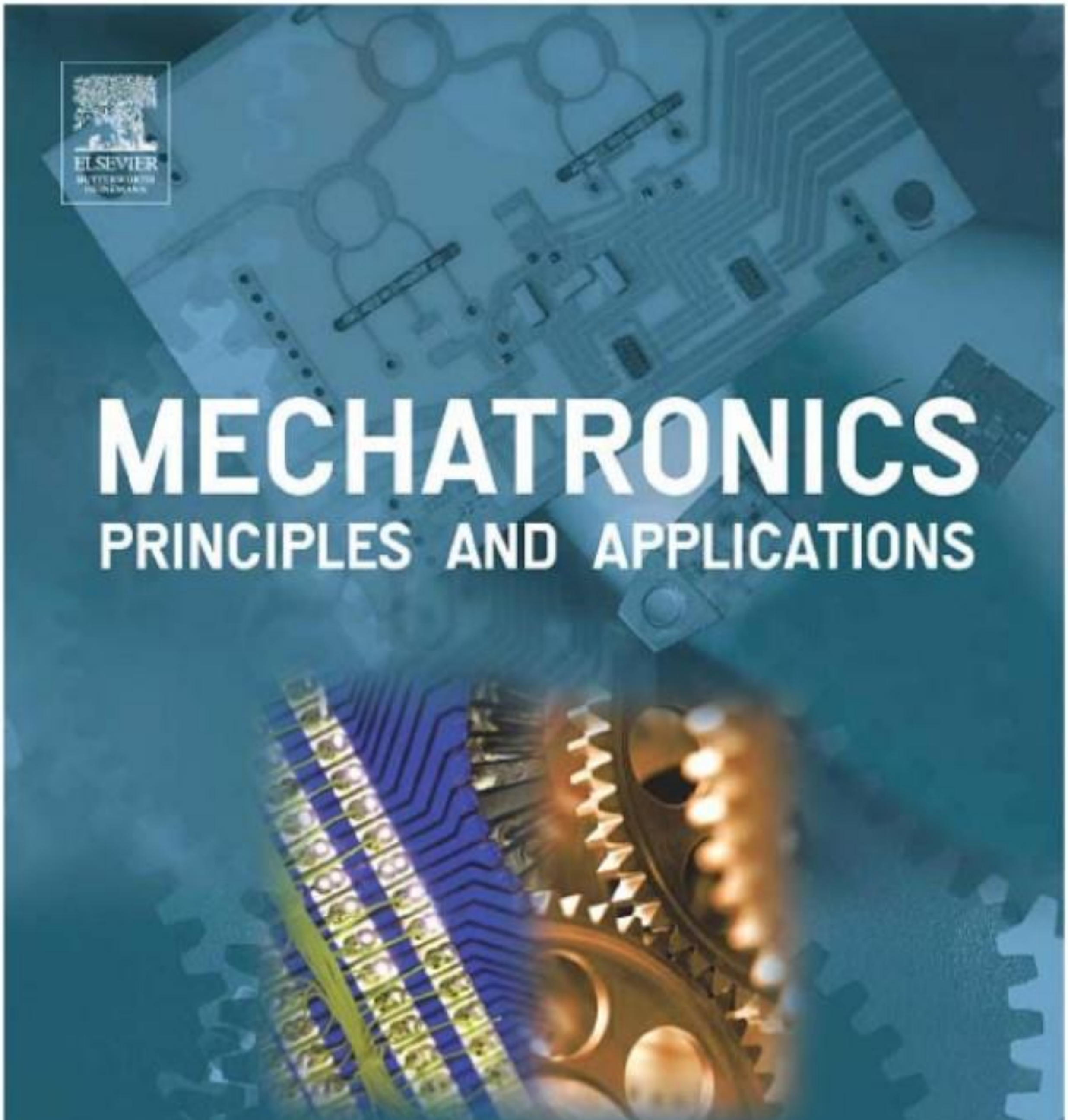
ELSEVIER
MECHATRONICS
PRINCIPLES AND APPLICATIONS
GODFREY ONWUBOLU

CHAPTER 1

Introduction to mechatronics

Chapter objectives

When you have finished this chapter you should be able to:

- trace the origin of mechatronics;
- understand the key elements of mechatronics systems;
- relate with everyday examples of mechatronics systems;
- appreciate how mechatronics integrates knowledge from different disciplines in order to realize engineering and consumer products that are useful in everyday life.

1.1 Historical perspective

Advances in microchip and computer technology have bridged the gap between traditional electronic, control and mechanical engineering. Mechatronics responds to industry's increasing demand for engineers who are able to work across the discipline boundaries of electronic, control and mechanical engineering to identify and use the proper combination of technologies for optimum solutions to today's increasingly challenging engineering problems. All around us, we can find mechatronic products. Mechatronics covers a wide range of application areas including consumer product design, instrumentation, manufacturing methods, motion control systems, computer integration, process and device control, integration of functionality with embedded microprocessor control, and the design of machines, devices and systems possessing a degree of computer-based intelligence. Robotic manipulators, aircraft simulators, electronic traction control systems, adaptive suspensions, landing gears, air conditioners under fuzzy logic control, automated diagnostic systems, micro electromechanical systems (MEMS), consumer products such as VCRs, and driver-less vehicles are all examples of mechatronic systems.

The genesis of mechatronics is the interdisciplinary area relating to mechanical engineering, electrical and electronic engineering, and computer science. This technology has produced many new products and provided powerful ways of improving the efficiency of the products we use in our daily life. Currently, there is no doubt about the importance of mechatronics as an area in science and technology. However, it seems that mechatronics is not clearly understood; it appears that some people think that mechatronics is an aspect of science and technology which deals with a system that includes mechanisms, electronics, computers, sensors, actuators and so on. It seems that most people define mechatronics by merely considering what components are included in the system and/or how the mechanical functions are realized by computer software. Such a definition gives the impression that it is just a collection of existing aspects of science and technology such as actuators, electronics, mechanisms, control engineering, computer technology, artificial intelligence, micro-machine and so on, and has no original content as a technology. There are currently several mechatronics textbooks, most of which merely summarize the subject picked up from existing technologies. This structure also gives people the impression that mechatronics has no unique technology. The definition that mechatronics is simply the combination of different technologies is no longer sufficient to explain mechatronics.

Mechatronics solves technological problems using interdisciplinary knowledge consisting of mechanical engineering, electronics, and computer technology. To solve these problems, traditional engineers used knowledge provided only in one of these areas (for example, a mechanical engineer uses some mechanical engineering methodologies to solve the problem at hand). Later, due to the increase in the difficulty of the problems and the advent of more advanced products, researchers and engineers were required to find novel solutions for them in their research and development. This motivated them to search for different knowledge areas and technologies to develop a new product (for example, mechanical engineers tried to introduce electronics to solve mechanical problems). The development of the microprocessor also contributed to encouraging the motivation. Consequently, they could consider the solution to the problems with wider views and more efficient tools; this resulted in obtaining new products based on the integration of interdisciplinary technologies.

Mechatronics gained legitimacy in academic circles with the publication of the first refereed journal: *IEEE/ASME Transactions on Mechatronics*. In it, the authors worked tenaciously to define mechatronics. Finally they coined the following:

> The synergistic combination of precision mechanical engineering, electronic control and systems thinking in the design of products and manufacturing processes.

This definition supports the fact that mechatronics relates to the design of systems, devices and products aimed at achieving an optimal balance between basic mechanical structure and its overall control.

1.2 Key elements of a mechatronic system

It can be seen from the history of mechatronics that the integration of the different technologies to obtain the best solution to a given technological problem is considered to be the essence of the discipline. There are at least two dozen definitions of mechatronics in the literature but most of them hinge around the 'integration of mechanical, electronic, and control engineering, and information technology to obtain the best solution to a given technological problem, which is the realization of a product'; we follow this definition. Figure 1.1 shows the main components of a mechatronic system. This book covers the principles and applications of mechatronic systems based on this framework. As can be seen, the key element of mechatronics are electronics, digital control, sensors and actuators, and information technology, all integrated in such a way as to produce a real product that is of practical use to people.

The following subsections outline, very briefly, some fundamentals of these key areas. For fuller discussions the reader is invited to explore the rich and established information sources available on mechanics, electrical and electronic theory, instrumentation and control theory, information and computing theory, and numerical techniques.

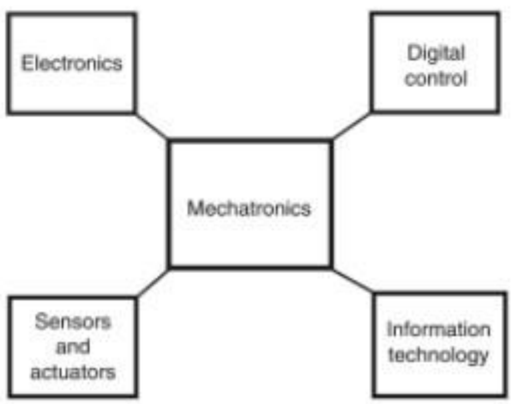

Figure 1.1 Main components of a mechatronic system.

1.2.1 Electronics

1.2.1.1 *Semiconductor devices*

Semiconductor devices, such as diodes and transistors, have changed our lives since the 1950s. In practice, the two most commonly used semiconductors are germanium and silicon (the latter being most abundant and cost-effective). However, a semiconductor device is not made from simply one type of atom and impurities are added to the germanium or silicon base. These impurities are highly purified tetravalent atoms (e.g. of boron, aluminum, gallium, or indium) and pentavalent atoms (e.g. of phosphorus, arsenic, or antimony) that are called the doping materials. The effects of doping the semiconductor base material are 'free' (or unbonded) electrons, in the case of pentavalent atom doping, and 'holes' (or vacant bonds), in the case of tetravalent atoms.

An n-type semiconductor is one that has an excess number of electrons. A block of highly purified silicon has four electrons available for covalent bonding. Arsenic, for, example, which is a similar element, has five electrons available for covalent bonding. Therefore, when a minute amount of arsenic is mixed with a sample of silicon (one arsenic atom in every 1 million or so silicon atoms), the arsenic atom moves into a place normally occupied by a silicon atom and one electron is left out in the covalent bonding. When external energy (electrical, heat, or light) is applied to the semiconductor material, the excess electron is made to 'wander' through the material. In practice, there would be several such extra negative electrons drifting through the semiconductor. Applying a potential energy source (battery) to the semiconductor material causes the negative terminal of the applied potential to repulse the free electrons and the positive terminal to attract the free electrons.

If the purified semiconductor material is doped with a tetravalent atom, then the reverse takes place, in that now there is a deficit of electrons (termed 'holes'). The material is called a p-type semiconductor. Applying an energy source results in a net flow of 'holes' that is in the opposite direction to the electron flow produced in n-type semiconductors.

A semiconductor diode is formed by 'joining' a p-type and n-type semiconductor together as a p–n junction (Figure 1.2).

Initially both semiconductors are totally neutral. The concentration of positive and negative carriers is quite different on opposite sides of the junction and a thermal energy-powered diffusion of positive carriers into the n-type material and negative carriers into the p-type material occurs. The n-type material acquires an excess of positive charge near the junction and the p-type material acquires an excess of negative charge. Eventually diffuse charges build up and an electric field is created which drives the minority charges and eventually equilibrium is reached. A region develops at the junction called the *depletion layer*. This region is essentially 'un-doped' or just intrinsic silicon. To complete the diode conductor, lead materials are placed at the ends of the p–n junction.

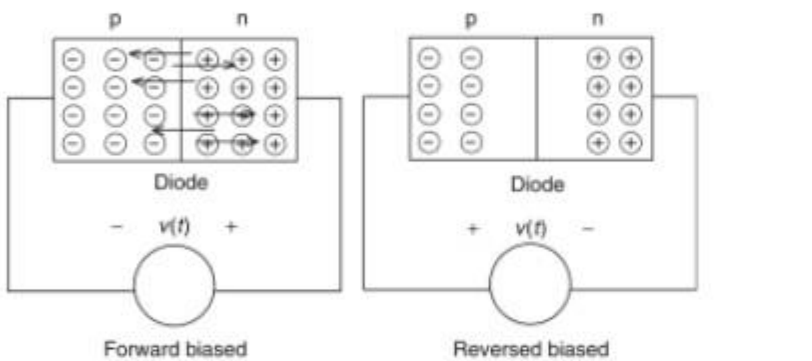

Figure 1.2 p–n junction diode.

Transistors are active circuit elements and are typically made from silicon or germanium and come in two types. The bipolar junction transistor (BJT) controls current by varying the number of charge carriers. The field-effect transistor (FET) varies the current by varying the shape of the conducting volume.

By placing two p–n junctions together we can create the bipolar transistor. In a pnp transistor the majority charge carriers are holes and germanium is favored for these devices. Silicon is best for npn transistors where the majority charge carriers are electrons.

The thin and lightly doped central region is known as the *base* (B) and has majority charge carriers of opposite polarity to those in the surrounding material. The two outer regions are known as the *emitter* (E) and the *collector* (C). Under the proper operating conditions the emitter will emit or inject majority charge carriers into the base region, and because the base is very thin, most will ultimately reach the collector. The emitter is highly doped to reduce resistance. The collector is lightly doped to reduce the junction capacitance of the collector–base junction.

The schematic circuit symbols for bipolar transistors are shown in Figure 1.3. The arrows on the emitter indicate the current direction, where $I_E = I_B + I_C$. The collector is usually at a higher voltage than the emitter. The emitter–base junction is forward biased while the collector–base junction is reversed biased.

1.2.2 Digital control

1.2.2.1 *Transfer function*

A transfer function defines the relationship between the inputs to a system and its outputs. The transfer function is typically written in the frequency (or *s*) domain,

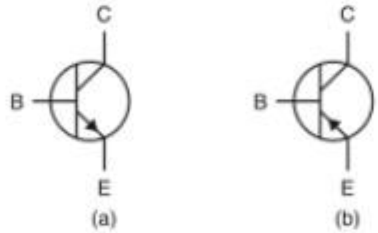

Figure 1.3 (a) npn bipolar transistor; (b) pnp bipolar transistor.

rather than the time domain. The Laplace transform is used to map the time domain representation into the frequency domain representation.

If $x(t)$ is the input to the system and $y(t)$ is the output from the system, and the Laplace transform of the input is $X(s)$ and the Laplace transform of the output is $Y(s)$, then the transfer function between the input and the output is

$$\frac{Y(s)}{X(s)}. \tag{1.1}$$

1.2.2.2 *Closed-loop system*

A closed-loop system includes feedback. The output from the system is fed back through a controller into the input to the system. If $G_u(s)$ is the transfer function of the uncontrolled system, and $G_c(s)$ is the transfer function of the controller, and unity (negative) feedback is used, then the closed-loop system block diagram (Figure 1.4) is expressed as:

$$Y(s) = \frac{G_c(s)G_u(s)}{1 + G_c(s)G_u(s)}X(s). \tag{1.2}$$

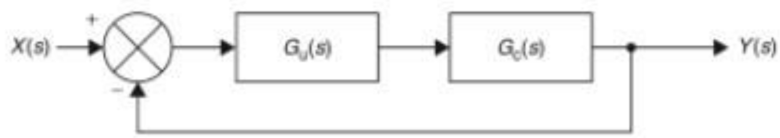

Figure 1.4 Block diagram of closed-loop system with unity gain.

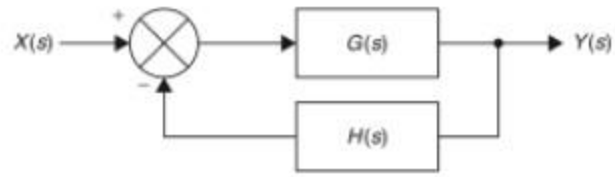

Figure 1.5 Block diagram of closed-loop system with transfer function in feedback loop.

Sometimes a transfer function, $H(s)$, is included in the feedback loop (Figure 1.5). For negative feedback this is expressed as:

$$Y(s) = \frac{G(s)}{1 + H(s)G(s)} X(s). \tag{1.3}$$

1.2.2.3 *Forward-loop system*

A forward-loop system (Figure 1.6) is a part of a controlled system. As the name suggests, it is the system in the 'forward' part of the block diagram shown in Figure 1.4. Typically, the forward-loop includes the uncontrolled system cascaded with the controller. Closing the loop around this controller and system using unity feedback gain yields the closed-loop system. For a system with controller $G_c(s)$ and system $G_u(s)$, the transfer function of the forward-loop is:

$$Y(s) = G_c(s)G_u(s)X(s). \tag{1.4}$$

1.2.2.4 *Open-loop system*

An open-loop system is a system with no feedback; it is an uncontrolled system. In an open-loop system, there is no 'control loop' connecting the output of the system to its input. The block diagram (Figure 1.7) can be represented as:

$$Y(s) = G(s)X(s). \tag{1.5}$$

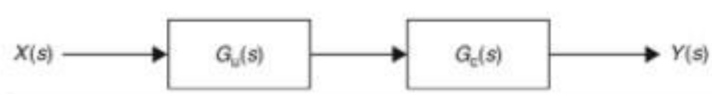

Figure 1.6 Forward-loop part of Figure 1.4.

Figure 1.7 Block diagram of open-loop system.

1.2.3 Sensors and actuators

1.2.3.1 *Sensors*

Sensors are elements for monitoring the performance of machines and processes. The common classification of sensors is: distance, movement, proximity, stress/strain/force, and temperature. There are many commercially available sensors but we have picked the ones that are frequently used in mechatronic applications. Often, the conditioned signal output from a sensor is transformed into a digital form for display on a computer or other display units. The apparatus for manipulating the sensor output into a digital form for display is referred to as a measuring instrument (see Figure 1.8 for a typical computer-based measuring system).

1.2.3.2 *Electrical actuators*

While a sensor is a device that can convert mechanical energy to electrical energy, an electrical actuator, on the other hand, is a device that can convert electrical energy to mechanical energy. All actuators are transducers (as they convert one form of energy into another form). Some sensors are transducers (e.g. mechanical actuators), but not all. Actuators are used to produce motion or action, such as linear motion or angular motions. Some of the important electrical actuators used in mechatronic systems include solenoids, relays, electric motors (stepper, permanent magnet, etc.). These actuators are instrumental in moving physical objects in mechatronic systems.

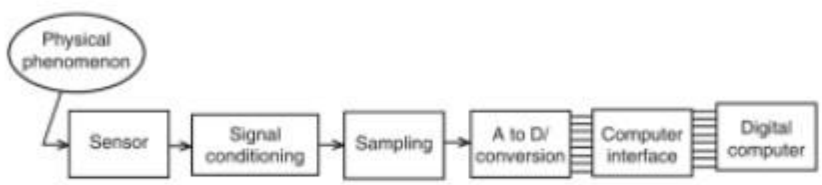

Figure 1.8 Measurement system.

1.2.3.3 *Mechanical actuators*

Mechanical actuators are transducers that convert mechanical energy into electrical energy. Some of the important mechanical actuators used in mechatronic systems include hydraulic cylinders and pneumatic cylinders.

1.2.4 Information technology

1.2.4.1 *Communication*

Signals to and from a computer and its peripheral devices are often communicated through the computer's serial and parallel ports. The parallel port is capable of sending (12 bits per clock cycle) and receiving data (up to 9 bits per clock cycle). The port consists of four control lines, five status lines, and eight data lines. Parallel port protocols were recently standardized under the IEEE 1284 standard. These new products define five modes of operation such as:

- Compatibility mode
- Nibble mode
- Byte mode
- EPP mode (enhanced parallel port)
- ECP mode (extended capabilities mode)

This is the concept on which the PC printer operates. Therefore, the code required to control this port is similar to that which makes a printer operate. The parallel port has two different modes of operation: The standard parallel port (SPP) mode and the enhanced parallel port (EPP) mode. The SPP mode is capable of sending and receiving data. However, it is limited to only eight data lines.

The EPP mode provides 16 lines with a typical transfer rate in the order of $500\,kB\,s^{-1}$ to $2\,MB\,s^{-1}$ (WARP). This is achieved by hardware handshaking and strobing of the data, whereas, in the SPP mode, this is software controlled.

In order to perform a valid exchange of data using EPP, the EPP handshake protocol must be followed. As the hardware does all the work required, the handshake only needs to work for the hardware. Standard data read and write cycles have to be followed while doing this.

Engineers designing new drivers and devices are able to use the standard parallel port. For instance, EPP has its first three software registers as Base + 0, Base + 1, Base + 2 as indicated in Table 1.1. EPP and ECP require additional hardware to handle the faster speeds, while Compatibility, Byte, and Nibble mode use the hardware available on SPP.

Compatibility modes send data in the forward direction at a rate of $50\text{–}150\,kb\,s^{-1}$, i.e. only in data transmission. In order to receive the data the