Electrovalves: principle of operation

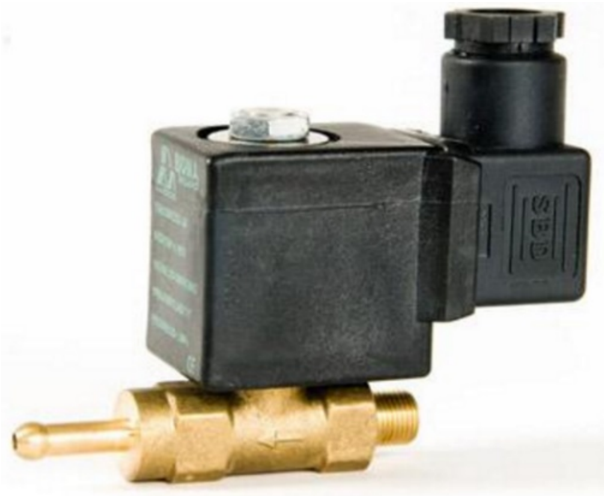
Introduction

In this document, the base of the operation and some tricks well to know about electrovalves are explained: the goal is not to provide a complete picture of the theory but to explain the physics behind the most common phenomena.

1 Principle of working

In order to understand the nature of the driving signals and the control strategy adopted for our valve driver current control systems, it is fundamental to learn how an electrovalve is made and what its principle of working is. There are several types of electrovalves and different manufacturing approaches. Our intention is not to deeply analyze such fields but to give the reader a complete picture of the working principle of this device, so that each phase of the injection process can be understood and linked to the intervention of the external electrical driving signal.

Figure 1. Assembled electrovalve



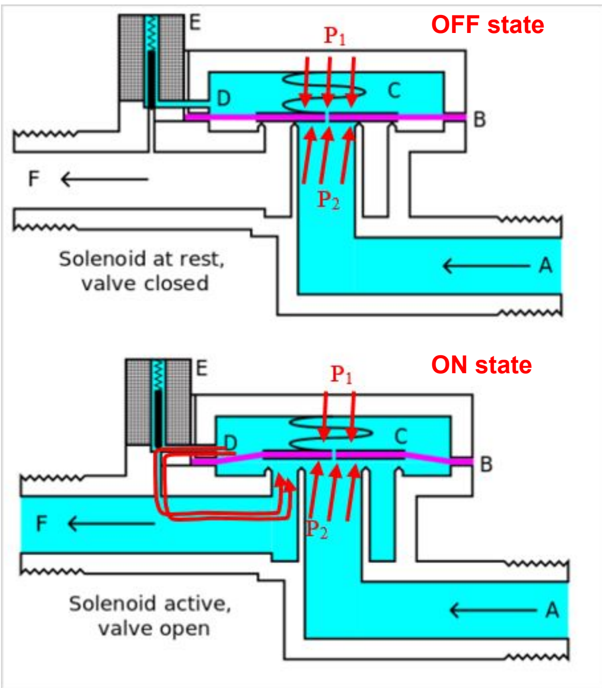
The Figure 1 shows an assembled electrovalve, while the Figure 2 is useful to describe what happens in the internal chambers during a normal toggling cycle. Although more complex interactions can happen, we are interested in the simplest one: the *ON-OFF* mechanism. The process can be easily described as follows:

- During the *OFF* state the diaphragm (also called as needle) *B* is closed and the fluid cannot flow from the input *A* to the output *F*, but it remains confined in the pressure chamber *C*;
- During the *ON* state the diaphragm B is opened and the fluid can flow from the input *A* to the output *F* passing through the pressure chamber *C*, where it was previously stored.

The presence of the pressure relief passage *D* is essential in order to avoid pressure becoming excessively high in the pressure chamber *C* during the *OFF* state. During such a period, liquid reaches the pressure chamber flowing from the input through the small aperture in the diaphragm. Thanks to the relief, pressure excess is controlled and system reaches equilibrium: therefore the forces acting on the small membrane from both sides are equal and the diaphragm keeps closed ($P_1 = P_2$ in the upper part of the Figure 2). Concluding the analysis of the *OFF* state, it can be observed that the fluid is in direct contact with the solenoid *E* through the pressure relief passage. Such a necessary interaction can produce corrosion of the core and it's one of the main causes of device aging. Therefore, it is really important choosing the right material for the core: it must have a high magnetic permeability and has to be compatible with the fluid.

The transition between *OFF* and *ON* state happens when the solenoid is excited by means of an electrical signal: such a signal generates a magnetic field inside the core, whose resulting force is able to shift the core needle upwards, allowing the fluid to expand itself downwards.

Figure 2. **ON-OFF electrovalve: toggling cycle**



A- Input side
B- Diaphragm
C- Pressure chamber
D- Pressure relief passage
E- Solenoid
F- Output side

As can be seen in the Figure 2, following the core shifting, the fluid expansion generates an additive force directed upwards on the diaphragm. Hence, the imbalance between P_1 and P_2 is able to win the elastic resistance and leads to the full opening of the diaphragm. Therefore, the fluid can now freely flow from the input *A* towards the output side *F*. As the pressure relief passage covers an important role during the *OFF* state, the presence of the spring is essential for the correct functionality of the device. In fact, when the electrical stimulus is *inverted* (instead of stopped) and the core is shifted downwards to the initial position, the pressure P_2 returns to low levels. Hence, the sum of P_1 and the pressure generated by the spring expansion are enough to shift the diaphragm downwards, thus closing again the pressure chamber. The spring must be designed in order to resist to the maximum fluid pressure that occurs at system startup. When the liquid is first introduced in the electrovalve through the input *A*, the spring must resist to the initial force pulse directed upwards, and stay still in the closed position. Otherwise, if the diaphragm opens at system startup, elastic force could be never able to recover the spring from the open position, leading to valve failure

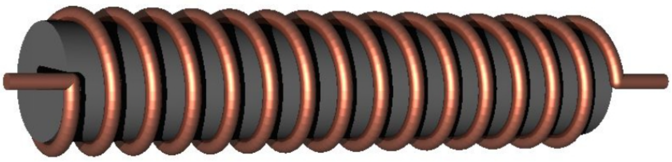
Before concluding our brief analysis on the working principle of an electrovalve, it is important to highlight one detail which can make a big difference in real applications: the inversion of the electrical signal that drives the electrovalve when switching from *ON* to *OFF* state and vice versa. In fact, once the core has been shifted upwards in the *OFF* → *ON* transition, we can not rely on gravity force (also because we don't know how the valve will be positioned in the mechanical system) to bring it to the initial position. Moreover, even if solenoid was placed vertically, it is most likely that pressure in the relief passage is so high to win gravity, keeping the core from going back to the initial position. Therefore, by inverting the electrical signal (and therefore the magnetic force), we are sure to correctly re-position the core at each valve cycle. In the following chapters we will also see how keeping the core always in movement by means of signals more complex than simple current inversion can be a good strategy in order to win the inertia of the two quiet points and optimize performances.

2 Ideal electrical model

Having explained what the working principle of an electrovalve is, our analysis will now be more focused on the electrical phenomena that occurs when driving the device. Starting from the basic physical principle which regulates its functioning, we will then analyze the electrical model of the electrovalve and all the parasitic effects that affect the ideal behavior.

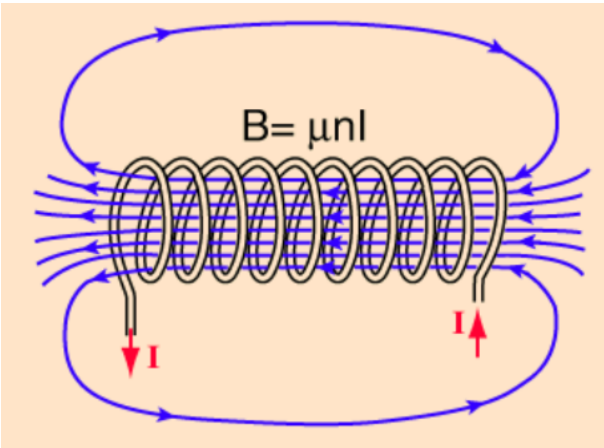
From the electrical point of view, the electrovalve represents essentially an inductive load due to the presence of the solenoid used for opening/closing the valve. As shown in the Figure 3, a solenoid is a coil made of a conductive material (e.g. copper or aluminum).

Figure 3. Solenoid with ferromagnetic core



Supplying a current intensity *I*, a magnetic field **H** and its induction vector **B** will be generated in both the internal section and the outside area surrounding the coil, as shown in the Figure 4. However, such a magnetic field **H** is concentrated into a nearly uniform axial field in the internal section, while it is weak and divergent outside of it.

Figure 4. Magnetic fiedl generated by a solenoid



The intensity of the internal magnetic induction is reported in the Eq. (1), showing that it is directly proportional to the current intensity *I*, the coil number *N* and the magnetic permeability *μ* of the material forming the inner section of the solenoid. Instead, it is inversely proportional to the solenoid length *l*.

Magnetic induction intensity in the inner section of a solenoid

$$B = \mu n I = \mu \frac{N}{l} I \quad (1)$$

The direction of the **B** vector is axial and parallel to **H**. The *magnetic flux* (equivalent of the electrical current in the electric circuits) inside a solenoid is defined as per Eq. (2):

Magnetic flux inside a solenoid

$$\Phi = N{*}B{*}S = \mu \frac{N^2 S}{l} I \quad (2)$$

Where *S* represents the internal section of the coil, and *B* has been extracted from the Eq. (1).

The *auto-induction coefficient*, better known as *inductance L*, is defined as the ratio between the magnetic flux Φ and the electrical current I (see Eq. (3)). It measures the solenoid capability of generating a magnetic field when supplied with an electrical current. Having fixed a target magnetic field, the higher the inductance, the less the current needed to generate it.

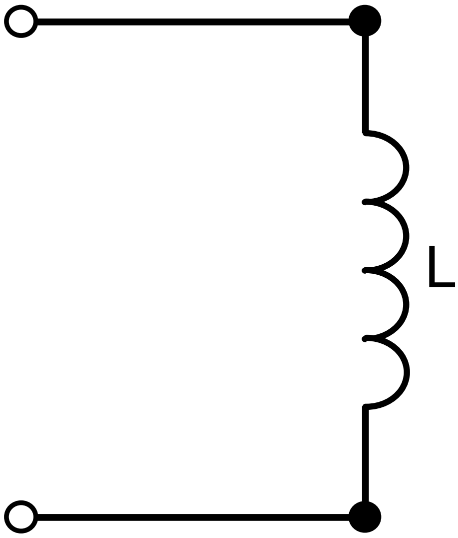
Definition of auto-induction coefficient (inductance) for a solenoid

$$L = \frac{\phi}{I} = \mu\frac{N^2 S}{l} \tag{3}$$

In order to obtain high values of inductance, *ferromagnetic* materials as permalloy can be used as the core of the solenoid, onto which the coil of conductive material is wrapped (see Figure 3). These materials are in fact characterized by the high values of magnetic permeability $\mu \approx 10^{-2}$ H/m ($\mu 0 \approx 10^{-6}$ H/m).

Hence, a first ideal representation for the electrovalve in an electric system would be a simple inductance L (see Figure 5), whose value depends on the inner core material (μ), inner section S, coil number N and solenoid length l.

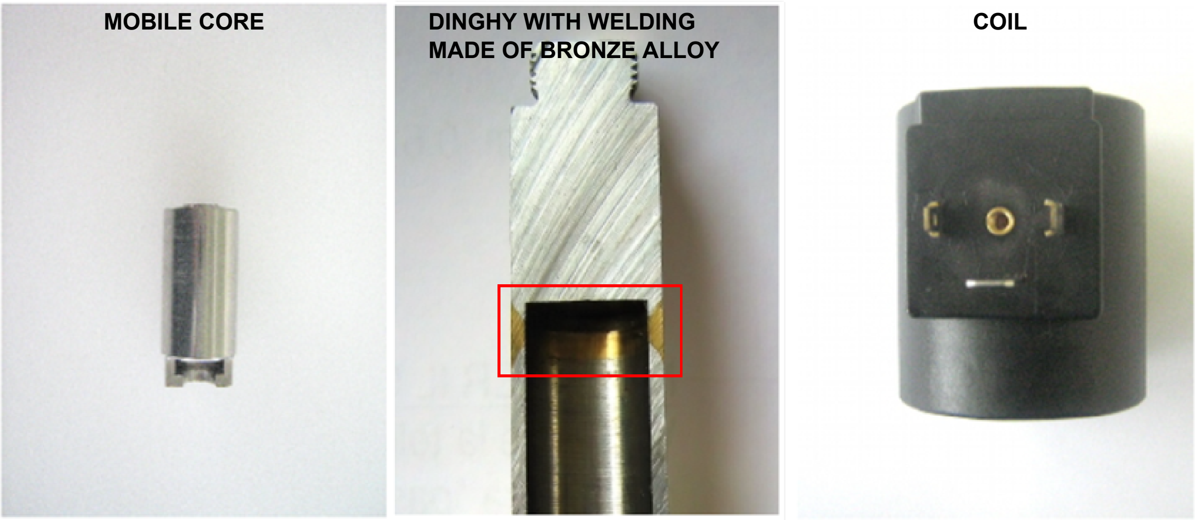
Figure 5. Ideal electrical model



3 Main components of the opening circuit

Such a simple representation is enough to explain the physical phenomena that occur during the opening/closure phase of the electrovalve. The “active” part of an electrovalve from the electrical point of view consists essentially of three components (see Figure 6):

Figure 6. Electrovalve active part: electrical point of view



The active part are the following:

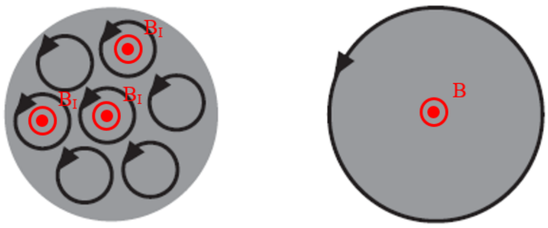
- *Mobile core (needle)*: it is the moving element that allows the opening/closure of the valve. It shifts up and down inside the *dinghy* under the effect of the magnetic field generated by the *coil*.
- *Dinghy*: it is the part containing the needle and acts as a guide for it. In order to direct magnetic field lines onto the *mobile core*, the magnetic circuit is interrupted by interposing a *nonmagnetic material* such as bronze alloy (the yellow welding highlighted in the Figure 6). This material prevents field lines from going outside the cavity by making them converging onto the axis of the device. In this way all (ideally) the magnetic force generated by the *coil* is directed on the *mobile core*.
- *Coil*: it is the actual solenoid described above, responsible for the generation of the magnetic field. It acts as a magnet used for attracting or repelling the *mobile core* during the opening/closure phase.

4 Losses and real electrical model

Having described the ideal electrical model and the ideal functioning, we can now focus on the side effects that affect efficiency and stability of the device. Firstly, losses can be analyzed and modeled; there are mainly four contributions:

- Losses due to *Joule effect*: the coil is made of a material with a finite conductivity. Therefore, when supply current flows through, it warms up and some power is lost as heat according to the well-known law $P = RI^2$. This effect can be modeled with a series resistor, because its entity depends on the value of the supply current I.
- Losses due to *dinghy* imperfection: although the usage of the *nonmagnetic material* helps in conveying as much field lines as possible on the mobile core, there will still be some field line which doesn't converge on the device axis. From our point of view these oblique lines represent a loss as they limit the efficiency intended as the capability of transferring all the energy generated by the magnet to the *mobile core*. This effect can be modeled with a series inductance which takes into account the magnetic flow lines that abandon the core and close their loop through external paths (air).
- Losses due to *magnetization currents (static phenomenon)*: the presence of the fixed core made of ferromagnetic material helps amplifying the strength of the magnetic field generated by the solenoid. However such an amplification doesn't come "for free" but requires a biasing. In magnetic circuits, this biasing is due to *magnetization currents* I_{MS}. The ferromagnetic core can be seen as a cylinder made of homogeneous material, whose atoms carry an elementary magnetic momentum $\mathbf{B_I}$. When the solenoid generates the magnetic field due to the supply current I, all these small elementary vectors tend to align in the direction of the magnetic field ($\mathbf{B} = \mu\mathbf{H}$ in isotropic materials), that is, in the axial direction. From the microscopic point of view, each atom acts like an elementary current loop generating the $\mathbf{B_I}$. However, all these internal currents are compensated by adjacent atoms, so that means the value of internal current is equal to zero. This is not true for the atoms lying on the external surface, which are not fully compensated by the neighbors, giving birth to a superficial current density: *the magnetization current* I_{MS}. Making it simple, an aliquot of the supply current I is lost as I_{MS}. This effect can therefore be modeled with a parallel resistor, whose value depends on the conductivity of the core, which subtracts current from the supply.

Figure 7. Magnetization current generation



The elementary current densities inside the magetized cylinder cancel each other mutually.
The unbalanced contributions are lying on the external surface.

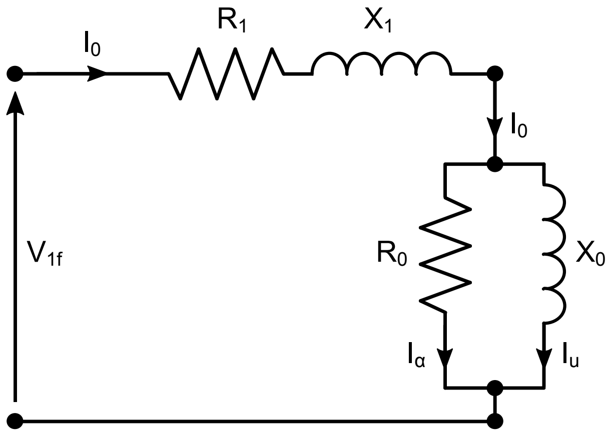
- Losses due to *parasitic currents (Foucault, dynamic phenomenon)*: in our application, the electrovalve is not used in a quasi-static context but it's driven by a PWM current signal that switches it on and off at a high frequency (up to 10 kHz) in order to "linearly" control the flow of the fluid transferred from the input to the output. In such a context, losses due to dynamic phenomena cannot be neglected. Faraday's law of induction (see) states that a variation of the magnetic flow Φ through the section of a material generates an induced electromotive force . Such a phenomenon is responsible for the generation of parasitic currents inside the ferromagnetic core, known as *Foucault currents (or Eddy currents)*, which contribute to the overall amount of losses.

Faraday's law of induction

$$\mathcal{E} = -\frac{d\Phi}{dt}$$

Putting all the effects together, the complete equivalent electrical model of the electrovalve would be the one shown in the Figure 8. I_0 represents the supply current; R_1 accounts for the losses by Joule effect; X_1represents the parasitic inductance accounting for the aliquot of magnetic flow which doesn't concatenate to the mobile core; R_0 accounts for the aliquot of supply current lost as magnetization current; X_0 represents the ideal load.

Figure 8. Real electrical model



5 Empirical electrical model used for MATLAB Simulink® simulations

In order to compare results obtained by simulations to the outcomes of the test bench, it is essential to work in the most similar condition as possible. Therefore, an empirical model (shown in the Figure 9) is usually used for the electrovalve in our Simulink simulations. Such a model is based on current regulated waveform measured on a real electrovalve at different current average values. For an optimal fitting, three contributions can be recognized: all of them can be represented as the evolution of a RL series circuit. Nonlinearities for the inductance values due to secondary effects like magnetostrain, coil whine and losses can be also taken into account by using different values for the inductances. Such a values, and features are strictly depending the specific electrovalve to be controlled; the Table 1 reports, in order to provide order of magnitude, an example of data extracted for a load used for braking application.

The *input stage* features a selector for choosing the supply connection mode (High Side / Low Side). Further details on the possible supply configurations and The *core* of the model consists of three RL series circuits in parallel. The Figure 10 shows how such a circuit can be modeled in Simulink and what are the two simple equations on which it is based. It is important to highlight the presence of the integrator *1/s*, whose output is used for closing the feedback loop and evaluating the voltage drop on the series resistor.

Figure 9. Empirical model used for simulations with MATLAB

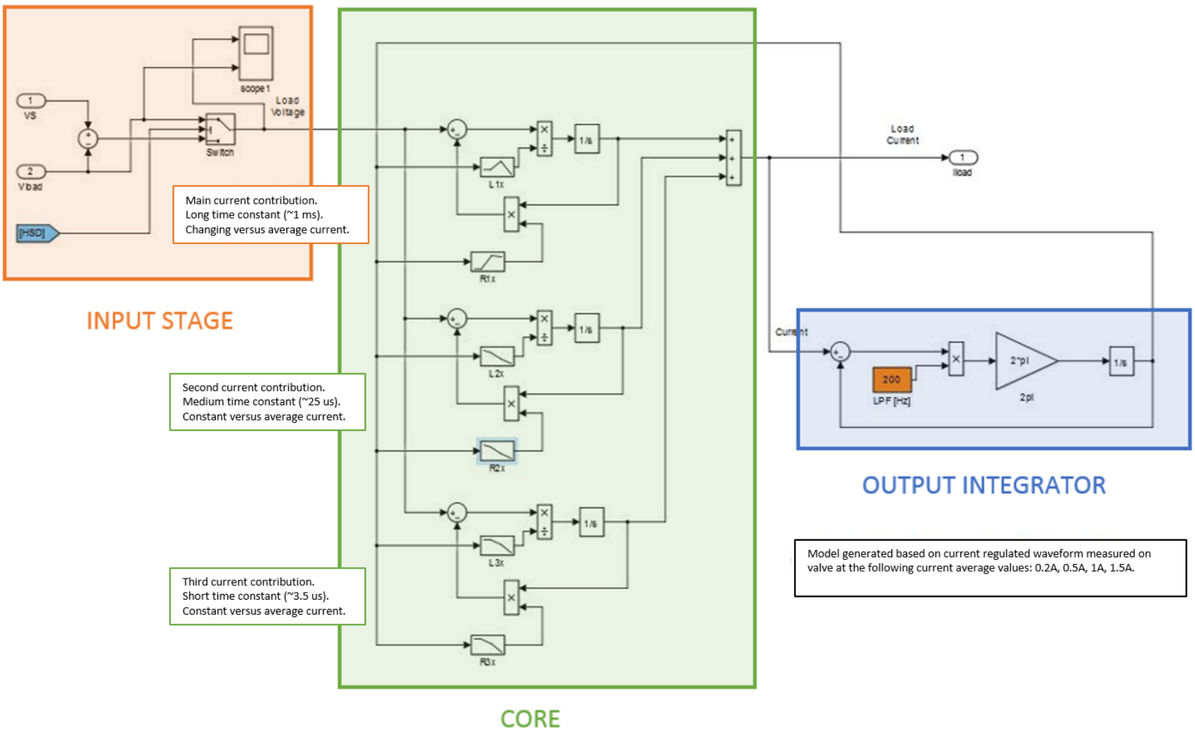
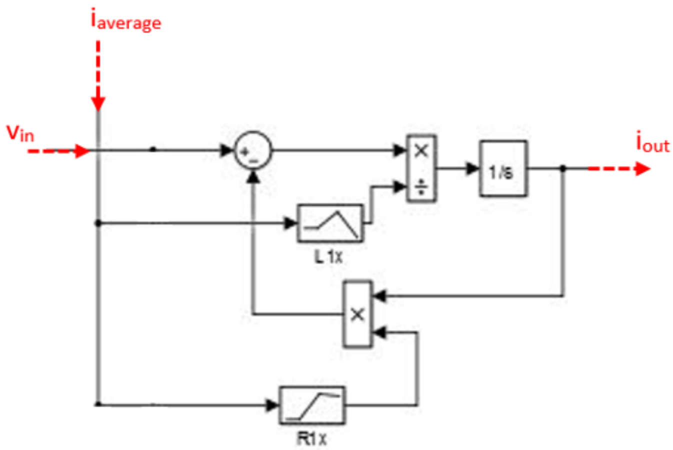


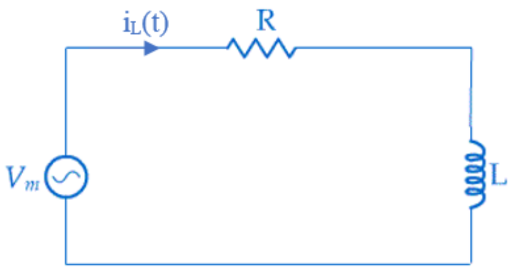
Table 1. Empirical model: nonlinear representation of resistors and inductances values

Component	Value				Unit
	@ Iavg = 0.2 A	@ Iavg = 0.5 A	@ Iavg = 1 A	@ Iavg = 1.5 A	
L_{1X}	4.54	4.54	5.07	4.28	mH
R_{1X}	4.54	4.54	5.4	5.35	Ω
L_{2X}	5	4.75	4	3.5	mH
R_{2X}	200	190	160	140	Ω
L_{3X}	0.98	0.98	0.91	0.77	mH
R_{3X}	280	280	260	220	Ω

Figure 10. Elementary Simulink block used for simulating the evolution of a RL series circuit



$$\begin{cases} i_L(t) = \frac{1}{L}\int_0^t v_L(t')dt' \\ v_L(t) = v_{in}(t) - Ri_L(t) \end{cases}$$

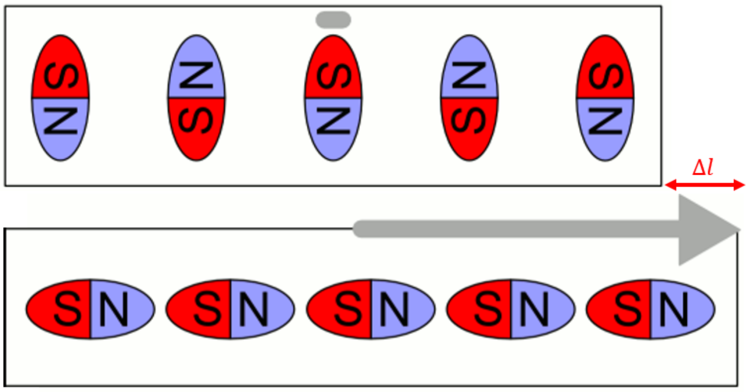


Moreover, looking at the model it can also be noticed the nonlinearity of L_{1x} and R_{1X}, whose value is determined according to the $i_{average}$ input. Such signal is generated by the *output integrator* block, that behaves as a *Low Pass Filter (LPF)*, cutting all the high frequency oscillations and extracting the average value of the output current. This value is then brought back in the feedback loop to determine the correct values for L_{iX} and R_{iX}, obtained by interpolating the samples based on a look up table similar to the example provided in the Table 1.

6 Acoustic noise sources: the magnetostrain effect

The magnetostrain effect is the physical phenomenon which consists of length variation of a material under the effect of a magnetic field. In our previous analysis, we have observed how the atoms of a ferromagnetic material modify their orientation when a magnetic field **H** is applied, in order to adapt the direction of the magnetic induction vector **B**, making it parallel to **H** (refer to the Figure 11).

Figure 11. Deformation of a material under the effect of the magnetic field applied



Summing all these microscopic contributions, a macroscopic deformation can be observed and measured according to the Eq. (4):

Definition of magnetostrain deformation

$$\lambda_\sigma = \frac{\Delta l}{l} \quad (4)$$

Where λ_σ is the magnetostrain deformation, is the length measured when no magnetic field is applied and Δl is the maximum length shift due to **H** (magnetic saturation). Both l and Δl are measured in the direction of magnetic application field **H** (in our case, the axial direction of the ferromagnetic core).

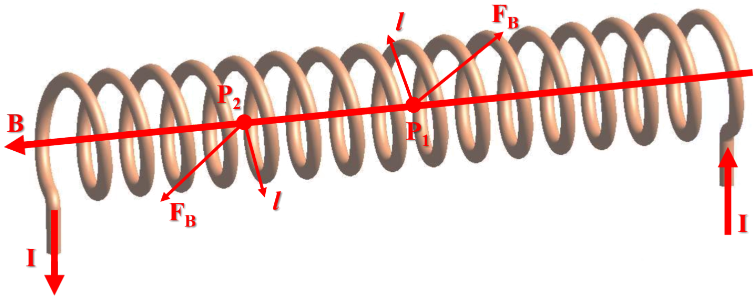
In ferromagnetic materials, λ_σ can be in the order of magnitude of 10^{-4} - 10^{-3}. Such a variation in the fixed core length produces, in case of variable magnetic field application, mechanical vibrations that are transferred to the coil, to which the fixed core is attached. From the electrical point of view, this phenomenon generates noise on the current that flows through the solenoid, and must be taken into account when designing the measuring interface of our ASIC. It is pretty intuitive understanding that this noise component will be mainly concentrated at the same frequency of the magnetic field applied to the core. Hence, in our application we expect to observe a noise component at the same frequency of the PWM signal used to toggle driving current signal: in fact, toggling current means toggling magnetic field then, toggling core length. Such a vibrations, generated by the needle, are usually responsible of the characteristics acoustic noise of the valve, which becomes more relevant the higher the current approaches the saturation current.

Although this effect can be usefully exploited in applications like ultrasound generators or MEMS switches, it represents a disturbance component in our field where having a “clean” measurement of the current flowing in the load is essential for achieving a good performance of the feedback control loop. A detailed description of the control algorithm and the hardware design of the CCCD will be provided, together with the explanation of the technical choices that have been made in order to reduce the impact of noise on performances.

7 Noise sources: the coil whine phenomenon

Coil whine or *coil noise* is a phenomenon that occurs in solenoids and represents, together with vibrations due to the core, an additive cause of the undesirable generation of audible sound, due always to vibration, induced not because of the needle, but because of the coil. It is not necessarily due to a malfunction but can occur also in healthy devices forced with a high amount of current. The Figure 12 helps explaining what the physical cause is.

Figure 12. Magnetic forces acting on the coil wires when the solenoid is supplied with a current



As previously analyzed, the solenoid shows axial magnetic field and magnetic induction vectors when supplied with a current of intensity *I*. In addition, a magnetic field can exert a force $\mathbf{F_B}$ on a wire supplied with a current according to the well-known law:

Magnetic force acting on a wire when a current flows through it

$$F_B = I\,\hat{l} \times B \quad (5)$$

where $\hat{\imath}$ is the versor pointing in the direction of the current flow in the examined point, *I* is the current intensity and **B** is the magnetic field. The Eq. (5) shows that the module of such a force depends on both the current and the magnetic field intensity, while its verse can be determined according to the right-hand rule.

Let us consider two points $\mathbf{P_1}$ and $\mathbf{P_2}$ on a generic section of the solenoid: as **B** is always axial and *I* is supposed constant in every point of the wire for the given working frequency, the only parameter that determines a variation in $\mathbf{F_B}$ is the direction of current flow, given by $\hat{\imath}$. As the latter is a versor, **FB** is only altered in direction but its module remains constant for all the possible $\mathbf{P_i}$. In particular, it can be noticed that in each point of the wire, the magnetic force vector points towards the direction of the versor normal to the external surface of the cylinder which encloses the solenoid. In other words, the magnetic force tends to “expand” the solenoid, causing its dilatation. This expansion is independent on the verse of the current because if *I* is inverted, **B** is inverted too, and therefore the verse of $\mathbf{F_B}$ doesn’t change.

In a dynamic context like our application, where the instantaneous current *I* varies in time, the module of $\mathbf{F_B}$ is not constant anymore but becomes a function of time. This intensity variation of the expansion force applied to the wire leads to an oscillation of the coil, which is expanding and collapsing at a frequency which depends on I(t). In many applications I(t) oscillates at frequencies for which the spectral components of the noise fall in the audible range and therefore we can hear the so-called *coil whine* or *coil noise*.

Revision history

Table 2. **Document revision history**

Date	Version	Changes
16-Apr-2021	1	Initial release.

Contents

List of tables

List of figures