SYMBOLS/UNITS

Z Impedance, measured in Ohms [Ω]
R Resistance, measured in Ohms [Ω]
L Inductance, measured in Henries [H]
f Frequency, measured in Hertz [Hz = s^{-1}]
ω Angular frequency, measured in Radian-Hertz [rad·Hz = rad/s]
B Magnetic field/magnetic flux density, measured in Tesla [T = Wb/m^2]
I Current, measured in Amperes [A]
μ Permeability, measured in Webers per Ampere-meter or Henries per meter [Wb/A·m = H/m]
ϕ Magnetic Flux, measured in Webers [Wb]
$\mathcal{F}$ Magnetomotive force (mmf), measured in Ampere-turns [A-t]
$\mathcal{R}$ Reluctance, measured in Ampere-turns per Weber [A-t/Wb]
F Force, measured in Newtons [N=kg·m/s^2]

INTRODUCTION

The solenoid is a commonly applied electrical component that can be seen throughout electrical and mechanical engineering. Between these two fields of engineering, they typically are used for different tasks but are unified in their fundamental functionality. This document is a general introduction to the theory behind solenoids, the complications which arrive from applying solenoids, and their two principle applications in mechanical and electrical engineering.

DISCUSSION

THEORY BASICS

A solenoid, on its own, is a fundamental circuit component comprised of a helical coil of wire through which an applied current may flow.[1] It is common for a solenoid to be comprised of many turns, that may be wrapped several times.[1] Since the solenoid is a wound coil of conductive wire, it exhibits inductive and resistive behavior in DC circuits.[2] However, in AC circuits it displays complex impedance as a function of the source's frequency.[2 3] The complex impedance is composed of a real impedance component, the resistance, and an imaginary impedance component, the reactance.[3] For an inductor, like the solenoid, the impedance can be denoted as a complex number, $Z_L = R_L + i\omega L$.[3] This behavior is only noticeable when dealing with an AC circuit, where $\omega = 2\pi f \neq 0$. However, solenoids are distinct from other inductors in the fact that they generate a 'near' uniform magnetic field when energized. The expressions for the magnetic field inside a solenoid and its inductance are:

$$B_{solenoid} = \frac{\mu N I}{l};$$
$$L_{solenoid} = \frac{\mu N^2 A}{l};$$
$$A = cross\ sectional\ area\ formed\ by\ each\ loop;$$
$$N = number\ of\ turns\ in\ solenoid;$$
$$l = length\ of\ the\ solenoid.$$

A visual representation of the resulting magnetic field produced, specifically, the magnetic field exhibited when the solenoid is left open to the air, can be seen in Figure 1(a). Where the linear arrangement of the coil turns produces a magnetic field comparable to that of a permanent bar magnet.[1] The symbolic representation of the solenoid typically depicts the helical coil wrapped around a magnetic core, seen in Figure 1(b), as their primary applications are derived from their use as electromagnets.[4]

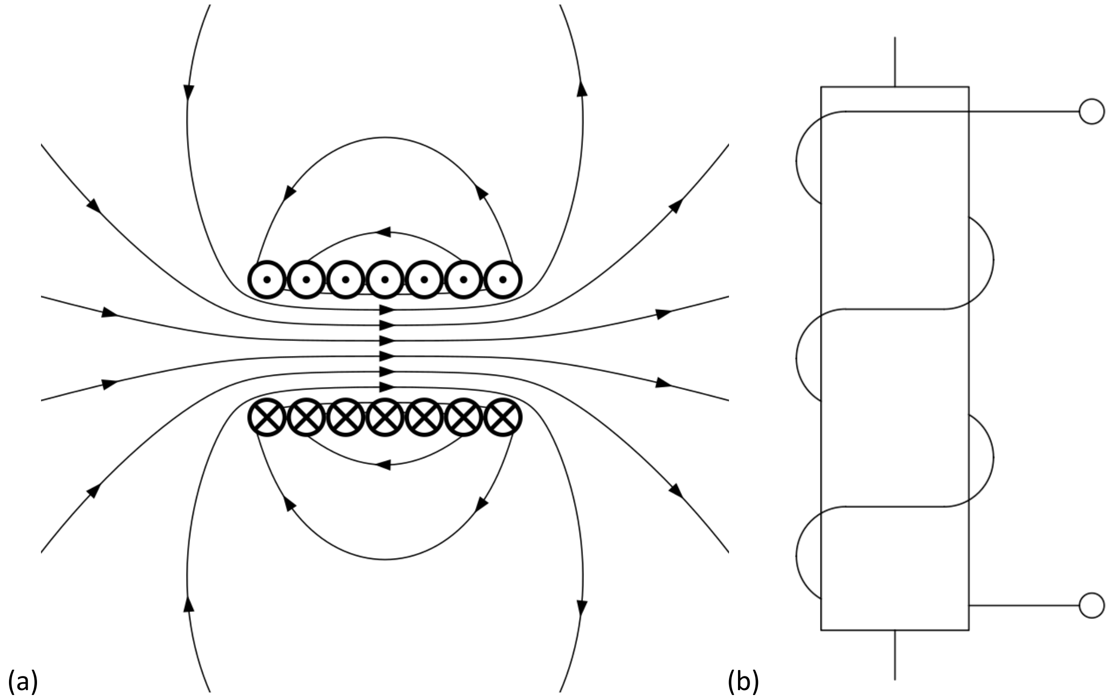

Figure 1(a) – Magnetic field inside a solenoid.[5]
(b) – Electromagnetic solenoid symbol.[5]

While solenoids can be used as inductors, they can produce this near-uniform magnetic field, which defines its application across lumped circuit analysis. Solenoids are typically undesirable for use as inductors, as an uncontained magnetic field can produce unwanted electromagnetic interference.[6] Amongst the lumped circuit applications for solenoids, it is their inductive quality that troubles any circuit meant to operate a solenoid-based electromechanical device.

PROTECTIONS

The applications for solenoids across engineering necessitate their ability to toggle between energized and de-energized states regularly, which adds some complexity as solenoids are inductors.[6] The voltage which develops across this inductor corresponds to the equation $V = L\frac{dI}{dt}$. Once the power supply is disconnected from the solenoid, an inductive kick occurs where the voltage across the solenoid dramatically increases in the reverse polarity to maintain continuity of current flow.[7] The voltage which develops is often referred to as the flyback voltage, which can damage the circuit controlling the operations of the solenoid as well as neighboring circuits due to impulsive interference.[6 7]

For a DC circuit, the inductive kick can be avoided with the inclusion of a back-biased diode (or flyback diode) in parallel with a solenoid.[6 7] An example of this protective diode configuration can be seen in Figure 2(a). When the power supply circuit is disconnected, the diode goes into conduction, allowing the gradual decay of current across the loop, protecting the control and neighboring circuits.[7] The diode that is used must be capable of handling the initial supply current.[7] The only drawback to using a flyback diode is that it extends the decay of the current traveling through an inductor, which can be problematic for high-speed solenoid devices, such as high-speed relays.[7] For these cases, a resistor or zener diode may be used instead to reduce the time of decay.[7] For an AC circuit, the inductive kick is prevented through the use of an RC “snubber” network, seen in Figure 2(b).[7]

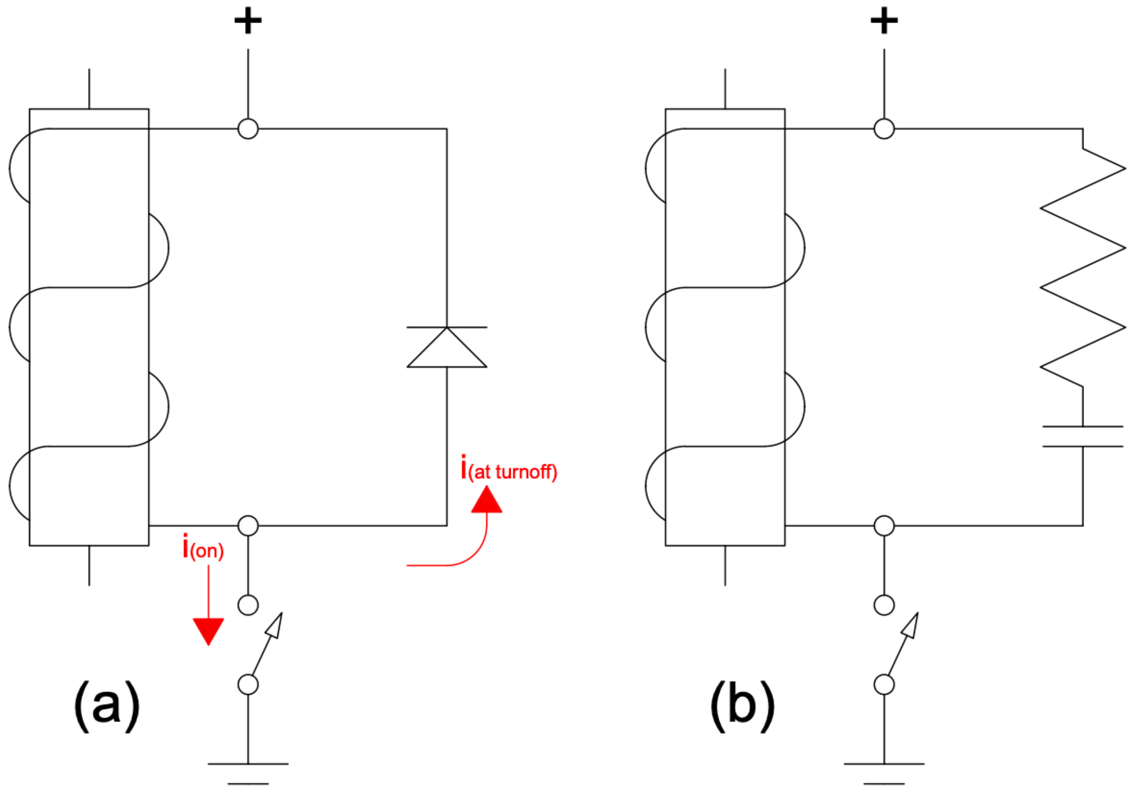

Figure 2(a) – Example of protective diode configuration for DC circuit.[5 7]
(b) – Example of protective RC "snubber" network for AC circuit.[5 7]

The necessity of these protections cannot be overstated, as they overcome one of the primary issues encountered when using an applied solenoid. Before we can jump to the specific applications of solenoids, we must also examine the theory behind their uses across engineering.

THEORY FOR ENGINEERING APPLICATION

The primary engineering applications for solenoids are derived from their uses in moving-iron devices, which are based on behaviors exhibited in magnetic circuits. A magnetic circuit, which can be considered as analogous to electric circuit, is comprised of an applied magnetomotive force (mmf) to a loop of magnetizable material, typically iron based.[8] This mmf is created by an energized N-turn coil, wrapped around parts of a ferromagnetic core that forms the magnetic circuit.[8] The magnetic field generated by the energized coil is channeled through the core, where the magnetic flux, $\phi = \int \vec{B} \cdot \overrightarrow{dA}$, becomes the electric analog to current.[8] The core's properties determine the reluctance of the magnetic circuit, which is comparable to resistance in lumped circuit analysis.[8] The analogy can best be shown by examining Table 1 and the following equations:

$$\mathcal{F} \triangleq NI;$$
$$\mathcal{R} \equiv \frac{\ell_\phi}{\mu A};$$
$$\mathcal{F}_\mathcal{R} = \mathcal{R}\phi \ (magnetic\ equivalent\ to\ Ohm's\ Law).$$

Magnetic quantity	***Symbol***	***Unit***	***Electric analog***
Flux	ϕ	Wb	Current
mmf	$\mathcal{F} = NI$	A-t	Voltage (emf)
Reluctance	$\mathcal{R}$	A-t/Wb	Resistance
Permeability	μ	Wb/A·m	Conductivity
Flux density	B	T	Current density
Magnetizing force	H	A-t/m	Field Strength

Table 1 – Magnetic and electric circuit analogues (t=turns).[8]

Magnetic fields are known to act on nearby magnetic materials, with the general trend of pulling the material into the densest part of the field. For a magnetic circuit, these attractive forces exist wherever an air gap is present and can be used to do mechanical work through the inclusion of a movable magnetizable material.[8] This force is called the gap force in magnetic circuits and can be expressed in a multitude of ways. In terms of magnetic flux and gap reluctance, $\mathcal{R}_g = \frac{\ell_g}{\mu_0 A_g}$, the gap force can be expressed as:

$$F_g = -\frac{1}{2}\phi^2 \frac{d\mathcal{R}_g}{dx}.$$

More commonly, the gap force is expressed in terms of the dimensions of the gap itself and the coil, which is driving the circuit as this is much more practical for design purposes. The mathematical expression for the gap force in terms of the solenoid physical qualities, currents, and the gap dimensions follow:

$$F_g = \frac{\mu_0 (NI)^2 A_g}{2\ell_g^2};$$
$$\ell_g = length\ of\ air\ gap\ in\ the\ magnetic\ circuit;$$
$$A_g = cross\ sectional\ area\ of\ air\ gap;$$
$$\mu_0 = permiability\ of\ free\ space.$$

It is from these equations that the gap force of magnetic circuits can be applied through the use of solenoids to do electromechanical work in mechanical and electrical applications.

APPLICATION IN MECHANICAL ENGINEERING

In engineering, the term solenoid (or solenoid valve) is commonly used to refer to the helical coil, the literal solenoid, when it is used as an electromagnetic actuator.[8] The principle design of this magnetic actuator is based around the construction of a magnetic circuit; a cross-sectional view of this circuit's most frequent incarnation can be seen in Figure 3.

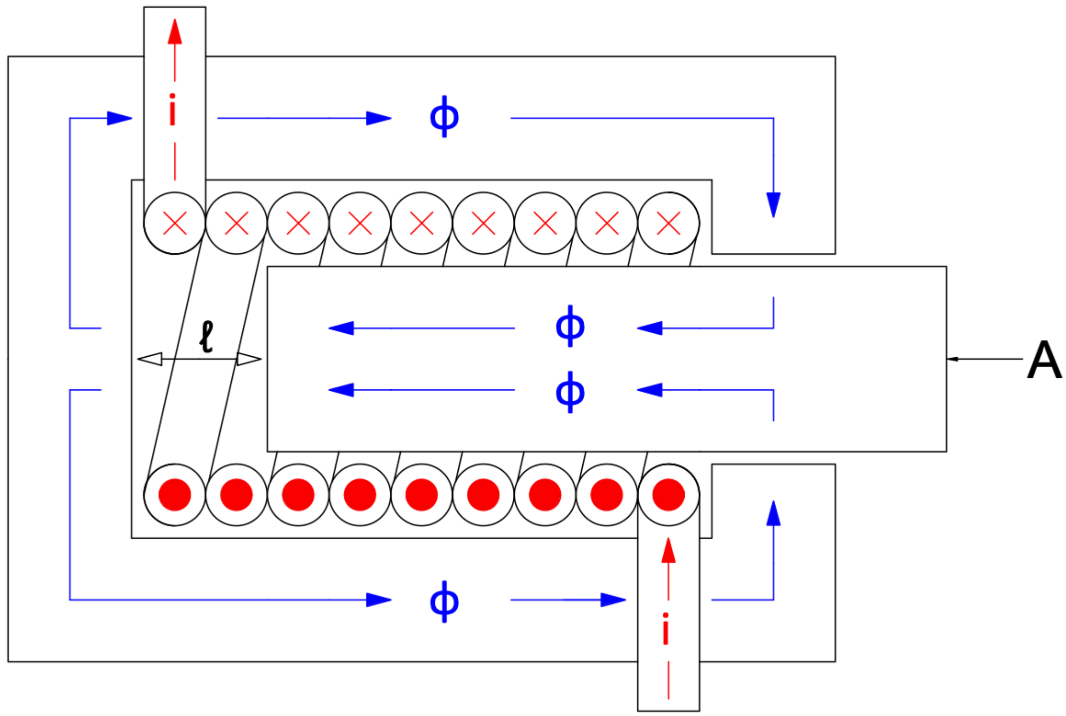

Figure 3 – Theoretical electromagnetic actuator (or solenoid) cross-section, with current and corresponding flux.[8 9]

The magnetic flux driven by the N-turn solenoid is channeled through the fixed ferromagnetic housing and the movable ferromagnetic shaft (or plunger), which is held inside the solenoid. An air gap, of length ℓ and cross-sectional area A, is left between the solenoids housing and the plunger to create a gap reluctance, which in turn creates the desired gap force. Typically, a spring is placed inside the air gap so as to return the plunger to its initial position before the circuit is activated by an incoming current, this creates a spring-loaded effect.[9] The plunger is generally aligned so that one end protrudes outside of the coil.[9]

When current is applied to the coil, a magnetic field develops parallel to the coil and the plunger. The ferromagnetic casing and plunger, contain this field, creating a magnetic flux, which is typically given the vector sign of the magnetic field vector. Since the field is contained in the ferromagnetic circuit, it can be approximated as constant. Meaning the flux at a particular point n can be expressed as $\phi_n = BA_n = \frac{\mu NI}{l} A_n$.

With the magnetic circuit completed, the plunger is connected to a device (or actuator head), allowing the conversion of electrical energy into mechanical work. Specifically, the magnetic actuator circuit configuration shown in Figure 3 is a basic configuration for a linear actuator, which is the most common configuration.[10] A key term to understand when designing or purchasing a linear solenoid, other than the voltage, current and power specifications, is the relative displacement of the plunger when the solenoid is energized. This displacement is commonly referred to as the stroke, as seen in Figure 4.[10] Along with the stroke of a solenoid, many vendors also list the force exerted by the solenoid.[11] However, the particulars of how a solenoid is marketed depend on its application.

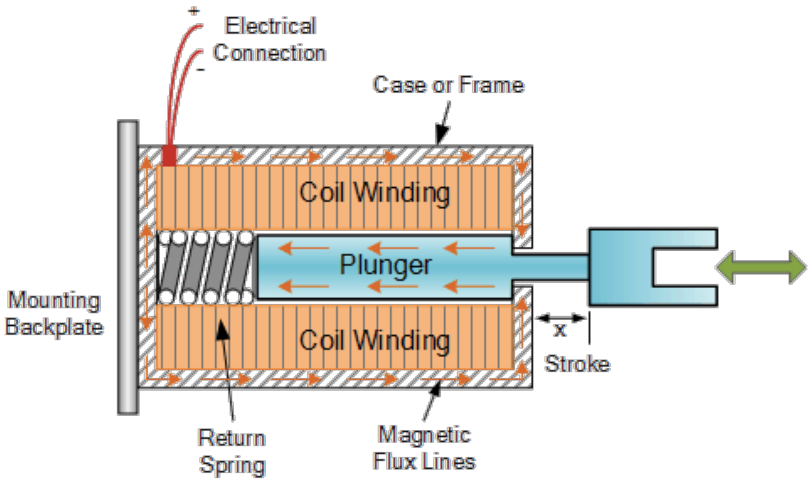

Figure 4 – Theoretical pull-type linear actuator diagram.[10]

There is a multitude of applications for linear actuators, which include their use in door-lock mechanisms, pressure and flow controls, and vehicle starter solenoids.[9]

APPLICATION IN ELECTRICAL ENGINEERING

A relay is, in its most basic sense, an electrically controlled switch designed so as to enable an electrical connection between two or more points of contact.[12] Electromechanically relays (EMR) are one of the most common types and operate using the same basic theory behind the operations of the linear actuator discussed earlier. The magnetic circuit is composed of a low reluctance, high permeability, iron core, which generally is comprised of two parts. The fixed portion of the iron core is called the yoke, and a movable spring-loaded portion called the armature.[12] The yoke and armature are connected at a pivot point, allowing the armature to undergo rotary motion.[12] An air gap is left between the armature and the opening of the solenoid. The solenoid is typically fixed to the yoke. An example of a possible relay can be seen in Figure 5.

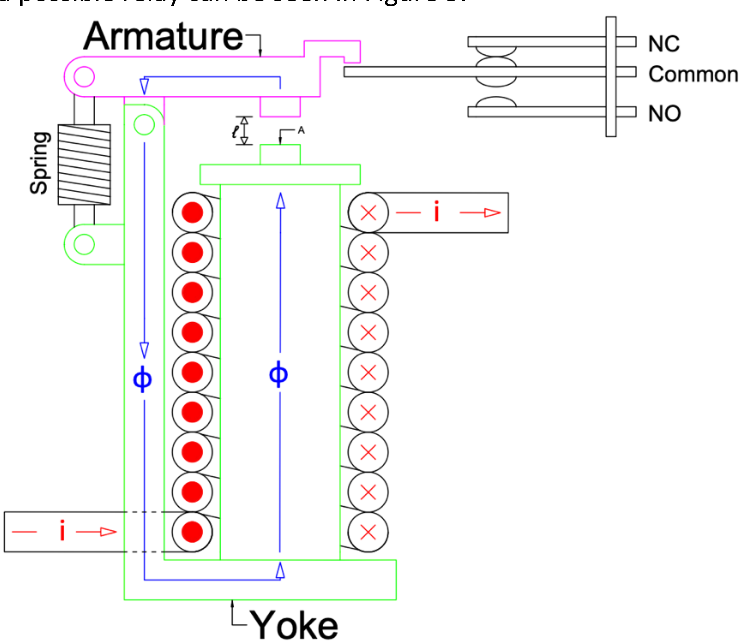

Figure 5 – Theoretical EMR schematic with currents and corresponding flux.[8 12]

The theoretical relay detailed in Figure 5, has a double-throw (DT) contact, also known as a change-over (CO) contact.[13] When the solenoid is off, the common and NC terminals are bridged; however, when the solenoid is energized, the armature applies a force to the common contact, disconnecting it from the NC terminal and bridging the common and the NO terminals.[12] Single-throw (ST) configurations are also available, where they control either an NC or NO switch.[13] There are also a single-pole (SP) and double-pole (DP), which refers to the number of switches a single relay possesses.[13] There are four primary arrangements, Single Pole Single Throw (SPST), Single Pole Double Throw (SPDT), Double Pole Single Throw (DPST), and Double Pole Double Throw (DPDT).[13] A schematic representation of the relay coil alongside these contact arrangements can be seen in Figure 6.

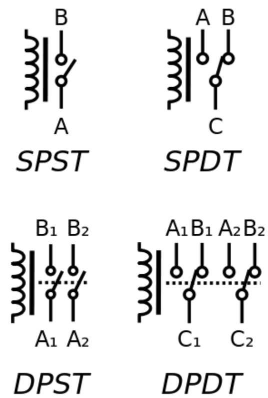

Figure 6 – Circuit schematic representation of a Relay.[14]

The resistance produced by bridging two relay contacts is a matter of great concern for electromechanical relays. Every electromechanical relay has some on-resistance (or contact resistance), which is expected to increase across the lifespan due to electrical arcing and sparking across contacts due to high inductive or capacitive loads across the relay terminals.[12] The major cause for contact failure is the occurrence of pitting on the contacts, where sparking results in the transfer and accumulation of material from one contact onto the center of another.[15] Special conductive materials can be used to reduce these effects and increase the lifetime of an electromechanical relay, each suited for a particular application on a case by case basis. A select number of possible contact materials, with their accompanying benefits, can be seen in Table 2. The maximum contact ratings for a relay can be found on the manufacturer's datasheets.[12]

Silver	High conductivity, however, is subject to sulphidation (reduces surface conductivity). Usually seen as the general-purpose material for relay contacts.[15]
Silver Nickel	High conductivity and increased resistance to pitting.[15]
Silver Cadmium Oxide	Less conductive but much more resistant to pitting. Can last much longer than silver contacts under the same conditions.[15]
Gold	High conductivity and does not oxidize. Generally used for low current switching due to the weak nature of the metal. Subject to low sulphidation. On-resistance does not increase during periods of disuse.[15]
Tungsten	Used in relays that have high voltage applications. Its' high melting point of 3380°C makes it highly resistant to erosion due to electrical arching.[15]

Table 2 – Table of relay contact materials and their advantages and disadvantages.

The lifespan of a typical electromechanical relay can be characterized in terms of electrical life expectancy as well as the mechanical life expectancy. Electrical life expectancy is characterized by the number of switching actions the contacts can undergo before the on-resistance exceeds the requirements of the circuit.[15] Mechanical life expectancy refers to the number of mechanical switching actions that can be completed in the lifespan of the relay and is typically more substantial than the electrical life expectancy.[15] The physical limitations of the electromechanical relay have led to the development of alternative approaches to electrically controlled switches. A prime example is the emergence of the Solid-State Relay (SSR). Generally, the SSR is much smaller, much quieter, and lacks the inductive behavior and the mechanical wear seen in electromechanical relays.[2 16 17] The SSR does not operate via an inductive solenoid nor a magnetic circuit. Instead, it is based on the use of semiconductors and electrical isolation between the control and load; which renders their design outside the prevue of this reference document.[16 17] The benefits of one type of relay over another is highly dependent on the application and the particulars of the circuit you want to create.

When selecting the appropriate relay for the circuit, you wish to create, it is worth noting some standard features to ask about when looking for a relay. These include coil and contact protections, the switching time, the voltage and current ratings, weather it is NC, NO, or closed contact, pricing, the life expectancy of the relay, and electrical isolation.[13] The major attractive elements for the electromechanical relays are their low-price points, the electrical isolation between the coil and contact circuits, and their high overload capacities.[16]

Relays, in general, are used in a multitude of applications, such as to realize logic functions, create time delay functions, as protective relays for faulty circuits, in the power grid, the automotive industry, and many more.[13]

CONCLUSION

Solenoids continue to be used throughout the field of engineering due to their simplistic construction, availability, and utility for key applications. The theory behind their operation is simple but profound. It is for this reason that the solenoid remains a key player in lumped circuit analysis, and it is likely to remain a central player.

REFERENCES

1. Knight, R.D., *Physics for Scientists and Engineers: A Strategic Approach with Modern Physics*, 4th Edition, p. 813-814, 2017.
2. Antonio R., “Privat Communication,” March 20-28, 2020.
3. Snoke, D.W., *Electronics: A Physical Approach*, p.67-68, 2015.
4. *Inductor Symbols -Solenoid, Chock and Coils Symbols*, https://www.electricaltechnology.org/2019/09/inductor-symbols.html.
5. https://commons.wikimedia.org/wiki/File:VFPt_Solenoid_correct3.svg, June 2010.
6. *Using Flyback Diodes in Relays Prevents Electrical Noise in Your Circuits*, https://resources.altium.com/p/why-you-should-use-a-flyback-diode-in-a-relay-to-prevent-electrical-noise-in-your-circuits, September 8, 2017.
7. Hill, W., and P. Horowitz, *The Art of Electronics*, 2nd Edition, p.52-53, 55, 1998.
8. Carlson, A.B., and D.G. Gisser, *Electrical Engineering: Concepts and Applications*, 1st Edition, p.631-631, 638-641, 655-657, 1981.
9. *THEORY OF SOLENOIDS, RELAYS, MAGNETOMOTIVE FORCE AND MAGNETIC FIELD STRENGTH*, http://www.tutorsglobe.com/homework-help/electrical-engineering/solenoids-and-relays-74071.aspx.
10. *Linear Solenoid Actuator Theory and Tutorial*, https://www.electronics-tutorials.ws/io/io_6.html.

11. https://www.moog.com/products/solenoids/solenoid-technologies/linear-solenoids/pull-solenoids/.
12. *Electrical Relay and Solid State Relays for Switching*, https://www.electronics-tutorials.ws/io/io_5.html.
13. John, *Working of Relays*, http://www.circuitstoday.com/working-of-relays, August 14, 2018.
14. https://en.wikipedia.org/wiki/File:Relay_symbols.svg, July 29, 2010.
15. *What is a Relay: Electromechanical or Electrical*, https://www.electronics-notes.com/articles/electronic_components/electrical-electronic-relay/what-is-a-relay-basics.php.
16. *Electromagnetic relays and Solid-State Relays (SSR), general technical descriptions, functions, Shutdown behavior, and design features*, https://releco.de/en/relays-general-technical-information.html.
17. Engineering Essentials: Relays and Contactors, https://www.machinedesign.com/mechanical-motion-systems/article/21834253/engineering-essentials-relays-and-contactors.