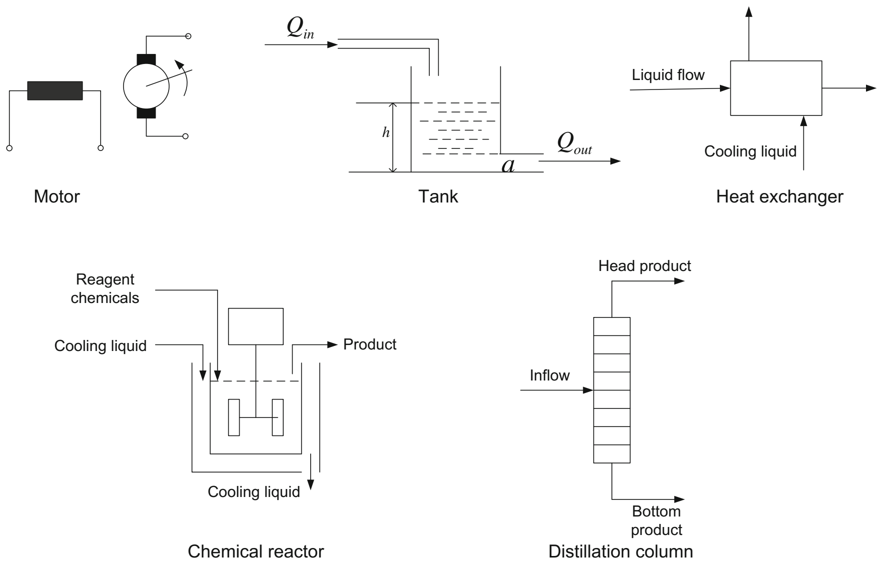


Fig. 1.4 Some typical control tasks

In the manufacturing production processes of material goods, mass and energy conversion takes place. Appropriate control is to be applied to ensure the suitable starting, maintenance and stopping of these processes. For example, in a thermal power station the chemical energy of the coal is converted to heat energy by burning. The heat is then used to produce steam. The steam drives the turbine, creating mechanical rotation energy. The turbine rotates the rotor of a synchronous generator in the magnetic field of the stator. This creates electric energy. All these processes must be operated in a prescribed way. The processes have to be started, and their performance has to be ensured according to the given technological prescriptions. For example, in electrical energy production, it has to be ensured that the voltage and frequency be kept at prescribed constant values within a given accuracy in spite of load changes during the day. Stopping the processes has to be executed safely.

To maintain the processes in a desired manner means keeping different physical quantities at constant values or altering them according to given laws. Such physical quantities could be, for instance, the temperature or pressure of a medium, the composition of a material, the speed of a machine, the angular position of an axe, the level in a tank, etc.

A process is a system which is connected to its environment in many ways. For example, a thermal power station converts the chemical energy of the fuel to electrical energy. The system consists of several pieces of interconnected equipment (furnace, turbine, synchronous generator, auxiliary equipment). The system converts the input quantity (fuel) to the output quantity (electrical energy), while it has multi-faceted relations with its environment (it produces waste material, transfers

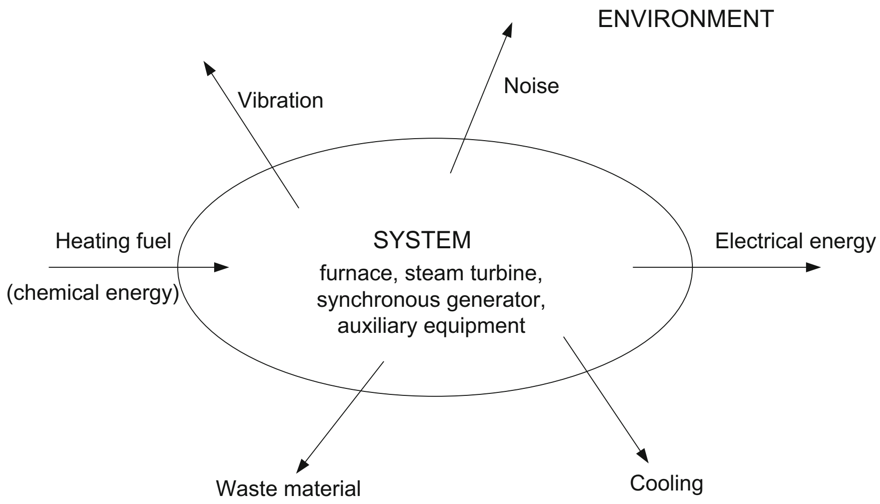


Fig. 1.5 The system and its environment

heat into the environment, produces mechanical vibration and noise, etc.). Figure 1.5 illustrates the relation of the system and its environment. If the operation of the turbine is investigated, then the relation of the system and its environment is considered in a different way (Fig. 1.6). In this case the system is the turbine, which converts the thermal energy of the steam into electrical energy.

The quantities going from the environment into the system are the inputs, while the quantities going from the system into the environment are the outputs. With control—by appropriately manipulating the input quantities—the output quantities are to be maintained according to the given requirements.

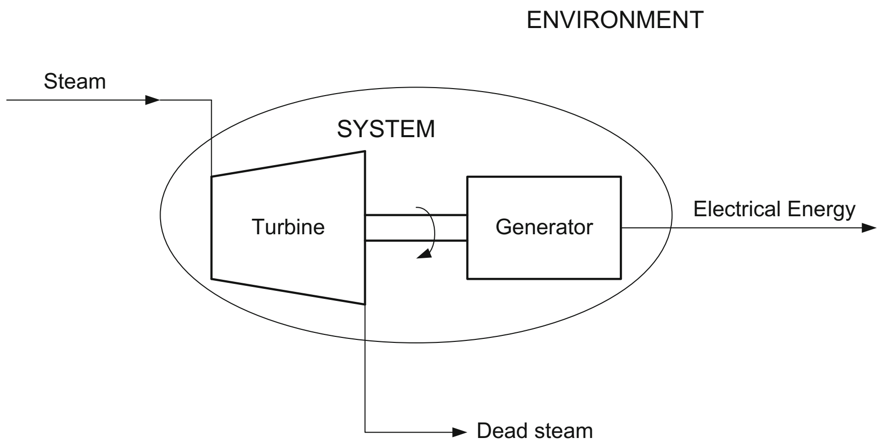


Fig. 1.6 The system and its environment (as a detailed part of the system in Fig. 1.5)

1.1 Basic Concepts

Control means the specific actions to influence a process in order to start it, to appropriately maintain it, and to stop it.

Control is based on information obtained from the process and its environment through measurements. Measuring instruments are needed to measure the different physical quantities involved in the control. Based on the knowledge of the control's aim and on the information obtained from the process and its environment, a decision is made about the appropriate manipulation of the process input. It is characteristic for control that high energy processes are influenced by low energy causes.

The methodology of control is that specifically designed external equipment is connected to the process and then, based on data obtained by measurements or in other ways, it directly modifies the input variables and in that way influences indirectly the output variables. The control system is the joint system made up of the interconnected plant to be controlled and the control equipment.

Control can be performed manually or automatically. In manual control the operator makes a decision and manipulates the input quantity of the process based on the observed output quantity. In automatic control automatic devices execute the functions of decision making and executing the manipulation. Taking a shower is a case of manual control (Fig. 1.1), Fig. 1.7 also illustrates manual control. The operator observes the level of the liquid in the tank and sets the required level by the valve position of the tap influencing the amount of the outlet liquid. Figure 1.8 shows an automatic level control in a tank. The level of the liquid is sensed by a floating sensor. If the level differs from its required value, the valve influencing the input flow will be opened more or less.

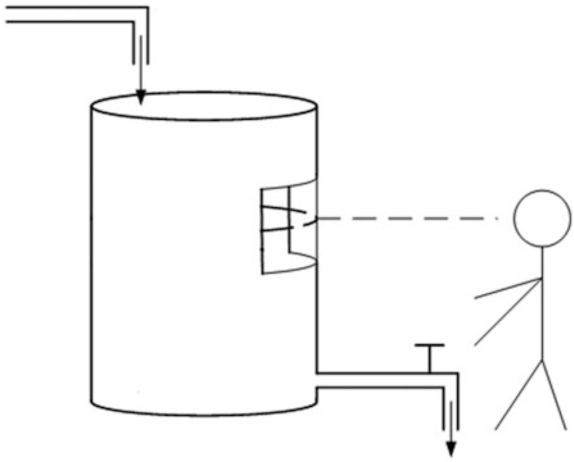


Fig. 1.7 Level control by hand

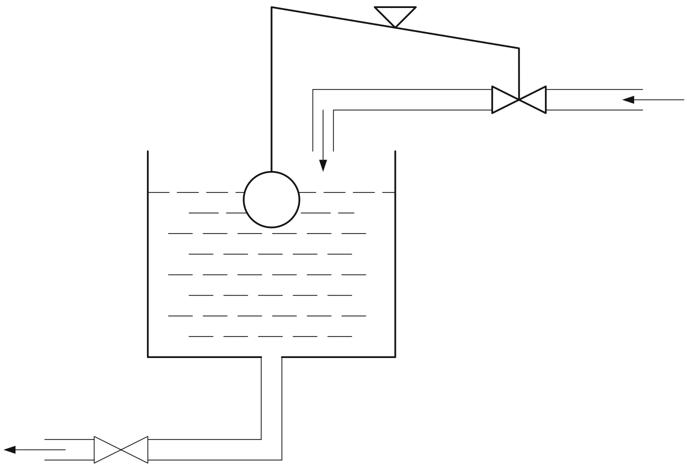


Fig. 1.8 Automatic level control in a water tank

Control engineering deals with the properties and behavior of control systems, with the methods for their analysis and design, and with the question of their realization.

1.1.1 The Basic Elements of a Control Process

A control process consists of the following operations (Fig. 1.9):

Sensing: gaining information about the process to be controlled and its environment
Decision making: processing the information and, based on the aim of the control taking decisions about the necessary manipulations
Disposition: giving a command for manipulation
Signal processing: determine the characteristics of intervention, acting
Intervention, Acting: the modification of the process input according to the disposition.

The individual operations are executed by the appropriate functional units.

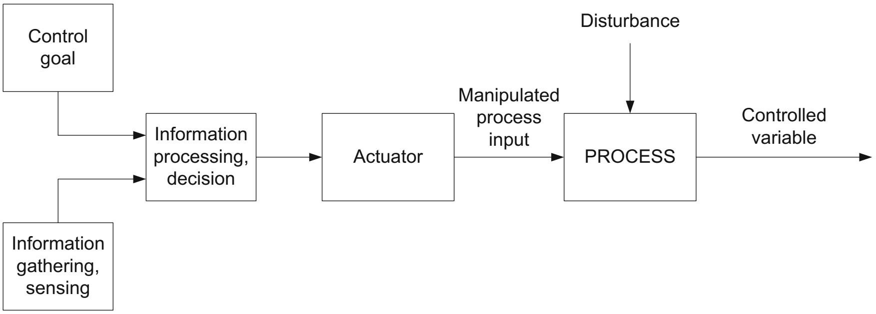


Fig. 1.9 Functional diagram of a control system

1.1.2 Signals and Their Classification

To control a process it is required to measure its changes. Changes of the process occur as consequences of external and internal effects. The features of the process which manifest its motion, and also the external and internal effects, are represented by signals. The signal is a physical quantity, or a change in a physical quantity, which carries information. The signal is capable of acquiring, transferring, as well as storing information. Signals can be observed by measurement equipment. Signals have a physical form (e.g., current, voltage, temperature, etc.)—this is the carrier of the signal. Signals also have informational content—which shows the effect represented by the signal (e.g., change of the current versus time).

Signals can be classified in different ways.

According to its temporal evolution

a signal is *continuous* if it is continuously maintained without interruption over a given range of time,

a signal is *discrete-time or sampled* if it provides information only at determined points in time in a given duration of time.

According to its set of value

a signal is *contiguous* if its set of value is contiguous,

a signal is *fractional* if its set of value is non contiguous and can take only definite values.

According to the form of representation of the information

a signal is *analog* if the value of the signal carrier directly represents the infor mation involved,

a signal is *digital* if the information is represented by digits which are the coded digital values of the signal carrier.

According to the definiteness of the signal value

a signal is *deterministic* if its value can definitely be given by a function of time,

a signal is *stochastic* if its evolution is probabilistic, which can be described using statistical methods.

The characteristic signals of a process are its inputs, outputs, and internal signals. Those input signals which are supposed to be used as inputs modifying the output of the process are called manipulated variables or control variables. The other input variables are disturbances.

1.1.3 Representation of System Engineering Relationships

The various parts of a control system are in interaction with each other. The relations of the individual parts can be represented by different diagrams. As was mentioned earlier, a piece of equipment which performs some control task is called functional unit (e.g., sensor, actuator, etc.). The symbols for the functional units also appear in the diagrams characterizing the connections of the elements of the control system.

A *structural diagram* gives an overview of the pieces of equipment forming the system and shows their connections. First of all it highlights those parts of the system which are substantial from the control viewpoint. Generally, a structural diagram uses the standard notation of the specific field under consideration.

Considering the performance of a control system, what is of primary interest is not the operation of the individual functional units, but rather the spreading effect of the information induced by their operation. An *operational block diagram* shows the connection and interaction of the individual control units disregarding their physical characteristics. In a block diagram the units are represented by rectangles. A line supplied with an arrow directed to a rectangle symbolizes the input signal, while a directed line going out of a rectangle represents the output signal. The direction of the arrow is also the direction of the flow of information. In the rectangles the functions of the structural units are indicated (e.g., sensor, actuator element, controller, etc.).

When realizing a control system, the requirements for the process and the aim of the control have to be formulated first. Then, to solve the control problem, the individual structural control units are chosen. These units are connected to the process and to each other according to the control structure. It has to be analyzed whether the control system meets the quality specifications. To do this it is required to examine the signal transfer properties of the individual elements and also the signal transfer in the interconnected system. In a *block diagram* the individual elements of the operational diagram are described by their signal transfer properties, i.e., by the mathematical formula giving the relationships between the outputs and inputs. These relationships can be mathematical equations, tables, characteristics, operation commands, etc. The signal transfer properties of the individual elements can be given by a mathematical description of the physical operation of the element, where the values of the parameters involved in the equations are also given. To indicate some frequently used operations, accepted symbols are written inside the rectangles (e.g., the symbol of integration). The symbols of summation and subtraction are shown in Fig. 1.10. A *chain of effect* is a set of connected elements along a given direction.

A block diagram can be considered as the mathematical model of the control system. In this model, mainly the signal transfer properties of the system are kept in view, other properties are ignored.

The static and dynamic behavior of the control system can be investigated based on the block diagram. The block diagram also provides the basis for the design of the control system.

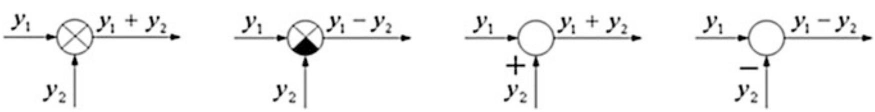


Fig. 1.10 Symbols of summation and subtraction

Of course, when the control system is actually implemented, in addition to its signal transfer properties, other aspects should also be taken into account (e.g., energy constraints, standardized solutions, etc.).

1.1.4 Open- and Closed-loop Control, Disturbance Elimination

If the information is not gained directly from the measurement of the controlled signal, an open-loop control is realized. If the information is derived by directly measuring the controlled signal, a closed-loop control or feedback control is obtained. Figure 1.11 gives the operational block diagram of a closed-loop control system.

An example of an open-loop control system is the control of a washing machine according to a time schedule of executing consecutive operations (rinsing, washing, spin drying). The output signal (the cleanness of the cloths) is not measured. An open-loop control is realized also if the heating of a room is set depending on the external temperature.

In the case of a closed-loop (feedback) control the controlled signal itself is measured. The control error, i.e., the deviation between the actual and the desired value of the controlled signal, influences the input of the process. The functional units are the sensor (measuring equipment), the unit providing the reference signal, the subtraction unit, the amplifier and signal forming unit, and the executing and actuator unit. The characteristic signals of the processes are measured by sensors. The measuring instruments provide signals which are proportional to the different physical quantities measured. The requirements set for the sensors are the following:

- reliable operation in the range of the measurements
- linearity in the range of the measurements
- accuracy

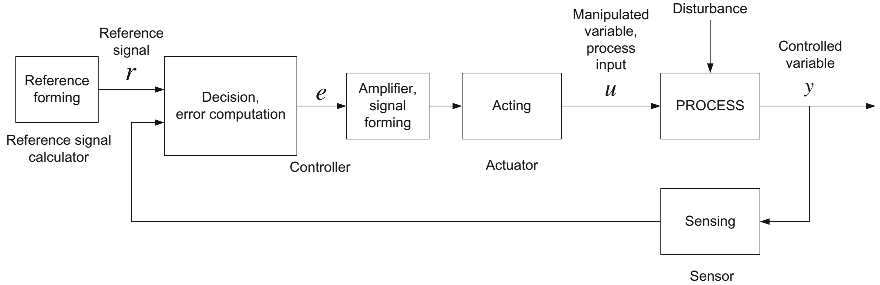


Fig. 1.11 Operational block diagram of the closed-loop control system

- small dead-time compared to the time constants of the process
- low measurement noise.

A sensor measures the physical quantity which is to be controlled and transforms it to another physical quantity which is proportional to the actual value of the controlled signal, and can be compared to the reference signal provided by the reference unit. The error signal operates the controller. The output signal of the controller is amplified, formed and operates the acting element (actuator) which provides the input signal (manipulated variable) for the process. The error signal gives the deviation of the actual output signal from its desired value. If it is different from zero, the system input is to be modified to eliminate the error.

The different functional units are selected according to practical considerations. The control system is built from the individual control elements (sensors which measure the given physical variables in the required range, controllers, actuators, miscellaneous elements) available on the market.

The basis of a closed-loop control system is *negative feedback*. The command for modifying the input of the process is performed based on comparing the reference signal and the actual value of the output signal to be controlled. (There are different schemes for realizing control systems, but all of them are based on negative feedback.)

Because of the dynamics of the plant and the individual elements of the control system, signals need time to go through the control loop. A well designed controller takes the dynamics of the closed-loop system into consideration and ensures the fulfillment of the quality specifications imposed on the control system.

Comparison of open-loop and closed-loop control

If the relationship between the control signal (manipulated variable) and the controlled signal (process variable) is known and reliable information is available on all the elements and all the disturbances in the control circuit, then open-loop control can ensure good control performance. But if our knowledge about the plant and about the disturbances is inaccurate, then the performance of the open-loop control will not be satisfactory. Open-loop control provides a cheap control solution, as it does not apply expensive sensors to measure the controlled quantity, but instead it uses apriori information or information gained about external physical quantities for decision making. In open-loop control there are no stability problems.

Closed-loop control is more expensive than open-loop control. The controlled variable is measured by sensor equipment, and manipulation of the input signal of the plant is executed based on the deviation between the reference signal and the measured output signal. Closed-loop control is able to track the reference signal and to reject the effect of the disturbances. As the actual value of the controlled signal is influenced by the disturbances, closed-loop control rejects the effect of the disturbances which are not known in advance, and also compensates the effect of the parameter uncertainties of the process model. If any kind of effect has caused the difference between the output signal and its required value, the closed-loop control is activated to eliminate the deviation. But because of the negative feedback stability problems may occur, oscillations may appear in the system. The stability of the control system can be ensured by the appropriate design of the controller.

If the disturbance is measurable, then closed-loop control is often supplemented by *feedforward* using the measured value of the disturbance. A block diagram of the feedforward principle is shown in Fig. 1.12. A signal depending on the measured disturbance variable is fed forward to some appropriate summation point of the control loop. This means an open-loop path which relieves the closed-loop control in disturbance rejection. This forward path tries to compensate the effect of the disturbance. This manipulation works in open-loop, the disturbance variable influences the controlled variable, but the manipulation does not affect the disturbance variable.

A classical example of feedforward compensation is the compound excitation of a direct current (DC) generator (Fig. 1.13). The armature voltage is the controlled variable, the excitation is the control (manipulated) variable. The load current (disturbance variable) decreases the armature voltage of the generator. With compound excitation, part of the excitation is created by the load current itself, thus the disturbance variable directly produces the effect of eliminating itself. In this way the armature voltage of the generator is greatly stabilized. For more accurate voltage control, an additional closed-loop configuration can be applied.

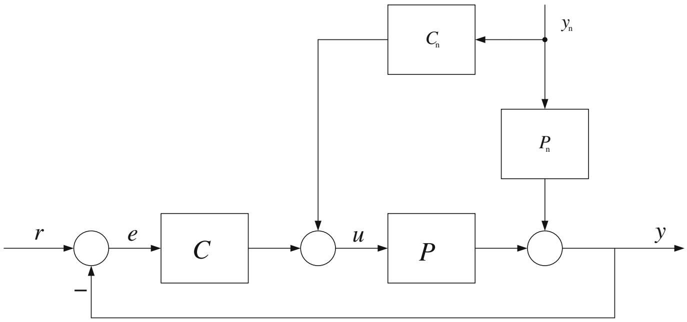


Fig. 1.12 Feedforward control (disturbance compensation)

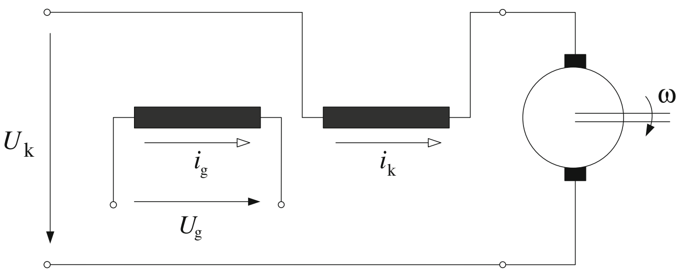


Fig. 1.13 DC generator with compound excitation

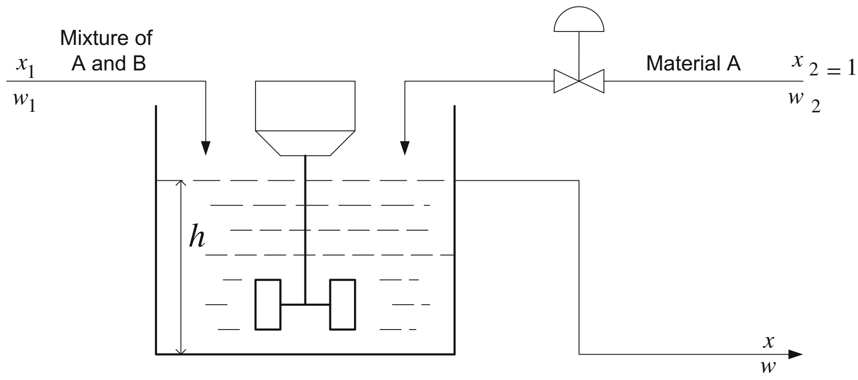


Fig. 1.14 Stirring tank

Let us consider the stirring tank in Fig. 1.14, where w_1 is the inflow quantity of the mixture of materials A and B flowing into the tank. In the mixture the partial rate of material A is x_1. w_2 is the inflow quantity of the pure material A, $x_2 = 1$. w denotes the amount of the outflow material of partial rate x. It is supposed that w_1 is constant, x_2 is constant, and the mixing process in the tank works ideally. The control aim is to keep the composition x of the outflow material (the controlled variable) at a prescribed value in spite of the variations in x_1 (disturbance variable). Manipulations can be executed by modifying the inflow quantity w_2 (control or manipulated variable) by setting the position of the valve. The control is realized by a closed-loop control, if x is measured and w_2 is set depending on this measurement (Fig. 1.15). An open-loop control is built if the composition x_1 of the inflow mixture material is measured, and the inflow amount w_2 is modified accordingly (Fig. 1.16). Figure 1.17 shows a feedforward solution, where both the composition x of the outflow material and the composition x_1 of the inflow mixture are measured, and the inflow quantity w_2 is set according to both measured values (In the figures, the standard symbols for the sensors, controllers, and valves are employed, see Appendix A.3).

The next example shows the speed control of a motor with open-loop and closed-loop control. In a CD player the disc has to be rotated at steady speed. A DC motor can be used as actuator. The angular velocity is proportional to the terminal voltage of the motor. Figure 1.18 shows the solution of the task in open-loop control. The terminal voltage of the motor is provided by a direct current power supply through an amplifier. The velocity is proportional to the terminal voltage. Figure 1.19 schematically presents the solution using closed-loop control. Figure 1.19a gives the structural diagram, while 1.19b shows the operational diagram. The speed of the motor is measured with a tachometer generator, whose output voltage is proportional to the velocity. The measured voltage is compared to the reference signal voltage set by the power supply, which is proportional to the prescribed value of the speed. The error signal operates the actuator DC motor.

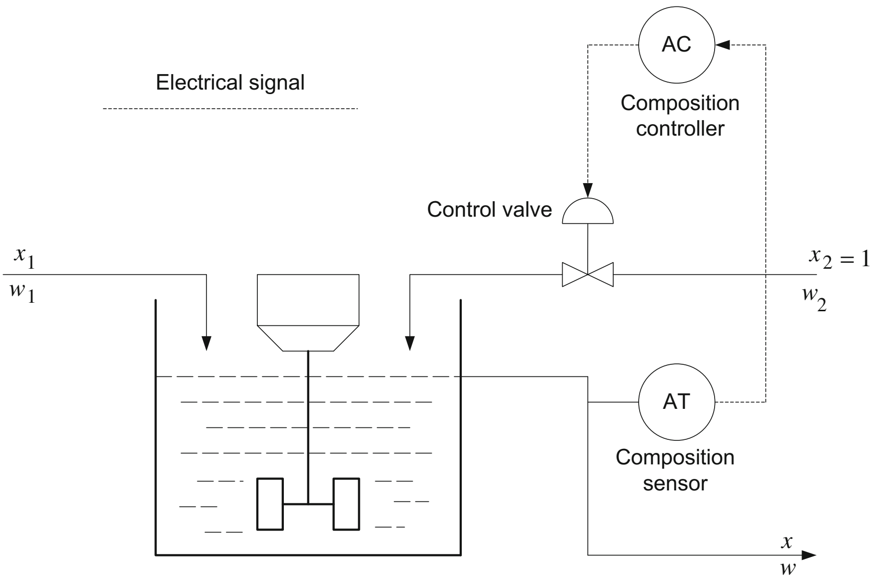


Fig. 1.15 Closed-loop composition control of the liquid in a tank

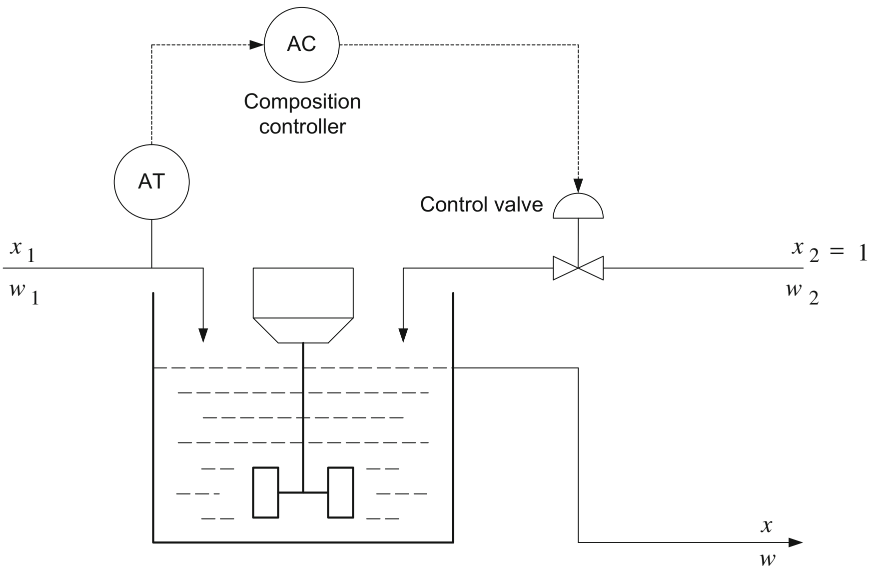


Fig. 1.16 Open-loop composition control of a liquid in a tank

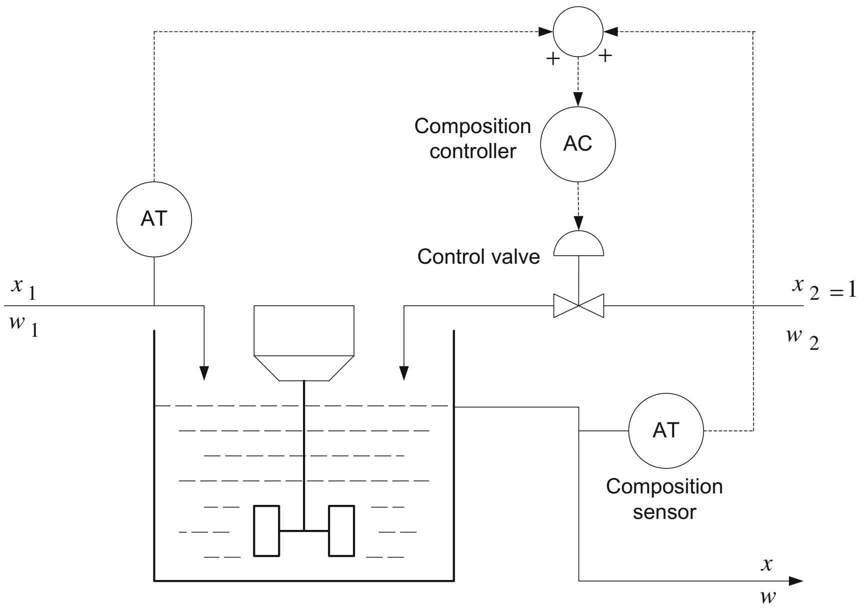


Fig. 1.17 Feedforward composition control of a liquid in a tank

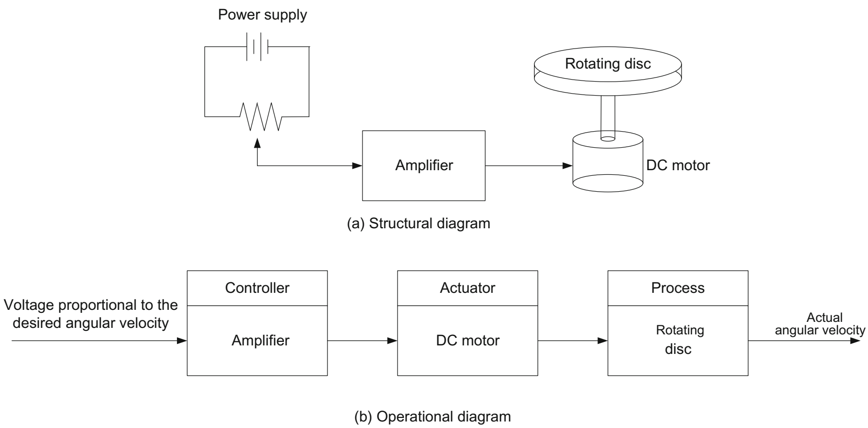


Fig. 1.18 Open-loop angular velocity control of a CD player

With closed-loop control more accurate and more reliable operation can be reached. Closed-loop control ensures not only reference signal tracking, but eliminates speed changes resulting from possible changes in the load, as well.

In practice, besides closed-loop control, open-loop control systems are also given an important role. When starting and stopping a complex system, a series of complex open-loop control operations has to be executed. Generally, intelligent

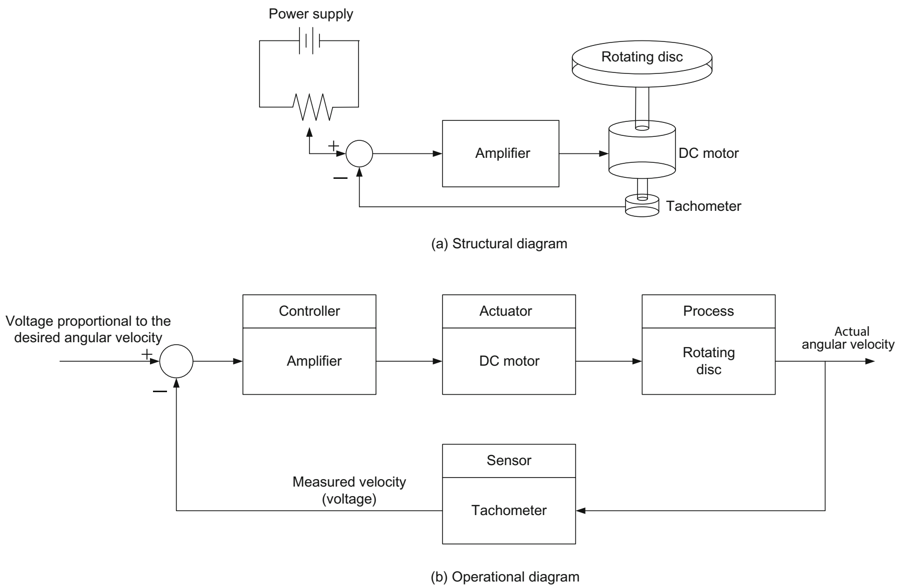


Fig. 1.19 Closed-loop angular velocity control of a CD player

Programmable Logic Controller (PLC) equipment is used to realize the open-loop control. To keep various physical quantities at their required constant values closed-loop control systems are applied.

1.1.5 General Specifications for Closed-Loop Control Systems

The main goal of a closed-loop control system is to track the reference signal and to reject the effect of the disturbances. Regarding the quality of the performance of the control system static and dynamic requirements are prescribed.

First of all a closed-loop control has to be stable, i.e., oscillations of steady or increasing amplitude in the loop variables are not allowed. After the change of the input signals a new balance state has to be reached. The problem of instability comes from the negative feedback realizing the closed-loop control. As after the appearance of the control error the manipulation of the process input can be executed only in a delayed fashion, it may occur that undesired transients do appear in the system (e.g., in Fig. 1.1 when taking a shower the water can be too hot or too cold, the desired temperature is not settled.) Stable behavior can be ensured by appropriate controller design. (The stability of a control system will be discussed in detail in Chap. 5).

Static specifications give the allowed maximum value of the steady error of the reference signal tracking, and the allowed remaining steady deviation in the output

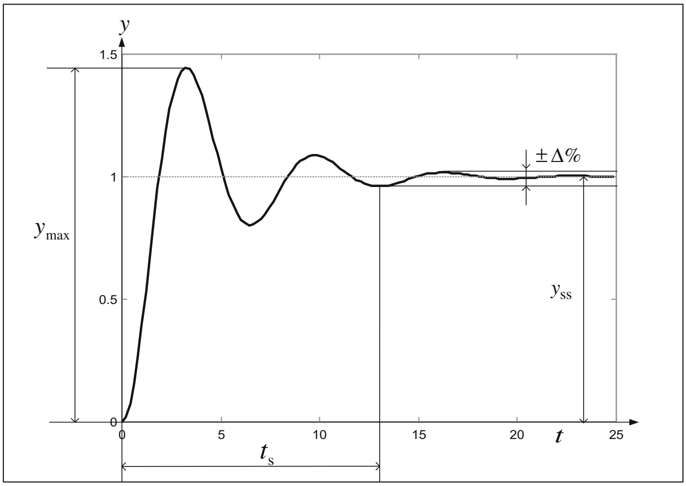


Fig. 1.20 Dynamic quality specifications

signal occurring as the effect of the disturbances, after deceasing of the transients, in steady state. It depends on the technology and on the process to be controlled whether deviations can be allowed at all, and if so, what their maximum possible value can be.

Dynamic specifications give prescriptions for the course of the transients. Let us consider the step response of the closed-loop control system (Fig. 1.20) with the indicated maximum value y_{max} and steady-state value $y_{ss} = y_{steady-state}$. The overshoot σ in percentages is expressed by

$$\sigma = \frac{y_{max} - y_{ss}}{y_{ss}} \cdot 100\%$$

There are processes where aperiodic performance is required (e.g., machine tools, landing of an airplane, etc.), while in other processes often an overshoot of 5–10% is tolerable.

The settling time t_s specifies the time it takes for the step response of the closed-loop control system to settle down within an accuracy of $\pm\,\Delta\%$ (generally $\pm(1–2)\%$) of its steady state value. Usually the number of allowed oscillations within the settling time is also prescribed.

The control signal in the control system is the output signal of the actuator. The control signal (or manipulated variable) can only take a restricted value corresponding to its physical realization (e.g., a valve setting the inflow liquid quantity in a tank can provide a maximum amount of liquid passing through in its totally open state, and is not able to provide more, in spite of possibly receiving such a command.). If a higher value were to be forced, the actuator would be saturated at only releasing its maximum possible amount, thus temporarily “opening” the control loop. The phenomenon of possible saturation of the manipulated variable (control signal) should be considered already in the control design phase, and it has to be ensured that the manipulated variable be within its specified range, or if nevertheless it exceeds it, effects substantially distorting the normal operation of the control loop should be avoided.

A control system is designed for the process to be controlled ensuring the quality specifications. The model of the process describing its signal transfer properties is obtained by mathematical description reflecting its physical operation. The values of the parameters in the equations are determined generally by measurements. Thus in their values uncertainties may occur. The closed-loop control has to operate appropriately (in a robust way) even if the actual parameters of the process and the parameters considered in its model do differ to some extent.

The requirements set for the closed-loop control system have to be realistic. For example, extremely fast settling can not be required from a slow heating process, as this would result in extremely high control signals. Instead, it is necessary to relax the strictness of the prescriptions in order to get a realizable solution.

Chapter 4 deals in more detail with the quality specifications set for a closed-loop control system.

1.1.6 Simple Control Examples

Next, some examples of closed-loop control will be presented.

Temperature control

Figure 1.21 shows a schematic structural diagram of a device producing warm water with a prescribed temperature. The water is circulating in tubes located in the stokehold of a furnace. The coal used for firing is delivered from the coal container to the heating equipment by a conveyor driven by an electrical motor. The velocity of the conveyor and thus the amount of the transported coal is controlled by the speed of the motor. On the basis of the difference between the prescribed temperature of the warm water and its measured actual value the controller sets the terminal voltage determining the speed of the electrical motor through a preamplifier and a power amplifier. Figure 1.22 shows a block diagram of the temperature control.

Speed control

Figure 1.23 shows the structural diagram of the speed control of a direct current (DC) motor with constant external excitation. The speed of the motor can be changed by the terminal voltage (manipulated variable). The machine driven by the motor produces a changing load for the motor (disturbance), and produces variation in the speed. The terminal voltage of the motor can be changed by an electronic unit

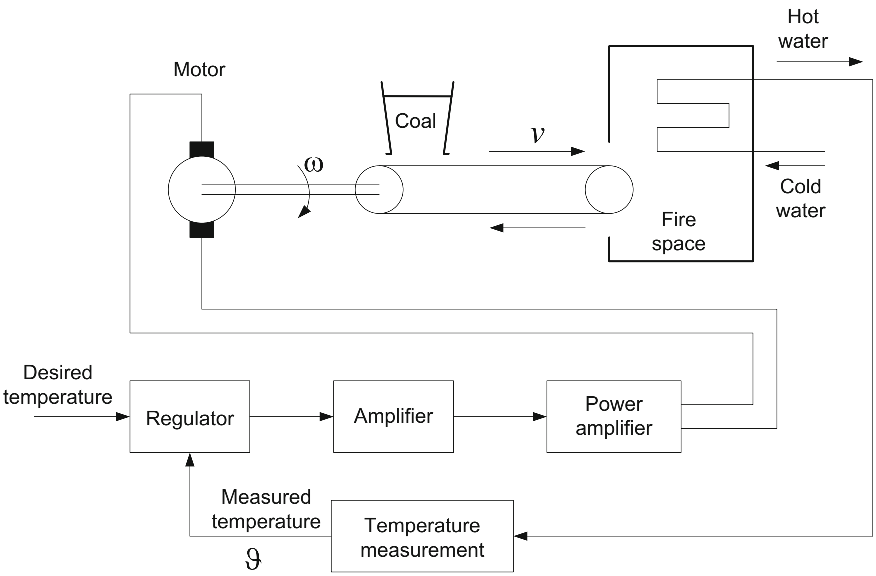


Fig. 1.21 Schematic structural diagram of temperature control

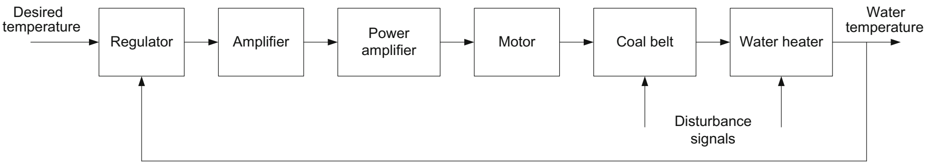


Fig. 1.22 Block diagram of temperature control

with thyristors. The speed of the motor is measured by a tachometer generator, which gives a voltage proportional to the speed (angular velocity). The error voltage is obtained by comparing this voltage with the reference signal voltage provided by the power supply. Its magnitude is amplified by the power amplifiers E1 and E2 and its shape is modified by a filter. Thus the manipulated variable is produced. The function of the manipulated variable is to change the firing angle of the thyristors. As a consequence, the terminal voltage, as the control signal, will be increased or decreased in order to reach the speed prescribed by the reference voltage of the power supply. A block diagram of the speed control is given in Fig. 1.24.

Level control, composition control, moisture control

Frequent tasks in industrial chemical processes are the following: level control in a tank, pressure control, temperature control, composition control of mixed materials, moisture control, etc. Figure 1.25 shows two solutions for liquid level control. In the upper figure the manipulation is executed by the control of the inflow. In the

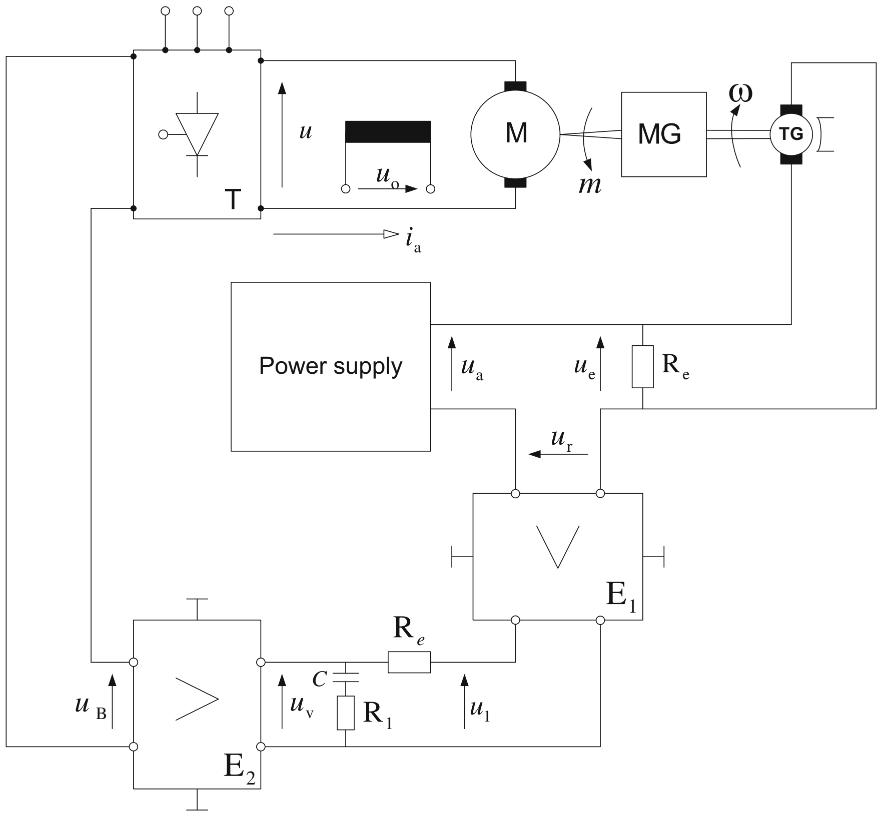


Fig. 1.23 Speed control of a DC motor

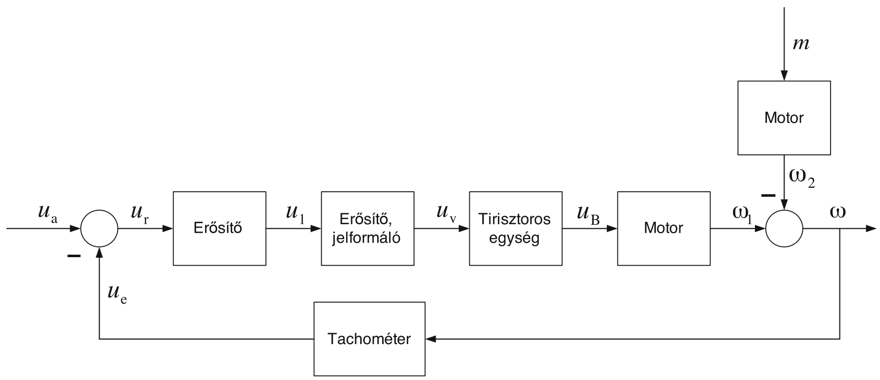


Fig. 1.24 Block diagram of speed control

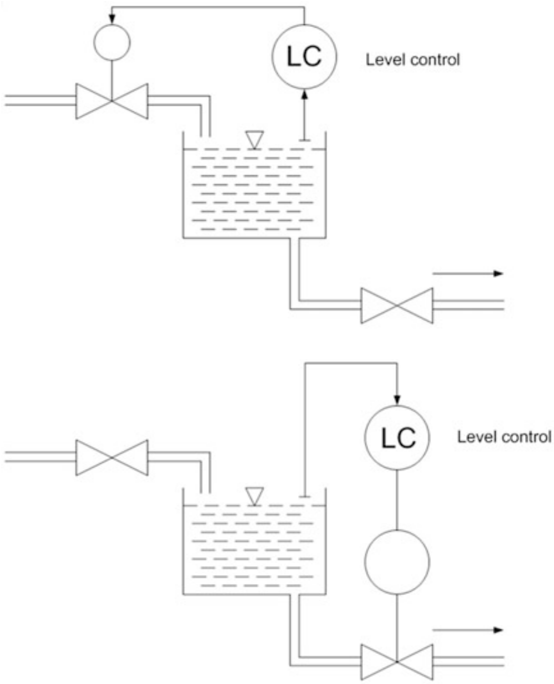


Fig. 1.25 Level control in a tank

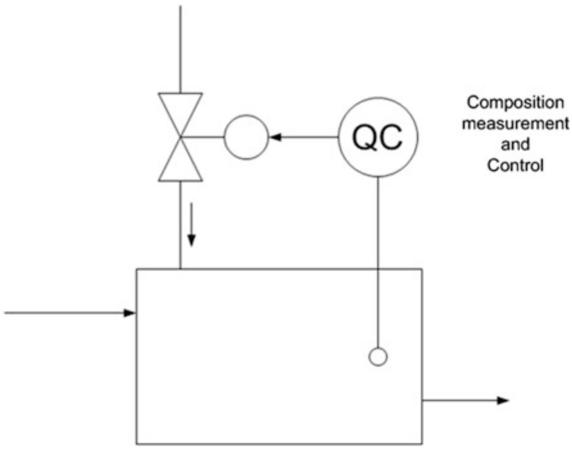


Fig. 1.26 pH control

lower figure the manipulation is executed by the control of the outflow. Figure 1.26 illustrates pH control. Figure 1.27 gives a schematic solution for the moisture control of a granular material in a drying process. The moisture content of the material is measured, and in case of its deviation from the desired value, the speed of the conveyor belt is modified or the inflow of the drying steam (or hot air) is changed.

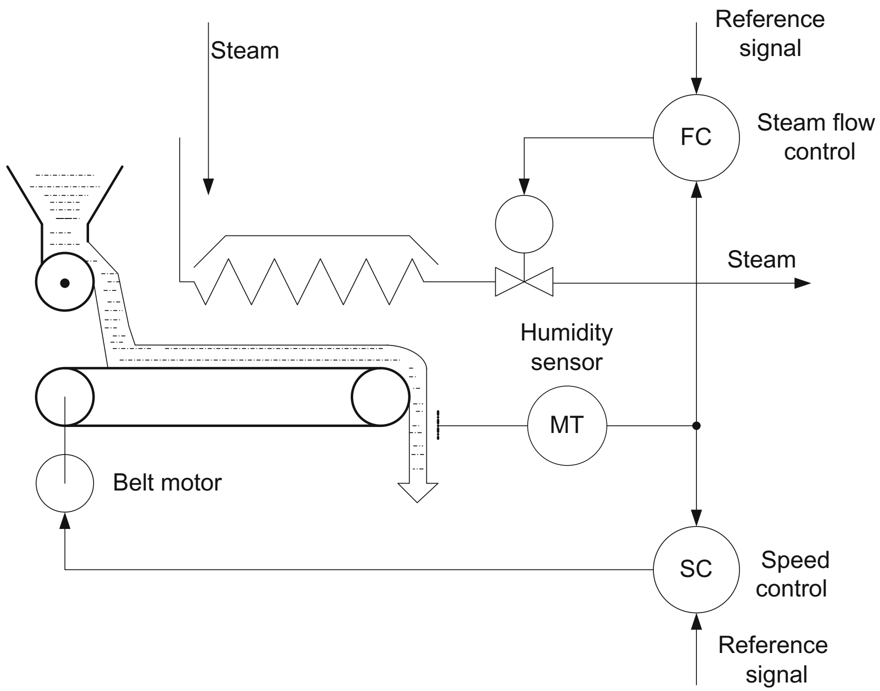


Fig. 1.27 Moisture control

1.2 On the History of Control

Control engineering even today is a developing discipline. New facilities and new techniques raise new theoretical questions, and open up the way to novel applications. Applying negative feedback is not a new principle, however: the ancient Greeks already used it. Looking back at the history of control engineering, some tendencies can be observed.

The application of negative feedback relates to the solution of engineering tasks. The development of control engineering is tightly connected to practical problems that waited for a solution in a stage of humanity's history. Some periods which had a significant influence at the development of control technique were

- the ancient Greek and Arab culture ($\sim$300 BC to $\sim$1200 AD),
- the industrial revolution (18th century, but the beginnings already around 1600)
- the beginnings of telecommunication (1910–1945)
- the appearance of computers, the beginning of space research (1957–)

Considering these eras we may establish that humanity was looking first for their place in space and time, and then tried to shape the environment to make life more comfortable; industrial production contributed to this. Then, using also communication, humans found their place and position in society, and then tried to get connected to the universe.

Already the ancient Greeks used several automata. One of the first closed-loop control systems was the water clock of Ktesibios in Alexandria (270 BC). The equipment used a float to sense the level of a tank and to keep it at a constant value. If the water level in the tank decreased, a valve opened and refilled the tank. A constant level ensured a constant value of the outflow of the water. The outflowing water filled a second tank. The level of this tank changed proportionally to the time. The Byzantine Philon (250 BC) also used a float controller to control the oil level of an oil lamp. Heron of Alexandria (first century AD) applied similar devices for level control, wine dosage, opening doors of churches, etc.

Arab engineers between 800 and 1200 AD used several controllers with floating balls. They initiated the on-off controllers, which operate by switching on and off the manipulating variable.

With the invention of mechanical clockwork, water clocks with floating balls were forgotten.

In the era of the industrial revolution many types of automatic equipment were invented. In these systems, the tasks of automatic level, temperature, pressure and speed control were carried out. Already from the beginning of the 17th century there were several control applications (speed control of windmills, temperature control of furnaces (Cornelis Drebbel), pressure control (Papin), etc.). The discovery of the steam engine (Savery and Newcomen, ~1700) indicates the beginning of the industrial revolution. The centrifugal controller of James Watt (Fig. 1.28) is considered the first industrial control system, which was applied to the speed control of a steam engine. The position of the centrifugal sensor depends on the speed of the steam engine. This sensor sets the position of the piston valve through the actuating lever, thus influencing the amount of steam inflowing to the

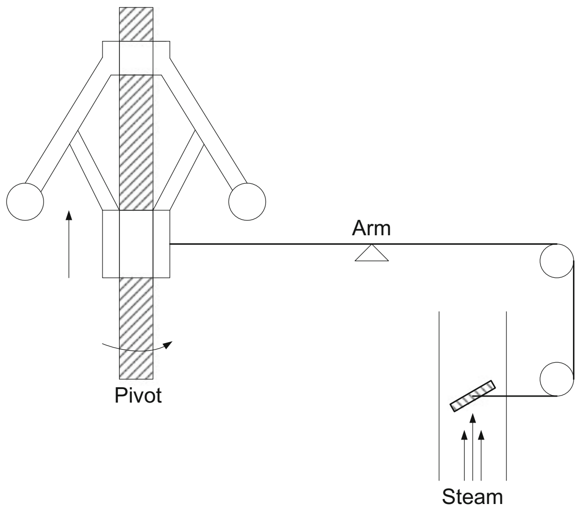


Fig. 1.28 Centrifugal controller

steam engine, changing its speed. (It is interesting to mention that almost another hundred years had to pass until Maxwell gave the exact mathematical description of the system with differential equations).

After the industrial revolution an essential step forward in the development of control engineering was the use of mathematical methods for the description of control circuits. This made possible a more rigorous and exact investigation of control systems.

A new era of control engineering started with the invention of the telephone, with the application of feedback operational amplifiers to compensate for the damping occurring in the transmission of the information.

During the Second World War a lot of high precision control systems were worked out, e.g., automatic flight control systems, radar antenna positioning systems, control equipment of submarines, etc. Then later on these techniques also gained applications in industrial production.

The general application of computers opened a new era in the development of control systems. The computer is no longer only an external device, to facilitate the control design, but becomes part of control systems in real time applications. The process and the process control computer are connected via peripherials, and the process control software calculates the control signal at every sampling time instant and forwards it to the process input. Thus the computer became a basic part of the control loop.

Industrial robots executing precision tasks appeared. The robot is a computer controlled automaton. Several times, human attributes have been imitated in robots, e.g. in robot manipulators the motion of the human hand is imitated. Mobile robots are aimed to be equipped with some intelligence, such as observing and avoiding obstacles moving in space.

Space research means a newer challenge for control systems. Tracking space-craft, placing artificial space objects in a given orbit requires extremely accurate, learning control systems which are able to adapt to changing circumstances. In these systems safe operation is extremely important.

Nowadays when realizing different control systems the control principles, the computer and communication systems and their interaction have to be considered together. The new technical possibilities facilitate new ways of control applications. The appearance of the new miniaturized sensors and manipulating elements opens new perspectives in control techniques. In industrial production processes, distributed control systems have appeared; a large number of control systems distributed in space are coordinated to ensure high quality production. These systems communicate, change information, forward commands and execute them in a coordinated way. Hardware and software elements (PLC-s, profibus, TCP/IP, industrial network standards, etc.) ensuring the operation at this level appeared.

Control theory deals with the construction and analysis and synthesis of closed-loop control systems. The classical period of control theory ($\sim$till 1960) gave the basic concepts of the operation, analysis and synthesis of closed-loop control systems based on negative feedback.

In the modern era of control theory (~1960–1980), the state space description of control systems and controller design methods based on this model have gained attention.

Nowadays design methods of robust reliable control systems which are less sensitive to parameter changes are in the forefront of interest. Control of non-linear systems, application of intelligent learning systems which are able to recognize environmental changes and adapt to them, application of distributed control systems using network connections and communication, open new perspectives in control theory and control engineering.

1.3 Systems and Models

Building a model is a significant part of analyzing a control system. The model describes the signal transfer properties of a system in mathematical form. With a model, the static and dynamic behavior of a system can be analyzed without performing experiments on the real system. Based on the model, calculations can be executed and the behavior of the system can be simulated numerically. A model of the system can also be used for controller design.

The choice of the elements of a control system is based on practical considerations. The operation of a control system can be followed in the structural diagram, which shows the connections and interactions of the individual units building the control system. The mathematical model of the elements of the control loop describes their signal transfer properties. In a control loop the signal transfer properties of all the elements are given by mathematical relationships. A block diagram can be considered as a mathematical model of the control loop. With a block diagram, the static and dynamic properties of the control system can be analyzed, and it can be determined whether the system satisfies the quality specifications.

The signal transfer properties of the individual elements can be given by mathematical relationships describing their physical operation. A deep understanding of the physical operation is required to derive its mathematical description. The parameters in the mathematical equations can be determined by calculations or by measurements.

The static and dynamic behavior of a system can also be obtained by analyzing the input signals and the output signals resulting from the effect of the input signals. For the execution of an experiment providing information for system analysis, it is important to choose the input signals appropriately. This procedure requires some form of a system model, and determines the parameters in such a way that the outputs of the system and that of the model be closest to each other in terms of a cost function. This procedure is called *identification*.

As the values of the parameters are generally determined by measurements, their values are not quite accurate, but usually the range of the parameter uncertainties can be given.

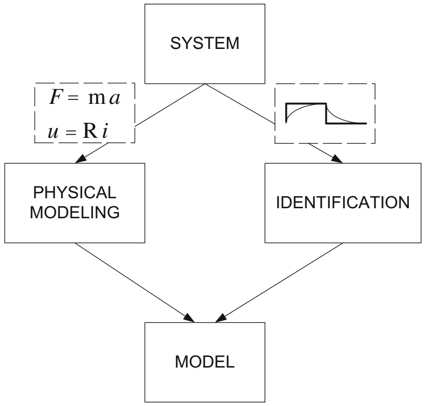


Fig. 1.29 Creation of a model of a system

To obtain a model of a system generally physical modeling and identification are used together (Fig. 1.29).

The model is reliable if its output for a given input approximates well the real output of the system. The domain of validity of the model can be obtained (e.g., in which range of the input signal it is valid).

1.3.1 Types of Models

A model is *static* if its output depends only on the actual value of its input signal. For example, a resistance where the input signal is the voltage and the output signal is the current is a static system. A model is *dynamic* if its output depends on previous signal values as well. An electrical circuit consisting of serially connected resistor and capacitor is a dynamical system, since the voltage drop on the capacitance depends on the charge, and thus on the previous values of the current.

A model can be *linear* or *non-linear*. The static characteristic plots the steady values of an output signal versus the steady values of an input signal. If the static characteristics are straight lines, the system is linear, otherwise it is non-linear.

A model can be *deterministic* or *stochastic*. The signals of a deterministic model can be described by analytical relationships. In a stochastic model, the signals can be given by probabilistic variables and contain uncertainties.

Spatially, a model can have either *lumped* or *distributed parameters*. Lumped parameter systems can be described by ordinary differential equations, while distributed parameter systems can be described by partial differential equations.

A model can be a *continuous-time* (CT) or a *discrete-time* (DT) model. A continuous-time model gives the relationship between its continuous input and output generally in the form of a differential equation. If the input and the output are sampled, the system is a discrete-time or sampled data system, where the relationship between the input and the output signals is described by a difference equation.

Considering the number of the input and the output signals, the model can be *Single Input Single Output* (*SISO*), *Multi Input Multi Output* (*MIMO*), *Single Input Multi Output* (*SIMO*) or *Multi Input Single Output* (*MISO*). Besides the input and the output signals *state variables* of the system can also be defined. The state variables are the internal variables of the system, whose current values have evolved through the previous changes of the signal in the system. Their values can not be changed abruptly when the input signals change abruptly. The current values of the input signals and that of the state variables determine the further motion of the system.

Our investigations will be restricted to the control of dynamic, linear, *SISO*, lumped parameter systems. The literature basically applies the following four methods to describe such systems:

- linear lumped parameter differential equations of order n
- state space equations
- the transfer function and frequency function
- time functions.

1.3.2 The Properties of a System

Some important system properties—which characterize the relationship between the input and the output—are linearity, causality and time invariance.

Linearity: A system is linear if the superposition and homogeneity principles are applicable to it. If for an input signal u_1 the output signal of the system is $y_1 = f(u_1)$, and for the input signal u_2 the output signal is $y_2 = f(u_2)$, then the superposition principle means that $y_1 + y_2 = f(u_1 + u_2)$; according to the homogeneity principle, a k-fold change in the input signal yields a k-fold change in the output signal: $k\,y = f(k\,u)$. It can also be stated that for the input signal $\alpha u_1 + \beta u_2$ the output signal is $\alpha y_1 + \beta y_2$.

Causality: at a given time instant the output depends on the past and the current input values, but it does not depend on future input values.

Time invariance: A system is time invariant if its response to the input signal does not depend on the time instant of applying the input signal: to an input signal shifted by a dead-time of τ, it gives the same response shifted by the dead-time τ (Fig. 1.30). In a time invariant system, for the delayed output the following relationship holds: $y_\tau(t) = y(t - \tau)$.

Linear time invariant systems generally are referred by the acronym *LTI*.

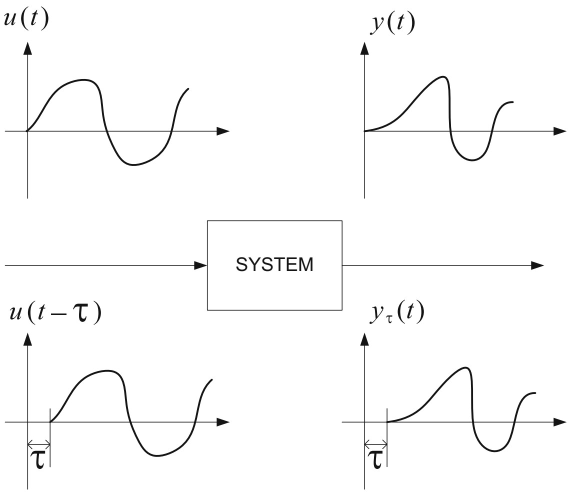


Fig. 1.30 Time invariant system

1.3.3 Examples of the Transfer Characteristics of Some Simple Systems

Next, some examples will demonstrate how to describe mathematically the signal transfer properties of physical systems, i.e., how to give the relationships between the input and the output signals. The description of the behavior of physical systems generally leads to differential equations.

Example 1.1 A mechanical system
Let us consider the mechanical system shown in Fig. 1.31, which can model a part of the chassis of a car. m denotes the mass, c_1 and c_2 are spring constants, and k is the damping coefficient of the oil brake. A concentrated mass is supposed. In the

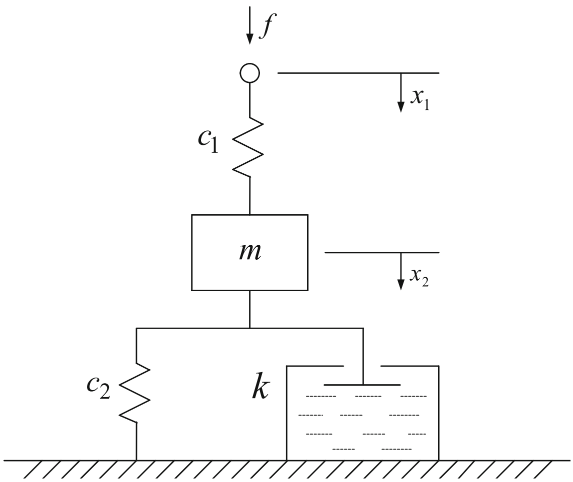


Fig. 1.31 Scheme of a mechanical system

springs, forces proportional to the position are created. The damping piston provides a braking force proportional to the velocity. The following force balance equations can be written. The force created by the upper spring is expressed as $c_1(x_1 - x_2) = f$. The equation expressing the balance of forces acting on the mass is

$$m\frac{d^2x_2}{dt^2} = c_1(x_1 - x_2) - c_2x_2 - k\frac{dx_2}{dt}.$$

It can be seen that the behavior of the system is described by a differential equation. By solving the differential equation, the motions x_1 and x_2 as function of time can be calculated as the responses to the given force. ■

Example 1.2 Direct current (DC) generator
Let us investigate the signal transfer of the externally excited DC generator shown in Fig. 1.32 between its input signal, the excitation voltage u_g, and its output signal, the armature voltage u_k. The resistance of the excitation coil is R_g and its inductance is L_g. The following differential equation can be written for the excitation circuit:

$$L_g\frac{di_g}{dt} + R_g i_g = u_g$$

Assume that the machine works within the linear section of its magnetic characteristic, thus L_g can be considered constant. The generator is not loaded. The terminal voltage of the generator is proportional to the excitation flux, or supposing a linear magnetic characteristics the terminal voltage is proportional to the excitation current: $u_k = K_g i_g$, where K_g is a constant depending on the structural data of the machine, its units are [V/A]. ■

Example 1.3 A chemical process
Let us consider the mixing tank shown in Fig. 1.33. A solution of concentration c_o is mixed with water to obtain a solution of concentration c_k. The amount q_v of the inflow water is constant, the amount q_o of the inflow solution is controlled by a valve. The concentration is given by the amount of the dissolved material in one liter of the solution expressed in grams. The input signal of the system is the

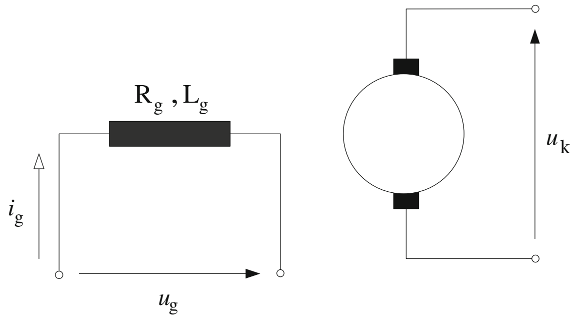


Fig. 1.32 Scheme of an externally excited direct current generator

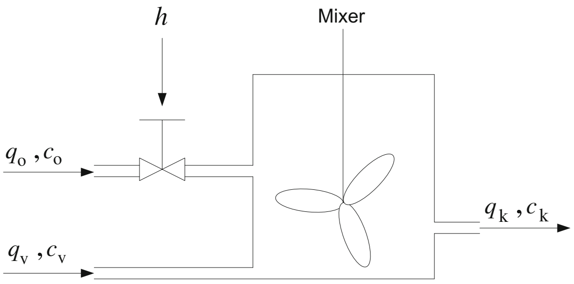


Fig. 1.33 Mixing tank

position h of the plunger, the output signal is the concentration c_k of the obtained solution. The amount of the inflow solution is proportional to the position of the plunger: $q_o = K\,h$. The amount of the outflow solution is the sum of the amount of the inflow solution and the inflow water: $q_k = q_o + q_v$. During time Δt the amount of the dissolved material getting into the tank of volume V is $q_o c_o \Delta t$, and at the same time dissolved material of amount $q_k c_k \Delta t$ leaves the tank. The change of the concentration is:

$$\Delta c_k = \frac{q_o c_o - q_k c_k}{V}\Delta t.$$

The differential equation of the system is obtained by taking the limit $\Delta t \to 0$:

$$\begin{aligned}\frac{dc_k}{dt} + \frac{q_k}{V}c_k &= \frac{dc_k}{dt} + \frac{q_o}{V}c_k + \frac{q_v}{V}c_k \\ &= \frac{dc_k}{dt} + \frac{K}{V}c_k h + \frac{q_v}{V}c_k = \frac{c_o K}{V}h\end{aligned}$$

The relationship is non-linear, as the product of the output signal c_k and the input signal h appears in the equation. But supposing $q_o \ll q_v$, then $q_k \approx q_v =$ constant, and a constant q_k can be taken into consideration in the differential equation. Thus a linear differential equation is obtained.

$$\frac{dc_k}{dt} + \frac{q_v}{V}c_k = \frac{c_o K}{V}h$$

■

1.3.4 Linearization of Static Characteristics

Investigation of non-linear systems is a difficult task. The analysis can be simplified if the non-linear characteristics are linearized in a given vicinity of a working point. Thus in the surrounding of the working point the non-linear system is approximated by a linear model supposing only small changes in the input signals.

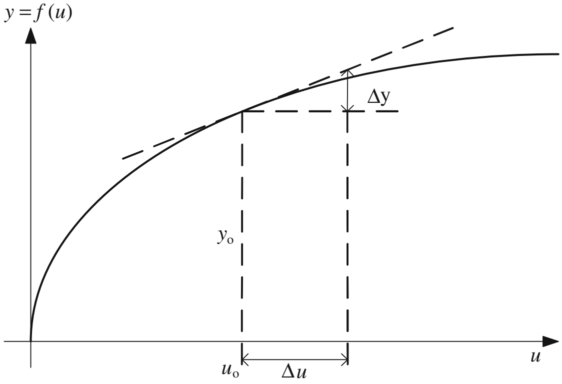


Fig. 1.34 A non-linear static characteristic with single input–single output

Let us consider the non-linear static characteristics $y = f(u)$ shown in Fig. 1.34. At the working point $u = u_o$; $y_o = f(u_o)$ the TAYLOR series of the function is:

$$y = y_o + \Delta y = f(u_o) + f'(u_o)\,(u - u_o) + \cdots$$

Neglecting the higher degree terms, the linearized model is given by

$$y - y_o = \Delta y = f'(u_o)(u - u_o) = f'(u_o)\Delta u$$

The linearized model replaces the static characteristics at the working point by the gradient. Of course the steepness depends on the working point.

Linearization in the case of several inputs

Let the output signal y be a function of the vector of the input variables $\mathbf{u} = [u_1, u_2, \ldots, u_n]^T$. Thus y is a scalar-vector function. Let the vector $\mathbf{u}_o = [u_{1o}, u_{2o}, \ldots, u_{no}]^T$ denote the working point. In a small vicinity of the working point the value of the output signal can be approximated by the TAYLOR expansion

$$y = y_o + \Delta y = f(\mathbf{u}_o) + \sum_{i=1}^{n} \left.\frac{\partial f(\mathbf{u})}{\partial u_i}\right|_{\mathbf{u}_o} (u_i - u_{io}) + \cdots$$
$$= f(\mathbf{u}_o) + \left[\left.\frac{d f(\mathbf{u})}{d\,\mathbf{u}}\right|_{\mathbf{u}_o}\right]^T (\mathbf{u} - \mathbf{u}_o) + \cdots$$

Neglecting the second and higher order derivatives, the small change in the function $f(\boldsymbol{u})$ around the working point can be given by the following linear relationship:

$$\Delta y = \sum_{i=1}^{n} A_i\,\Delta u_i.$$

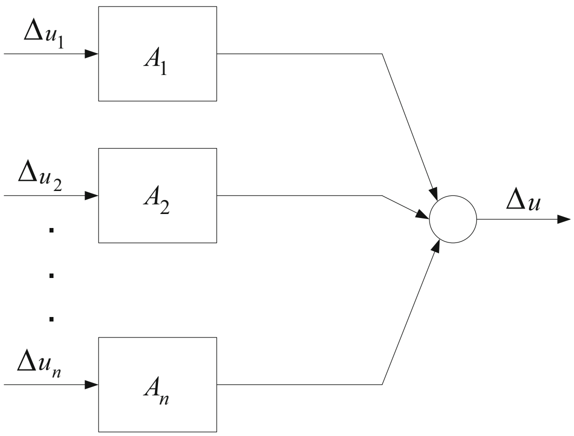


Fig. 1.35 Linearization of multi-input single-output static characteristics

The linearized block diagram is shown in Fig. 1.35. The A_i coefficients are the so called static transfer coefficients of the linearized model, whose values depend on the working point.

Example 1.4 Linearization of the moment equation of a DC motor

The moment m in a direct current (DC) motor is proportional to the product of the flux φ in the excitation coil and the armature current i (Fig. 1.36). The product of these two changing variables results in a non-linear relationship.

$$m = m_o + \Delta m = k\varphi i = k\varphi_o i_o + \left.\frac{\partial m}{\partial \varphi}\right|_{\varphi_o, i_o} \Delta\varphi + \left.\frac{\partial m}{\partial i}\right|_{\varphi_o, i_o} \Delta i$$

Determining the derivatives and considering that the value of the moment in the working point is $m_o = k\varphi_o i_o$, the change of the moment around the working point can be calculated according to the following relationship: $\Delta m = k i_o \Delta\varphi + k\varphi_o \Delta i$. ■

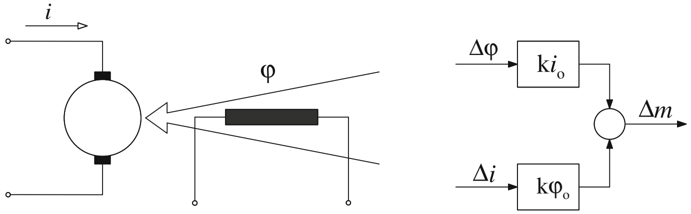


Fig. 1.36 The moment in the DC motor is proportional to the product of the excitation flux and the armature current

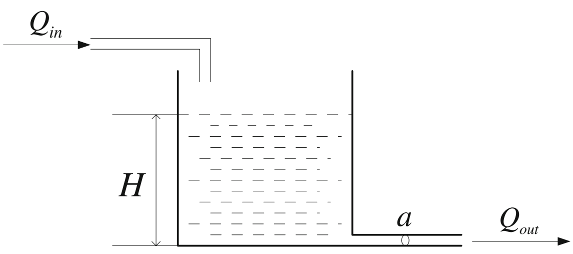


Fig. 1.37 Setting the liquid level in a tank

Example 1.5 Linearization of the tank equation
In a tank, the increase of the liquid level depends on the difference between the flow rate of the input liquid and that of the output liquid (Fig. 1.37). Let us denote the input flow by Q_{in}, and the output flow by Q_{out}, respectively. The cross section of the tank is denoted by A, and the cross section of the outflow tube is denoted by a. The liquid level is H. The change of the liquid level is described by the following differential equation:

$$A\frac{dH}{dt} = Q_{in} - Q_{out}$$

The output liquid flow depends on the velocity v of the outflow, which is proportional to the square root of the level.

$$Q_{out} = av = a\sqrt{2gH} = \beta\sqrt{H}$$

In steady-state, the level does not change, so the input and output flows are equal: $Q_{in} = Q_{out}$.

The steady-state value of the level will be $H = Q_{in}^2/\beta^2$. The static characteristic of the tank, viz., the relationship between the liquid level and the input flow, is non-linear (Fig. 1.38).

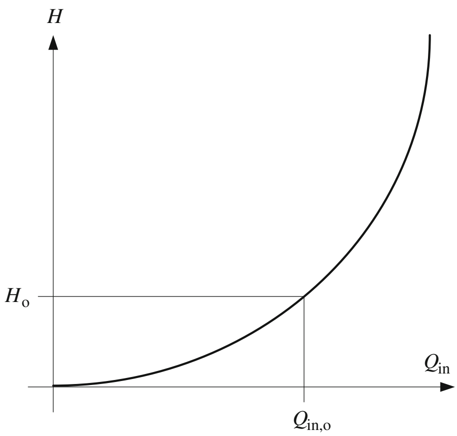


Fig. 1.38 Static characteristics of the tank

Let us denote the values of the working points by the index zero, and the changes around the working point with lower case letters

$$H = H_o + h$$
$$Q_{in} = Q_{in,o} + q_{in}$$

The outflow can be expressed with the first order TAYLOR approximation of the square root expression as

$$Q_{out} = \beta\sqrt{H} \approx \beta\sqrt{H_o} + \beta\frac{1}{2\sqrt{H_o}}h$$

The differential equation expressed with the working point values and the small changes around them is:

$$A\frac{\mathrm{d}\,(H_o + h)}{\mathrm{d}t} = Q_{in,o} + q_{in} - \beta\sqrt{H_o} - \frac{\beta}{2\sqrt{H_o}}h$$

As the derivative of a constant H_o working point value is zero, and $Q_{in,o} = \beta\sqrt{H_o}$, for the small changes around the working point the following differential equation can be given:

$$A\frac{\mathrm{d}h}{\mathrm{d}t} = q_{in} - \frac{\beta}{2\sqrt{H_o}}h$$

This is a linear differential equation whose parameters depend on the working point. ■

1.3.5 Relative Units

The transfer factors (gains) of the elements in a control system have dimensions. In the previous example of the liquid tank, the units of the working-point-dependent transfer gain resulting from the static characteristics is cm/(l/min). In the case of a motor, the output signal is the speed, the input signal is the voltage, thus the dimension of the transfer gain is (rad/s)/V. If the actual values of both the input and the output signals are related to their maximum values, the signals can be given with dimensionless relative values, which are between 0 and 1. The signals to be compared should be normalized identically. For example, the maximum values of the reference signal, the controlled signal and the error signal have to be the same.

Quantities with the dimension of time can also be given with relative values, if they are related to a maximum value chosen for the time variable.

As an example, let us consider the construction shown in Fig. 1.39. The DC motor M moves the rod R through transmission gears. The input signal of the motor

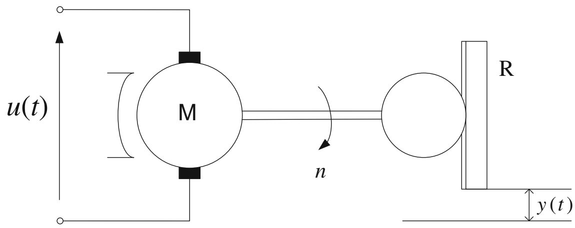


Fig. 1.39 Position control

is its terminal voltage $u(t)$, and its output signal is the position $y(t)$ of the rod (plunger). Neglecting the transients the displacement of the rod is proportional to the integral of the speed of the motor, and the speed is proportional to the terminal voltage. If the application of a terminal voltage of 200 V produces displacement of the rod by 5 cm within 10 s, then after time t the displacement is

$$y(t) = \frac{5\ \text{cm}}{10\ \text{s} \cdot 200\ \text{V}} \int_0^t u(t)\,\text{d}t = 2.5 \cdot 10^{-3} \frac{\text{cm}}{\text{V s}} \int_0^t u(t)\,\text{d}t$$

as the effect of the input voltage $u(t)$.

Let us take $t_{\text{max}} = 50$ s as the unit of time, $y_{\text{max}} = 20$ cm as the unit position and $u_{\text{max}} = 200$ V as the unit of voltage. The relative units related to their basic units are:

$$t_{\text{rel}} = \frac{t}{t_{\text{max}}} = \frac{t}{50\ \text{s}}; \quad y_{\text{rel}} = \frac{y}{y_{\text{max}}} = \frac{y}{20\ \text{cm}}; \quad u_{\text{rel}} = \frac{u}{u_{\text{max}}} = \frac{u}{200\ \text{V}}$$

With relative units the displacement of the rod can be given by the following relationship:

$$y_{\text{rel}}(t) = \frac{\frac{5\ \text{cm}}{20\ \text{cm}}}{\frac{10\ \text{s}}{50\ \text{s}} \cdot \frac{200\ \text{V}}{200\ \text{V}}} \int_0^{t_{\text{rel}}} u_{\text{rel}}(t)\,\text{d}t_{\text{rel}} = 1.25 \int_0^{t_{\text{rel}}} u_{\text{rel}}(t)\,\text{d}t_{\text{rel}}$$

1.4 Practical Aspects

The design and implementation of a control system is an iterative task. First the requirements set for the control system have to be formulated. Then based on the physical operation of the process, its mathematical model is established, whose parameters are determined by measurements and identification procedures. The controller is designed for the process model considering the given requirements. Then the operation of the control system is checked by simulation. If necessary, the controller is redesigned. During the implementation, the adjustment of the controller is refined.

In a control problem three basic tasks may occur.

It is necessary to create the model P of the process: the signal transfer properties of each element have to be determined based on the physical relationships describing the behavior of the element, or from its input and output measurement data by *identification* (Figs. 1.40 and 1.41).

If the input signal and the element P are known, the output signal can be determined and the behavior of the element can be analyzed (Fig. 1.42).

If the element P is given and the course of its required output signal is prescribed, then the task is to determine the input signal which ensures this behavior. The input of the plant is created by a control circuit. This is the synthesis or controller design task (Fig. 1.42).

Control engineering is an interdisciplinary area of science. The operation of the process is to be understood, to do this there is a need of knowledge of physical, chemical, biological, etc. phenomena. Mathematical knowledge is required for system modeling as well as the analysis and synthesis of control systems. To investigate the operation of control systems, knowledge is needed about signals, systems, and the behavior of systems with negative feedback. During the design, rational considerations and basic restrictions also have to be taken into account. The design has to cover economic, safety, environmental protection, etc. aspects as well. To fulfill a more complex control task, the coordinated work of different professionals is needed.

During the realization, the state of the system has to be observed—the considered output signal has to be measured by the appropriate measuring equipment, it is required to manipulate the process input—an actuator has to be selected. The measurement noise of the sensors, the signal ranges of the actuators, the limits of the produced actuating effects, all have to be taken into account. Several times the measured data have to be transferred across longer distances, thus data transfer has to be ensured. There are standards, so called protocols for data transfer which have

Fig. 1.40 Identification

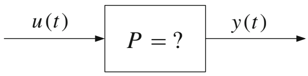


Fig. 1.41 Analysis

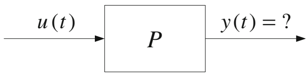
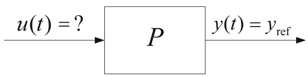


Fig. 1.42 Synthesis



to be considered. The control signal has to be determined with an appropriate calculation algorithm, and has to be forwarded to the input of the process. In the design of the control algorithm the disturbances acting on the process, the uncertainties in the process parameters and also the restrictions due to practical realization have to be taken into account. During the control, real data are elaborated and real time signal transfer is realized. In signal transfer, non-deterministic signal delays do appear, which may distort the operation. The connection and exchange of information between the individual elements have to be addressed using appropriate interface elements.

Besides the continuous-time control systems computer control systems have gained more and more applications. The process and the process controller computer are connected via A/D (analog to digital) and D/A (digital to analog) converters. The computer executes the essential control functions in real time, repeatedly at the sampling instances. In industrial process control systems, distributed control systems are implemented, where spatially distributed control systems operate in an aligned fashion, communicating with each other.