United
Electronic
Industries

The High-Performance Alternative™

Everything You Ever Wanted to Know about Data Acquisition

—

Part One — Analog Inputs

by Bob Judd
Director of Marketing
United Elecronic Industries, Inc

United Electronic Industries, Inc,
27 Renmar Ave, Walpole, MA 02081
(508) 921-4600

Web: www.ueidaq.com **Email:** info@ueidaq.com

Editor's Note

This document does not contain complete descriptions of any of the topics. The goal of this document is only to provide a general understanding of each topic. More detailed investigations can then be initiated in areas where more detail is required or desired.

Table of Contents

1.1 Preamble

This is the first of a three part series designed to introduce the key aspects of computer-based data acquisition and control to new users. It should also serve as a useful review or reference document for existing DAQ customers. The presentation is provided in three parts. Part 1 – Analog Inputs, covers the myriad of topics related to making measurements with a computer. Part 2 – Analog Outputs and Digital I/O, provides a discussion of analog output technology and topics related to digital I/O such as counter/timer. The third installment in the series is titled, Part 3 – Special DAQ functions, and covers such topics as communications interfaces (e.g., CAN-bus or ARINC-429) and special transducer interfaces (e.g., Synchro/ resolvers and quadrature encoders).

1.2 Introduction

People have been acquiring scientific data for thousands of years. From the ancient Greeks and Mayans up until very recent times, it's always been done the same way. A person looks at a scientific instrument and writes down what he/she observes. This continued on unchanged until the early 20th century when the paper-based chart recorder became available. Finally, data could be acquired and stored automatically.

Things stayed largely the same until the introduction of the digital computer, which provided the platform required to not only acquire and store data, but to also analyze and report it. There were a host of computer based data acquisition (and to a lesser degree, control) systems in the 1970s, but the industry really was born the day IBM released the PC. Though pitiful by today's standards, the original PC had a variety of features that made it an ideal data acquisition platform. The key features that helped the PC revolutionize the Data Acquisition industry included:

- It was inexpensive (relative to other computers like the HP-9825)
- It was easy to program (with a built-in basic interpreter)
- It had built-in, standardized I/O slots that would hold a DAQ board
- It was an almost overnight standard for computing.

The industry was born, and by the mid 1980s there were a wide variety of firms making data acquisition and control interfaces for the PC. "In the beginning", the PC-based data acquisition firms were basically divided into three categories: (1) plug-in DAQ[1] board vendors, (2) external box data acquisition vendors, and (3) software vendors.

The original PC-bus later became designated as the ISA-bus, and remarkably, there are still ISA bus applications being built today. It is a simple, robust, and inexpensive interface that has certainly stood the test of time. However, most of today's plug-in board business is based on the PCI bus (or variants such as cPCI and PXI) and is starting to follow the lead of the consumer PC vendors into PCI Express. The external box vendors now have Ethernet and USB standards to work with as well as some less used, but very viable, interfaces such as Firewire, CAN, and perhaps the oldest standard in computing, RS-232.

Software has progressed, too, from the original version DOS-based, interpreted Basic programs to extremely powerful applications such as MATLAB and LabVIEW, that are both easy to use and able to take advantage of today's powerful computers.

Today, most data acquisition companies (UEI included) provide board level, external box, and software. Depending on the application, board or box-level products may be more appropriate. This is the topic of a later chapter.

---

[1] At around this time, the industry was searching for an abbreviated way to say "Data Acquisition." The term "DAQ" was born and is now used interchangeably with data acquisition.

1.3 Data Acquisition, Data Logging & Control

Before continuing in the discussion of data acquisition, it may be useful to discuss what we mean by data acquisition and to better define how we will differentiate data acquisition from data logging and data recording.

We will use a very broad brush in our definition of data acquisition. Any computer system that either monitors or controls parameters in the outside world will meet our definition of data acquisition. The remainder of this chapter briefly discusses what we mean by data logger, data recorder, as well as the “& Control” technology that has become assumed, though not mentioned, when considering Data Acquisition/DAQ systems.

1.3.1 Data Logging

For purposes of this article, we will consider Data Logging as a special case within Data Acquisition. In common usage, a “data logger” is a self-contained data acquisition device that requires no connection or real-time interaction with a host PC in order to perform its function. Wouldn’t a lap-top PC with a PCMCIA DAQ card easily fit this description? How about a desktop PC with a number of PCI DAQ boards installed? Can’t a Programmable Automation Controller be configured to acquire and store data? If it acquires data and stores it in a digital format, we’ll call it data acquisition.

1.3.2 Data Recording

We will also consider Data Recorders as a special case with data loggers and, therefore, data acquisition. Historically, data recorders were data loggers designed specifically to capture higher speed data, often audio or vibration inputs. They also frequently would capture various communications signals such as serial or ARINC-429.

Figure 1. Modern Data Acquisition System Product

1.3.3 “— & Control”

Though not mentioned explicitly, data acquisition to most people involved in the industry has always meant data acquisition and control. There are a wide variety of vendors who consider themselves DAQ and though I have been in the industry for over 20 years, I am not aware of a single successful vendor that does (or did) not offer analog or digital output capability. Common usage of Data Acquisition / DAQ implicitly implies the “& Control” part of the system.

1.4 Board- vs. Box-based Systems

PC-based DAQ systems are available with a wide variety of interfaces. Ethernet, PCI, USB, PXI, PCI Express, Firewire, Compact Flash and even the venerable GPIB, RS-232/485, and ISA bus are all popular. Which one(s) is/are the most appropriate for a given application may be far from obvious.

Perhaps the first question to address when considering a new DAQ project is whether the application is best served by a plug-in board system (e.g., UEI’s PD2 series of PCI boards) or an external “box” based system (e.g., UEI’s PowerDNA Cubes or various USB devices available from many vendors). This issue has been a source of much confusion (and competition) over the years, and the decision may be less well defined today than ever.

In the early days of PC-based DAQ, the rule of thumb was: High speed measurements were performed by board solutions, high accuracy was the domain of the external box. Of course, there was a “gray” area in between that could be addressed by either form factor.

Today’s gray area is much larger than ever before. Board level solutions offering 24-bit resolution are now available as are 6.5 digit DMM boards. On the box side, USB 2.0 is theoretically capable of delivering 30 million 16-bit conversions per second and Gigabit Ethernet will handle more than twice that. Though internal plug-in slot data transfer rates have increased 10 fold in recent years, the typical data acquisition system sample rate has not. Planes and cars don’t go much faster now than in 1980 and temperatures and pressures are still relatively slow changing phenomena.

Since most application accuracy and sample rates are perfectly within the capabilities of both board and box level solutions, other considerations will determine which solution is best for a given application. Some of these key factors as well as why they are key are listed below.

1.5 Tradeoffs and Considerations

1.5.1 Distance from the PC to the Sensor or Measurement

This is a more important consideration than many people realize and it’s important for two reasons. First, running long wires from your test system and sensors can be a very expensive proposition, especially in large systems. Running a single communication cable, on the other hand, is inexpensive. Also, each foot of wire connecting your sensor or output to a remote host computer increases your susceptibility to noise. Quiet measurements of 18-bits or greater are almost impossible to obtain using long connection wires. Mounting the DAQ system close to the signal source, however, reduces this noise potential.

1.5.2 Portability

Some systems need to be portable. There are many small, external box devices that meet this need better than trying to drag a desktop or tower PC around. However, don’t overlook PXI when portability is a requirement. There are a variety of compact 4- and 6-slot chassis available.

1.5.3 Number of I/O Channels

Most people assume the external systems allow for more expandability and may be a better selection for a large system than a plug-in board system. That is often true and most of today’s desktop and tower PCs only include a few I/O slots. However, though considerably more expensive than a standard desktop PC, there are a large variety of server and industrial computer chassis providing as many as 16 I/O slots. PXI chassis with up to 18 slots are also available.

1.5.4 PC Obsolescence

External box systems certainly have the edge here. Even if your next PC is functionally identical to your existing computer, do you really want to remove all your I/O boards and install them in your new computer? Also, as technology changes, the slots inside computers change. If your current system has 4 PCI boards in it, are you sure your next PC will have homes for them? Of course, there is no guarantee your next computer will have the same external connections as your current PC, but the probability is almost certainly higher.

1.5.5 Preferred Host Computer

It's no secret that laptop computers are becoming ever more popular and their capabilities have expanded to the point where they're not just for road warriors any longer. Your options for developing a plug-in board-based DAQ system around your laptop are pretty limited. There are a variety of PCMCIA/PC-Card options available as well as a number of Compact Flash-based devices, but their capabilities and expandability are certainly limited. However, most new laptops come with Ethernet and USB ports, and many include Firewire as well.

1.5.6 Price

The "old" rule of thumb was that all else being equal, a plug-in board based system was likely to carry a smaller price tag. This is no longer the case with some of the lowest cost DAQ interfaces ever released offering USB or Ethernet interfaces.

1.5.7 Pure Speed

The internal buses will almost, by definition, be faster than those based upon an external communications link. After all, if the computer itself can't keep up with the speed of an external communications port, the extra speed is unlikely to be useful. However, only the highest speed applications are beyond the capability of USB, Ethernet, or Firewire.

1.6 Popular PC Interfaces

The remainder of the chapter discusses the popular computer interfaces used in today's DAQ products and briefly touches on the advantages and disadvantages of each.

1.6.1 Ethernet (100Base-T)

Originally released in 1980, Ethernet has become the standard network of PCs worldwide. Oddly, it is only fairly recently that we have seen general acceptance of Ethernet as a computer interface in DAQ/Measurement systems. Theories abound explaining Ethernet's slow migration into DAQ, perhaps the most common is that previously many engineers felt Ethernet systems were too difficult to configure and only trained IT personnel should dare. Of course, as the technology advanced, things got simpler, and today most teenagers are perfectly capable of installing a LAN in their houses and even most of us old-timers will have an easier time setting up a network than programming the VCR!

Standard Ethernet's 100 Mbps data transfer rate is fast enough for all but the fastest DAQ applications and its 100-meter range is also sufficient for the vast majority of systems. Ethernet systems can also be quite portable, since the only tie required to the host computer is a CAT5 cable.

Ethernet-based systems are easily expanded as ports may be added with extremely low cost, off the shelf routers. However, users should be careful to keep track of total system bandwidth requirements as all of the devices on a single Ethernet port share the bandwidth.

Ethernet ports are included on virtually all computers sold these days and most evidence points to this continuing for the foreseeable future. The IEEE has worked very hard to maintain backward compatibility among Ethernet revisions and so even as the Ethernet specification progresses, Ethernet equipment purchased today should be useful for many years to come. Ethernet communication is generally considered very secure and is therefore used by some of the largest manufacturing and office facilities.

Interoperability of Ethernet based DAQ devices from multiple vendors has not always been stellar. However, most Ethernet based DAQ (as opposed to Instrument) systems are single vendor and this has not been a major issue in the DAQ space. The LXI Consortium has developed a specification that ensures simple and seamless multi-vendor interoperability.

1.6.2 Gigabit Ethernet (1000Base-T)

As the name implies, Gigabit Ethernet is a version of Ethernet that supports 1 Gigabit per second data transfer rates. Other Ethernet specifications, such as deployment range and data types, remain unchanged. One thing to note is that to take advantage of the Gigabit bandwidth, your system needs to be developed accordingly. This means using either Cat5e or Cat6 cables, as well as a adding a Gigabit port to your computer and Gigabit routers/switches.

Gigabit is still a new technology, so most Ethernet-based DAQ products do not yet support the faster bandwidth (and in many, if not most, applications, the extra bandwidth is not required). However, many new computers' standard Ethernet interface/ports are now 1000Base-T capable and low cost, off-the-shelf Gigabit routers/switches are also available (e.g., I just saw a 5-port switch advertised for $29.99).

Ultimately, most networks of the future will probably be developed as Gigabit, but in most cases there may be little reason to retrofit existing networks or installations.

1.6.3 Fiber Ethernet (100Base-FX)

Boasting the same speed capability as standard Ethernet, the fiber optic implementation extends the range of the system to 2 kilometers (6,560 feet). Fiber interfaces are far from standard equipment on today's PCs, but 100Base-T to 100Base-FX converters are readily available as are PCI plug-in boards with 100Base-FX interfaces. For applications requiring even larger distances, single mode fiber links extend the useful range up to 20 km (12.4 miles). Single mode fiber systems, however, come with a fairly high price tag.

In addition to their ability to extend control beyond standard Ethernet distances, the Fiber interfaces have a number of other advantages. First, and probably foremost, is that they are almost immune to electrical and magnetic interference. If your application needs to communicate reliably in a noisy environment, therefore, fiber may be the way to go. Fiber also provides virtually absolute electrical isolation. If there's a good chance your DAQ system is going to take a big electrical hit and you want to make sure your host PC doesn't get fried, look to fiber. Finally, from a security point of view, fiber doesn't radiate any electrical or magnetic fields that can be "sniffed out" by uninvited guests.

1.6.4 Firewire (IEEE-1394a and b)

Initially developed by Apple Computer (with support from others), Firewire is a high speed serial interface. The Firewire specification is maintained by the IEEE and is known as IEEE-1394. The original spec, released in 1995, supported 400 Mbps transfers and is also known as Firewire 400. In 2002, IEEE-1394b was released and supports data transfer rates up to 800 Mbps (a.k.a. Firewire 800). The "b" version also extends the maximum distance between devices. Though the distance extends beyond the original 4.5 meters, the maximum data transfer rate is reduced.

The original target markets for Firewire were video and audio products. In these areas, Firewire has been very successful and has a significant market share. Firewire also has the basic requirements to make it an excellent backbone for data acquisition systems. However, at approximately the same time Firewire was being promoted, USB was coming on line. It appears that USB has “won” the battle for DAQ though the reasons are not intuitively obvious. At this time, there are a wide variety of DAQ vendors and products actively promoting USB devices, while with a few exceptions, Firewire success has been confined to the original target market of Audio and Video.

1.6.5 GPIB (IEEE-488)

Originally developed by HP and designated HPIB, the GPIB bus remains the dominant interconnection standard between computers and instruments though Ethernet and USB are beginning to make inroads. However, as prevalent as GPIB is in T&M applications, it has never had a substantial impact on the data acquisition market. There are a number of GPIB DAQ products available, but their market penetration is very small relative to PCI, PXI, Ethernet, and USB based products.

1.6.6 PCI (Peripheral Component Interconnect) Bus

The PCI bus is arguably the most common DAQ interface used today. Though Ethernet, USB, and PXI are all significant and growing, and PCI Express is “looking for a fight”, PCI is still the workhorse. It is very fast relative to almost all of the external box interconnection systems. PCI slots (of some sort) are also included in virtually all desktop and tower PCs.

PCI was originally developed by Intel as an interface to connect various functions on motherboards. It wasn’t long before it was generalized as a replacement for the aging 16-bit ISA bus that had dominated early PCs. With its “blazing” 33 MHz clock rate, a full 32-bit data path and Windows 95’s excellent support (including plug and play) it wasn’t long before the PCI bus had totally eclipsed the ISA bus in new “consumer” computers.

NOTE: Though you’d be unlikely to ever find a new computer from one of the major consumer suppliers with an ISA slot, the ISA bus market is surprisingly vibrant. ISA DAQ boards installed in industrial computers are still the backbone of many systems!

Though PCI has remained the industry standard since the middle 90’s, it has not remained stagnant. The PCI “standard” has moved from 33 MHz to 66 MHz. It also grew from a 32-bit bus to 64-bits. Also, as the industry moved from +5 VDC to +3.3 VDC logic, the specification was revised so as to support both. Throughout all of this, the spec has done a remarkably good job of maintaining backward compatibility. Cards designed in the mid 90s may still be used in many PCs purchased today.

The original PCI 33 MHz, 32-bit spec provided maximum transfer rates of 133 Megabytes per second. This was (and remains) fast enough for all but the highest speed data acquisition applications. Most DAQ boards today only take advantage of 32-bit transfers and support both 3.3 and 5 VDC interfaces. A new version of the PCI spec eliminates +5V support, but it is not yet known if this spec will become a common standard or will be eclipsed by other technology (e.g., PCI Express).

1.6.7 PCI Express

The latest of the computer interfaces to become common on standard computers, PCI Express is the first “all new” plug-in, general purpose computer bus to become popular since PCI. PCI Express slots are now found in most new desktop and tower PCs.

PCI Express abandoned the parallel data transfer architecture of PCI and PCI-X. Instead, PCI-Express is based upon multiple very high speed (~2 Gbps) serial paths. The serial nature of PCI Express becomes evident when you look at a PCI Express board and notice how small the board's PC interface is and how few "golden fingers" the boards have. Though 2 Gbps is quite fast, the PCI Express spec is not done there. PCI Express allows up to 16 of these serial links in each direction. The total possible data transfer rate of a full PCI Express implementation is 32 Gbps in each direction.

It is too early to determine the ultimate impact of PCI Express on the DAQ market. There are a variety of DAQ boards supporting PCI Express at this time, but only time will tell whether it, or some alternative, becomes the next de facto plug-in board standard.

1.6.8 PCI-X

PCI-X (sometimes confused with PCI Express) is a recent variant on the 64-bit PCI specification. The original PCI-X spec bumped the bus clock speed to 100 MHz and then 133 MHz. A new version of the specification bumps the clock rate to as high as 533 MHz. Even the most recent specification maintains backward compatibility with slower PCI boards, but of course the legacy boards cannot take advantage of the higher speeds. PCI-X has never become a significant factor in the DAQ market, though there may be a number of PCI-X DAQ boards available.

1.6.9 PXI

The PXI bus is electrically identical to PCI. PXI chassis, however, are developed specifically with measurement/DAQ applications in mind. All boards (including the CPU module) plug into the front of a PXI chassis. This allows much easier installation as no PC cover need be removed.

Also, the connectors of the boards plugged in are at the front of the chassis, which makes them much easier to get to in most applications.

The PXI specification covers much more than simply the computer interface and mechanical structure. The PXI backplane also offers a number of powerful triggering capabilities and mandates various "good neighbor" requirements so that boards from multiple vendors may all be easily integrated.

If there is a downside to the PXI market, it's that the CPU modules are specific to the PXI form factor. PXI CPUs don't take advantage of the huge economies of scale the consumer PC makers have, so a PXI CPU is likely to cost much more than a computer with equivalent horsepower from a company like Dell, Gateway, or HP. Of course, in many applications, the convenience of the PXI form factor more than makes up for the added cost. It should be noted that it is possible to control a PXI chassis from an independent host PC by installing a special "gateway" module in slot 0. Though this does allow an off the shelf, COTS computer to serve as a PXI controller, the Gateway interface systems are typically more expensive than the host computer and the price advantage of going to a COTS host are lost.

PXI has been very well received by the market and PXI products are available from a very large number of vendors. The specification is controlled by the PXI Systems Alliance. For more details about PXI, please see http://www.pxisa.org.

1.6.10 PXI Express

PXI Express is an implementation of PCI express. There are a variety of PXI Express products available, though as a very new specification, it is difficult to predict the ultimate acceptance of PXI Express products.

1.6.11 RS-232

People have been writing eulogies for the venerable RS-232 since I was a young engineer in the early 80s. However, the last survey result I saw indicated it was still the single most common interface between a PC and an external DAQ device. RS-232 is slow, fairly subject to noise and fairly short range, yet it remains ubiquitous. However, for the first time, new PCs have replaced the once common RS-232 port with USB ports and most external “consumer” devices have abandoned the RS-232. Could this finally be the end of RS-232? Time will tell.

1.6.12 RS-422/485

RS-422 and its networkable sibling, RS-485, has also been around a long time. Though slow by most standards, it is fast enough for many applications and has an excellent 1200-meter range. RS-485 has been especially well accepted by manufacturers (and buyers) of slow, low channel count, remote DAQ modules.

1.6.13 USB

USB has totally supplanted RS-232 as the communications interface of choice for most consumer items such as printers, cameras and the like. Though the initial releases (v1.0 and 1.1) were slow for a number of applications, version 2.0 has all the bandwidth most DAQ applications will ever require. Virtually every new computer includes multiple USB ports. It has also become very popular in the DAQ marketplace and there are a large number of vendors offering USB-based DAQ.

USB’s simple plug-and-play installation, combined with its 480 Mbps data transfer rate, make it an ideal interface for many data acquisition applications. Also, the popularity of USB in the consumer market has made USB components very inexpensive. New, low cost, USB DAQ devices are now available at previously unheard-of prices.

USB’s 5-meter range is perhaps its largest detraction as it limits the ability to implement remote and distributed I/O systems based on USB. There is also concern among some in the industrial arena that the USB connection’s lack of a locking cable mechanism might allow a USB cable to vibrate out of its connector. Whether this is a real possibility or not is certainly open to debate.

1.6.14 Wireless

An entire article could be written on the various options of wireless technologies and how they COULD apply to DAQ. However, the reality of the market as it stands right now is that the dedicated wireless-based DAQ market is still very small. There are a number of vendors who now offer wireless DAQ interfaces based upon such standards as Zigbee and 802.15.4 and variants of the 802.11 standard. There are also a growing number of customers implementing systems based on standard copper Ethernet devices, such as UEI’s PowerDNA Cube. Rather than using CAT-5 connections between the host PC and the DAQ device, however, they are opting to use standard off-the-shelf wireless routers from firms like Linksys and Netgear.

The next chapter starts the discussion of data acquisition I/O, and in particular, the all important analog input. Subsequent chapters will discuss other I/O types such as analog output, digital I/O, and counter/timer.

1.7 Analog Inputs

The analog input is the backbone of the data acquisition market. Though other I/O types play key roles in many applications, the vast majority of DAQ systems include analog input and a good percentage of them require only analog input.

We will define analog input by exclusion. Any input that is not digital, that is not defined as two states, (e.g., high/low or one/zero) will be considered analog. Common analog inputs include such measurements as temperature, pressure, flow, strain as well as the direct measurement of voltage and current.

Analog inputs are “measured” by a device called an A/D (Analog to Digital) converter (sometimes also referred to as an ADC). Though we’ll discuss A/D converter technology in the next section, it may be useful to mention here that *A/D* and *analog input* are often used interchangeably when referring to a DAQ product. In common usage, an *analog input board* and *A/D board* are the same thing. Similarly, an *analog input channel* and an *A/D channel* perform the same function.

1.8 A/D Converters

An A/D converter does exactly what its name implies. It is connected to an analog input signal, it measures the analog input and then provides the measurement in digital form suitable for use by a computer. The A/D converter is the heart of any analog input DAQ system as it is the device that actually performs the measurement of the signal.

1.8.1 Resolution

The resolution of an A/D input channel describes the number or range of different possible measurements the system is capable of providing. This specification is almost universally provided in term of “bits”, where the resolution is defined as: $2^{(\# \text{ of bits})} – 1$. For example, 8-bit resolution corresponds to a resolution of one part in $2^8 – 1$ or 255. For resolutions above 12-bit, the “- 1” term becomes virtually insignificant and it is dropped. A resolution of 16-bits corresponds to one part in 2^{16} or 65,536. The minimum difference in a measurement is one bit. This one bit is frequently referred to as the Least Significant Bit or LSB.

When combined with an input range, the resolution determines how small a change in the input is detectable. To determine the resolution in engineering units, simply divide the range of the input by the resolution. A16-bit input with a 0-10 Volt input range provides $10\text{ V} / 2^{16}$ or 152.6 microvolts. Table 1 provides a comparison of the resolutions for the most commonly used converters in DAQ systems.

Table 1. **Common ADC Converter Resolutions**

	8-bit	**12-bit**	**16-bit**	**18-bit**	**24-bit**
Distinct Levels	256	4,096	65,536	262,144	16,777,216
Resolution, ±10 V scale	78.4 mV	4.88 mV	305 µV	76.4 µV	1.192 µV
Resolution in °C, K type TC, ±0.5 Volt Full Scale input range (~25 °C)	97.7	6.10	0.38	0.10	0.00149
Dynamic Range in dB	48.2	72.2	96.3	108.4	144.5

Table 1 shows the resolutions of the more commonly used A/D converters in levels as well as in a number of common engineering units.

In general, the higher the A/D converter's resolution, the more accurate it will be. However, a DAQ device's overall measurement accuracy relies on much more than the accuracy of the A/D converter. A high resolution product is not always an accurate product. For example, many audio input products offer 24-bit resolution, but offer only 1% (about 7-bit) overall accuracy. This lack of accuracy is typically fine for an audio measurement, but could be unworkable in a strain measurement. Overall system accuracy concerns will be left to a subsequent section while this section returns its focus back to A/D converters.

1.8.2 A/D Converter Types

There are a variety of different types of A/D converters used in data acquisition. A detailed description of A/D converters could take an entire book and is beyond the scope of this note. However, a cursory knowledge of the different types of converters and, in particular, their relative strengths and weaknesses should prove beneficial. The most commonly used A/D converters in today's DAQ products are the: Successive Approximation, Delta Sigma (a.k.a. Sigma Delta), Flash, and Dual Slope/Integrating. The text below describes the various converter types while an overview of each type's key parameters is depicted in Table 1.

1.8.3 Successive Approximation (SA)

These converters remain the backbone of the DAQ industry. They typically provide resolutions in the 10 to 18-bit range, and depending on the resolution, offer sample rates up to tens of Megasamples per second. The basic underlying technology of SA involves comparing the input to the output of an on-chip D/A converter. Based upon whether the D/A converter output is higher or lower than the input, the D/A converter output is raised or lowered. In this manner, the SA converter ultimately zeroes in and when the D/A converter output is equal to the input (within one LSB), the iteration ends and the current digital word is written to the A/D output. The majority of UEI "Cube" I/O layers and all of UEI's PCI/PXI analog input boards use successive approximation converter technology.

1.8.4 Sigma Delta (also known as Delta Sigma)

Some manufacturers refer to this type of converter as delta sigma, while others call it a sigma delta. Sigma Delta appears to be the more popular designation at this time. Regardless of the "chicken or the egg" nature of what the converters are called, the sigma delta converter is rapidly becoming the standard by which other converters are judged and is seen in more and more DAQ devices each day. Perhaps the most differentiating feature of sigma delta converters is that they provide resolution up to 24-bits. Another interesting feature of converters is that they inherently trade off sample rate with resolution. Many converters provide an ability to sample at high speeds at low(er) resolutions or to slow down and sample at a rate that provides the maximum possible resolution.

Without going too much into the technology, the is conceptually fairly simple. A lower resolution converter inside the converter (typically 1-bit) is dramatically oversampled, i.e., sampled faster (in this case, orders of magnitude faster) than the sample rate your DAQ system desires. The large number of "one-bit" A/D conversions is then digitally integrated. The resolution is determined by the number of "internal" samples taken. A high number of internal samples provides a higher resolution with a lower overall converter sample rate. If you sample fewer times, you receive your A/D converter data faster, but at lower resolution. The converters are not without their share of design issues, but these are being addressed by the various manufacturers at a high rate and they will almost certainly supplant the Successive Approximation converter in the reasonably near future. A number of UEI analog input boards take advantage of converter technology, including the 25-channel, 24-bit resolution DNA-AI-225.

1.8.5 Flash

Flash converters can be extremely fast, though they do not typically offer high resolution. Sample rates in the gigahertz range are possible, but flash converters are rarely seen with resolutions greater than 8 bits. A flash converter is really nothing more than a string of comparators with highly trimmed input resistors. One side of each comparator input is set at one bit of resolution greater than the next. That means any given input to the A/D converter will always fall between the reference inputs of two adjacent comparators, driving one comparator output high, with the adjacent output low. This data is then taken into a decoder, and the position where the switch from 0 to 1 occurs defines the A/D output. As mentioned previously, it is a simple concept, but as a Flash converter requires at 2^n comparators, the amount of circuitry required on the chip gets huge as the converter resolution increases. An 8-bit flash converter requires 255 comparators. That’s a lot of analog circuitry on a single chip, even at today’s densities. Try to add one bit of resolution and you push the total number of comparators required to 511. A very difficult challenge indeed.

1.8.6 Dual Slope / Integrating

Once the king of the high resolution A/D converter market, dual slope converters are seldom seen in new designs. Like the flash converter, the longer you are willing to wait for your conversion, the more resolution is achievable. Unlike the flash type, the dual slope converter is typically set to sample at a single rate and the user is not allowed to “fiddle” with the sample rate versus resolution equation. Though seldom used in new designs, there are a variety of existing DAQ products based upon these converters that provide an excellent combination of high accuracy and noise immunity in applications where high sample rate is not critical.

1.9 Input Channel Configuration

Most multi-channel DAQ boards are based upon a single A/D converter. A type of switch called a multiplexer is then used between the input channels and the A/D converter. The multiplexer connects a particular input to the A/D, allowing it to sample that channel. The multiplexer then switches to the next channel in the sequence and another sample is taken. This goes on until all the desired channels have been sampled. This configuration provides inexpensive, yet accurate measurements. The primary disadvantage of this system is that even if the switching and sampling are very fast, the samples are actually taken at different times. **Figure 2** depicts a typical, multiplexed input configuration.

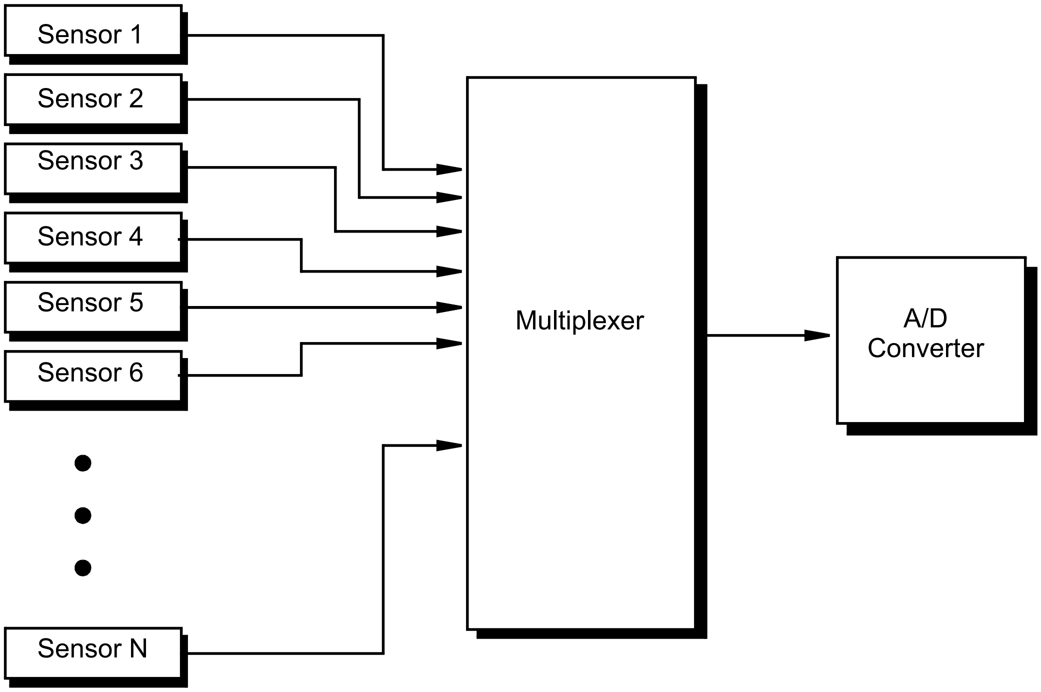


Figure 2. Typical Multiplexer/ADC DAQ System

Most multi-channel analog input devices use a multiplexer and a single A/D converter as shown in the block diagram above.

Though the multiplexed configuration minimizes cost, it does have a number of disadvantages. Perhaps the most important of these is that this configuration does not sample all channels at the same instant. The multiplexed systems sample a channel, and then switch the multiplexer and sample again. The time differential between samples is typically referred to as the sample “skew”.

In many, if not most applications, the time skew between samples on different channels is not problematic. However, in some applications, these skews (a.k.a. phase shifts) between signals cannot be tolerated and a standard multiplexed configuration cannot be used. The ability to sample inputs at the same instant in time is typically referred to as simultaneous sampling.

1.10 Simultaneous Sampling[1]

There are two common ways to achieve simultaneous sampling. The first is to simply place a separate A/D converter on each channel. They may all be triggered by the same signal and will thus sample the channels simultaneously.

---

[1] Simultaneous sampling is somewhat of a misnomer. Samples can never truly be simultaneous as there is always a certain skew between samples. However, this skew can be reduced to levels low enough that they are insignificant in the particular application. The error or skew between samples is commonly referred to as the aperture uncertainty and is typically measured in nanoseconds (ns), though some higher speed devices may offer sub-nS aperture uncertainty. As an example, the 4-channel, 250 kHz DNA-AI-205 offers a maximum aperture uncertainty of 30 nS.

The second is to place a device called a sample & hold (S&H) or track & hold (T&H) on each input. In “sample” mode, the device behaves like a simple unity gain amplifier. That is, whatever signal is provided on the input is also provided at the output. However, when commanded to “hold”, the S&H effectively freezes its output at that instant and maintains that output voltage until released back into sample mode. Once the inputs have been placed into hold mode, the multiplexed A/D system samples the desired channels. The signals it samples will all have been “held” at the same time and so the A/D readings will be of simultaneous samples. The second way to provide simultaneous sampling is to provide an independent A/D converter on each channel. Either system should provide good results.

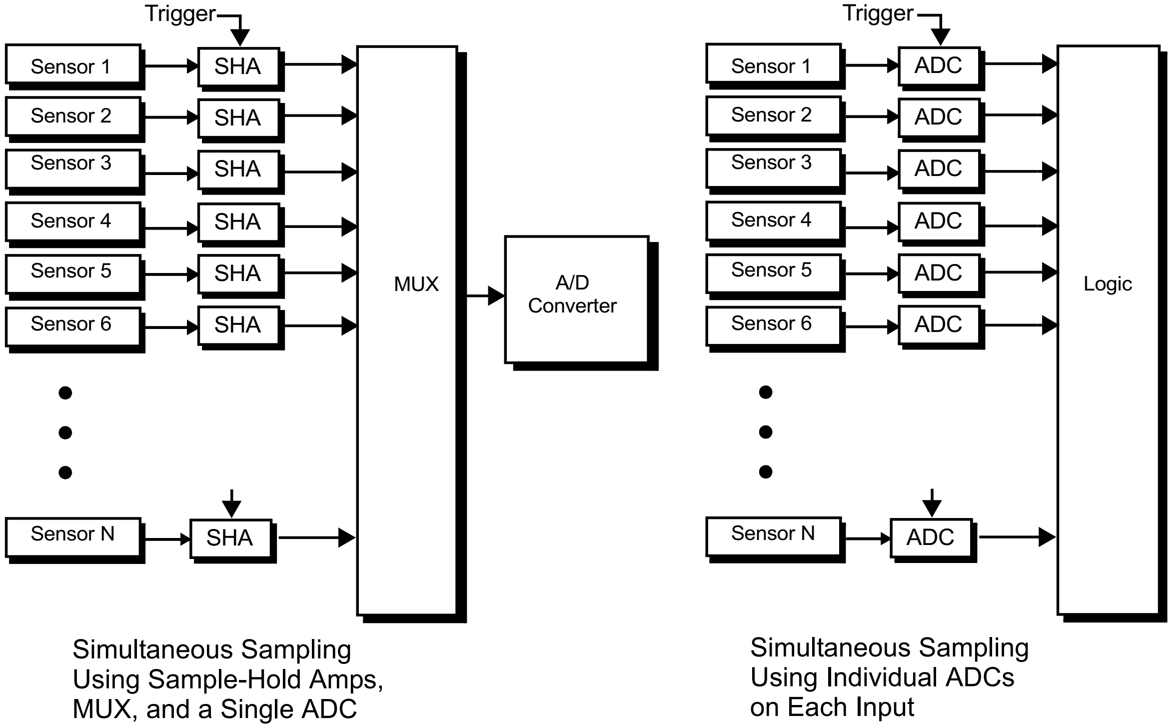


Figure 3. Simultaneous Sampling Techniques

Both of the simultaneous sampling configurations are shown in **Figure 3**. Depending on the application, an A/D per channel or the use of simultaneous S&H amplifiers may be a better design choice, though it will seldom matter to the actual DAQ users which configuration has been implemented. UEI’s DNA-AI-205 and DNA-AI-225 use a single A/D converter per channel, while the PCI and PXI-bus based PD2-MFS series boards utilize a single A/D with multiple S&H amplifiers.

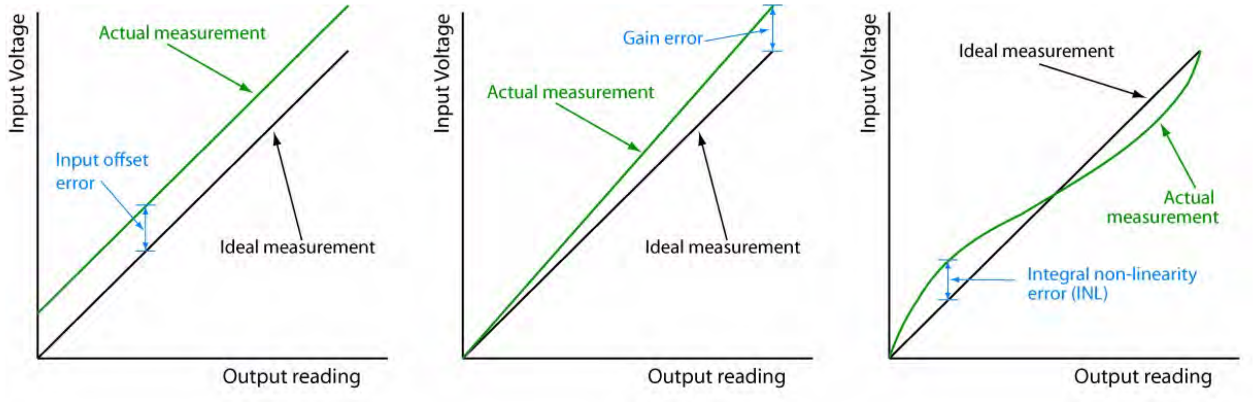
1.11 Accuracy Specifications

DAQ system accuracy is often equated to resolution, but they are not the same specification. Just because an A/D input can “resolve” a one microvolt signal does not mean the input is accurate to one microvolt. In fact, it is common that systems with sub-microvolt resolution provide overall accuracy on the order of millivolts. For example, a 24-bit audio input with an input range of ±2.0 V has 0.238 microvolt resolution. However, it might only provide overall DC accuracy of ±20 millivolts. The 20 mV accuracy is perfectly adequate for the audio application, but is not likely to be good enough for a temperature or pressure measurement. Of course, this is an extreme case. Though absolute accuracy is not always a critical issue, it is in many, if not most, applications. The key is to remember that high resolution does not always ensure high accuracy.

While resolution is almost uniformly specified in bits, accuracy specifications are offered in a wide assortment of ways. No one way is correct, and depending on the application, one specification method might provide more insight than another. An in-depth discussion of input accuracy specifications could be quite long, so we will only touch on a few of the key issues. Should you have any questions regarding input accuracy, I suggest you contact our applications group. It is not only the simplest way to get the information you require, it's a great way to evaluate our commitment to customer service and technical support.

A final note before beginning to describe the various components that lead to DAQ system inaccuracy is that it is important to be reasonable in specifying your input system. If your sensor provides an overall accuracy of ±1%, it's entirely unlikely you need a 24-bit A/D input. Over-specifying your system will simply add cost and complexity. The one caveat in this note is that you must keep your eyes open to the future. Is your system likely to change? Is your current DAQ application a short term project with the possibility of wanting to use the existing system in future applications? If so, you may want to consider specifying a system that is likely to offer the capabilities you anticipate in the future, and purchase appropriately.

The primary error contributors of an analog input system are: Input Offset, Gain Error, Non-Linearity and Inherent System Noise. There are additional error factors, but for most applications and in most systems, they will be small compared to the "big four" and can be ignored. **Figure 4** shows the relationship between a "perfect" input system and the effect each of the errors has on the measurement. We will also provide a brief textual description of the primary error components below.



Offset Error Gain Error Integral Non-Linearity Error

Figure 4. Graphical Descriptions Of Offset, Gain, and INL Errors

1.11.1 Input Offset

Assuming all other errors are zero, input offset is a constant difference between the measured input and the actual input voltage. For example, if the input offset voltage was +0.1 volt, measurements of perfect 1, 2 and 5-volt input signals would provide readings of 1.1, 2.1 and 5.1 volts, respectively. In a real system, the other errors are never zero, which complicates the measurement of input offset. Most analog input specifications define the input offset as the measurement error at 0 volts.

Note that some DAQ products, such as the DNA-AI-207, provide an "auto zeroing" capability. This function drives the input offset error to zero, or at least to a level low enough that its contribution is no longer significant relative to other errors.

The UEI DNA-AI-207 provides an auto-zero capability that reduces the input offset error to an insignificant level relative to the board's 18-bit resolution, resulting in the "ideal" measurement curve shown in **Figure 4**.

1.11.2 Gain Error

It is easiest to illustrate this error by first assuming all other errors are zero. Gain error is the difference in the slope (in volts per bit) between the actual system and an ideal system. For example, if the gain error is 1%, the gain error at 1 volt would be 10 millivolts (1 * .01), while the error at 10 volts would be ten times as large at 100 millivolts.

In a real world system where the other errors are not zero, the gain error is usually defined as the error of the measurement as a percentage of the full scale reading. For example, in our 0 – 10 volt example range, if the error at 10 V (or more often at a reading arbitrarily close to 10 volts such as 9.99 V) is 1 millivolt, the gain error specified would be 100*(.001 / 10) or .01%. For higher precision measurement systems, the gain error is often specified in parts per million (or ppm) rather than percent as it's a bit easier to read. To calculate the error in parts per million, simply multiply the input error divided by the input range by one million. In our example above, the 0.01% would be equivalent to 1,000,000 * .001 / 10 or 100 ppm.

Though many products offer auto-calibration, which substantially reduces the gain error, it is not possible to eliminate it completely. The automated gain calibration is almost always performed relative to an internally supplied reference voltage. The reference voltage will drift over time and any error in the reference will translate into a gain error. It is possible to create references with arbitrarily small errors. However, as the gain error gets small relative to other system errors, it becomes economically unfeasible to improve the reference accuracy. In addition to the cost penalty involved in providing the "pseudo perfect" reference, one of the errors, if not the largest, in most references is the drift with temperature. The only way to eliminate this drift is to maintain the reference temperature at a constant level. This is not only expensive, but it also requires a significant amount of power, which increases overall system power consumption.

1.11.3 Non-Linearity

As its name implies, non-linearity is the difference between the graph of the input measurement versus actual voltage and the straight line of an ideal measurement. The non-linearity error is composed of two components, integral non linearity (INL) and differential non linearity (DNL).

Of the two, integral non-linearity is typically the specification of importance in most DAQ systems. The INL specification is commonly provided in "bits" and describes the maximum error contribution due to the deviation of the voltage versus reading curve from a straight line. Though a somewhat difficult concept to describe textually, INL is easily described graphically and is depicted in **Figure 4**. Depending on the type of A/D converter used, the INL specification can range from less than 1 LSB to many, or even tens, of LSBs.

Differential non-linearity describes the "jitter" between the input voltage differential required for the A/D converter to increase (or decrease) by one bit. The output of an ideal A/D converter will increment (or decrement) one LSB each time the input voltage increases (or decreases) by an amount exactly equal to the system resolution. For example, in a 24-bit system with a 10-volt input range, the resolution per bit is 0.596 microvolt. Real A/D converters, however, are not ideal and the voltage change required to increase or decrease the digital output varies.

DNL is typically ±1 LSB or less. A DNL specification greater than ±1 LSB indicates it is possible for there to be "missing" codes. Though not as problematic as a non-monotonic D/A converter, A/D missing codes do compromise measurement accuracy.

1.11.4 Noise

Noise is an ever present error in all DAQ systems. Much of the noise in most systems is generated externally to the DAQ system and “picked-up” in the cabling and field wiring. However, every DAQ system has inherent noise as well. This noise is commonly measured by shorting the inputs at the board or device connector and acquiring a series of samples. An ideal system response would be a constant zero reading. In almost all systems, however, the reading will bounce around over a number of readings[1]. The magnitude of the “bounce” is the inherent noise. The noise specification can be provided in either bits or volts, and as peak-to-peak or Root Mean Square (RMS).

The key consideration with noise is to factor it into the overall error calculations. Note that a 16-bit input system with 3 bits RMS of noise is not going to provide much better than 13-bit accuracy. The three least significant bits will be dominated by noise and will contain very little useful information unless many samples are taken and the noise is averaged out.

1.11.5 Calculate the Total Error

To determine overall system error, simply add the offset, linearity, gain, and noise errors together. Though it can be argued that it is unlikely all three of the offset, linearity and gain errors will contribute in the same direction, it is certainly risky to assume they will not. It is seldom prudent to ignore Murphy’s law!

Max Error = Input Offset + Gain Error + Non-Linearity Error + Noise

A final note is that in most systems, Input Offset, Gain Error, and Non-Linearity all vary over time, and in particular, over temperature. If you require a very accurate measurement and your DAQ system will be subject to extreme temperature fluctuations, be sure to consider the errors caused by temperature change in your calculations.

1.12 Sample Rate

1.12.1 How fast is fast enough?

“How quickly must I sample my input signal?” is a fairly common question among DAQ system designers, and especially those without formal training in either DAQ systems or sample theory. The simple answer is the system must sample fast enough to “see” the required changes in input. In a purely input system, the minimum required sample rate is typically defined by Nyquist sampling theory. Nyquist found that to recreate a waveform, you need to sample at least twice as fast as the highest frequency component contained in the waveform. For example, if your input signal contains frequency components up to 1 kHz, you will want to sample at least at 2 kHz, and more realistically, at 2.5 - 3 kHz.

As with input resolution and accuracy, there is a tendency among DAQ system designers, particularly those new to the industry, to “over-specify” the system input sample rate. There are very few applications where it is necessary to sample a thermocouple more than 10 times a second, and most will probably be adequately served at a tenth that rate. Avoid the temptation to over-sample as it often increases system cost, memory requirements, and subsequent analysis costs without adding any useful information.

Note that the above pertains mostly to input-only systems. Control systems represent an entirely different set of considerations. Not only must the input sampling rate be high enough, but the CPU must have the “horsepower” to perform the calculations fast enough to keep the system stable and the output devices must have the speed and accuracy required to achieve the desired control results. A discussion of control theory is well beyond the scope of this note, but there we will add a few notes that may be helpful.

---

[1] *If the average of the readings is not zero, it indicates the system has an input offset error.

First, if you need any sort of deterministic control, and/or a hiccup in your control algorithm would be problematic, or your system update rate is more than 10 updates per second[1], you will likely need to consider using a real-time or “pseudo real-time” operating system. UEI offers support for QNX, RT Linux, RTAI Linux, RTX and XPC. Many users also find that though it is not a fully deterministic real-time OS, Linux-based applications have low enough latencies to be used in some higher speed control applications.

1.13 *DAQ “System” Considerations

Be careful to examine the analog input systems you are considering to determine if the sample rate specification provided is for each channel or for the entire board. As discussed previously, most DAQ input boards use a multi-channel multiplexer connected to a single A/D converter. Most “product” descriptions (e.g., 100 kilosample/second, 8-channel, A/D board), specify the total sample rate of the board or device. This allows sampling of one channel at 100 kS/s, but if more channels are required, the 100 kS/s is shared among all channels. For example, if two channels are sampled, each may only be sampled at 50 kS/s each. Similarly, 5 channels could be sampled at 20 kS/s each. If the specification does not specify the sample rate as “per-channel”, it is likely the sample rate must be divided among all channels sampled.

Another sample rate factor should be considered when various input signals contain widely varying frequency content. For example, an automotive test system may need to monitor vibration at 20 kS/s and temperature at 1 S/s. If the analog input only samples at a single rate, the system will be forced to sample temperature at 20 kS/s and will waste a great deal of memory/disk space with the 19,999 temperature S/s that aren’t needed. Some systems, including all of UEI’s “Cube” based products, allow inputs to be sampled at different rates, while products from many vendors do not.

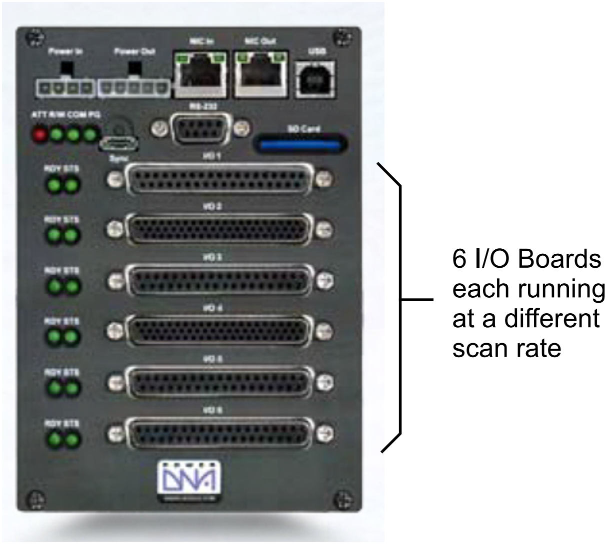


Figure 5. UEI Cube with 6 I/O Boards

---

[1] To ensure 10 updates/second in most Windows/Vista environments, you would also need to disable many automated functions such as Windows Update or Automated backups.

The final sampling rate concern is the need to sample fast enough, or provide filtering to prevent aliasing. If signals included in the input signal contain frequencies higher than the sample rate, there is the risk of aliasing errors. Without going into the mathematics of aliasing, we will just say that these higher frequency signals can and will manifest themselves as a low frequency error. **Figure 6** provides a graphical representation of the aliasing phenomenon. A real life example of aliasing is common in movies. The blades of a helicopter/airplane or the spokes of a wheel appearing to be moving slowly and/or backwards is an example of aliasing. In the movies it doesn't matter, but if the same phenomenon appears in the measured input signal, it's a pure and sometimes critical error.

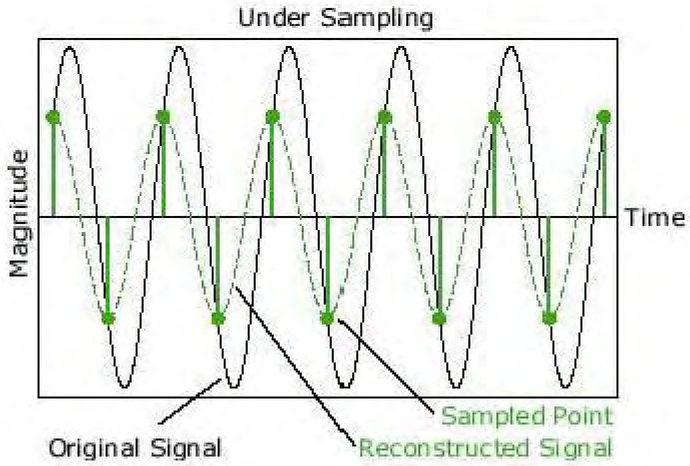


Figure 6. Aliasing

As shown in **Figure 6**, aliasing is possible when the sampling rate is too slow. The problem may be corrected by sampling faster, or by inserting an anti-aliasing low-pass filter, which removes any undesirable higher frequency signals.

There are really two solutions for aliasing. The first, and often simplest, is to sample at a rate higher than the highest frequency component in the signal measured. Some measurement purists will say that you can never be sure what the highest frequency in a signal will be, but in reality many, if not most, systems designers have very good *a priori* knowledge of the frequencies included in a given input signal. People do not use anti-aliasing filters on thermocouples because they are almost never needed. With a good idea of the basics of the signals measured, it is usually a straightforward decision to determine if aliasing might or might not be a problem.

In some applications, such as audio and vibration analysis, aliasing is a very real concern and it is difficult to guarantee that a sample rate is faster than every frequency component in the waveform. These applications require an anti-aliasing filter. These filters are typically 4-pole or greater filters set at one half the sample rate. They prevent the higher frequency signals from getting to the system A/D converter, where they can create aliasing errors.

1.14 Input Range

Though a fairly obvious specification, Input Range is not one to ignore. If you have a ±15 volt signal, you're not going to be happy with the results from a DAQ board with a fixed ±5 volt input range. Many systems these days have software programmable input ranges, though some still provide fixed input ranges. Either way, the signal measured should be smaller than the maximum input range of the DAQ input device.

Other than the obvious matching of the input range to your signal issue, one other input range consideration is whether the DAQ device is capable of setting different channels at different ranges. A single input range system measuring diverse signals (e.g. a thermocouple and a ±10 volt input) will require the system be set at a gain low enough to cover the largest input's full scale range. This in effect reduces the resolution available on signals with smaller full scale input ranges.

1.15 Differential and Single-ended Inputs

A differential input measures the voltage difference between two input terminals, while a single ended input measures the difference between and input and the DAQ device's ground. The two different configurations are shown in **Figure 7**. Many data acquisition products have the ability to be configured as either differential or single-ended. Selecting the single-ended mode doubles the number of channels available, but sacrifices the performance advantages offered by the differential input configuration.

1.15.1 Differential Mode Advantages

Differential inputs offer better noise immunity than single-ended for two reasons. First, much of the noise in a DAQ system is picked up when electromagnetic waves (usually referred to as EMI) in the local environment are coupled into system cables. Keeping the two wires in a differential input in close proximity means both wires are subject to the same EMI. Since the EMI pickup is identical on both wires, it does not show up as a difference in voltage at the two inputs. Since the differential input only measures the voltage difference between the inputs, the EMI is ignored.

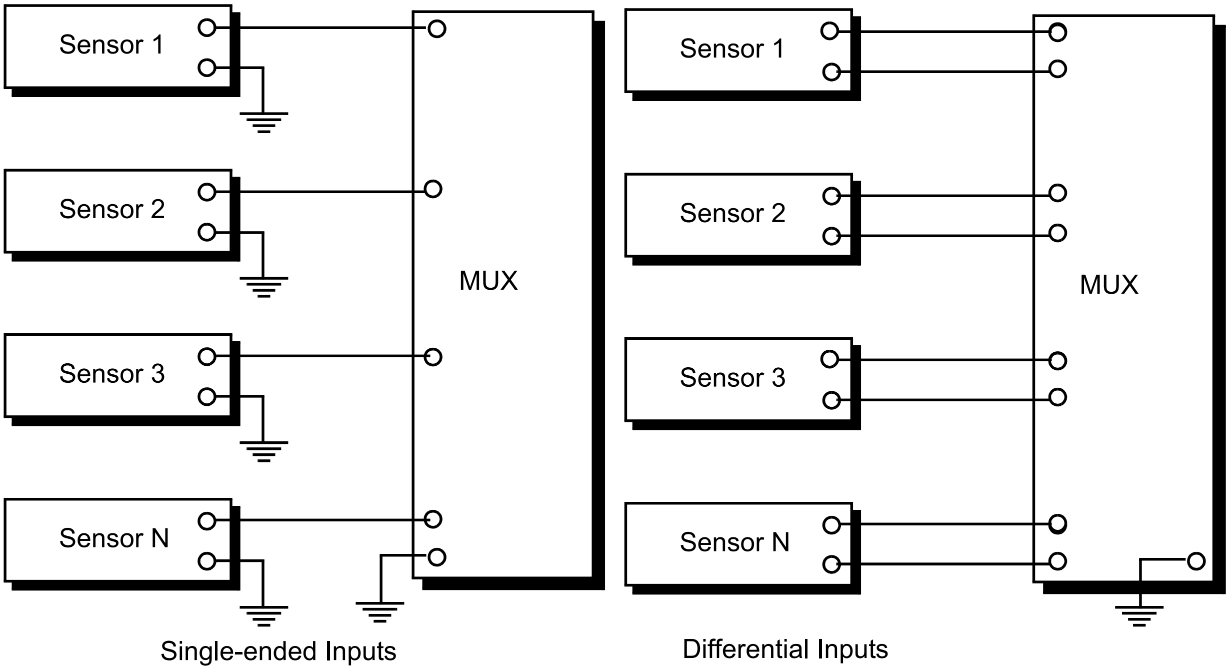


Figure 7. Single-ended and Differential Analog Inputs

Figure 7 depicts standard single ended and differential input modes.

The noisier the environment, the more important it is to have differential inputs. Though not a firm rule, at 16-bit resolution or higher, it is usually recommended that you use differential inputs. Most vendors do not offer interfaces with higher than 16-bit resolution in single ended mode as the noise picked up by the high resolution, single ended input is almost certain to disappoint the customer.

Differential signals also have the advantage that within a limited range (referred to as the Common Mode Range, see below) they float relative to the DAQ device's ground. Many, if not most signals connected to a typical analog input come from other ground referenced devices. However, the ground of the signal source is almost never at exactly the same voltage as the ground of the DAQ inputs. A difference in the "ground voltage" at the DAQ system and at the sensor or signal source is largely ignored by differential inputs, but creates a tug of war that can "move" the DAQ system ground around enough to cause significant measurement error.

1.15.2 Common Mode and CMRR

The difference between the "average voltage" of the two differential inputs and the input ground is referred to as the signal's Common Mode. Mathematically, the Common Mode voltage is defined as:

$$Vcm = \frac{1}{2}(V_{hi} + V_{low})$$

Where Vhi is the voltage of the signal connected to the V+ (or VHi) terminal and Vlow is the voltage on the V- (or Vlow) terminal.

The range of input signals where the input is able to ignore or "reject" the Common Mode Voltage is called the Common Mode Range. Common mode range is typically specified in volts (e.g. ±10 V). If both inputs remain within this range, the differential input will work properly. However, if either input extends beyond the range, the differential input amplifier will saturate and create a substantial and often unpredictable error.

To keep your signals within the common mode range, you must ensure that V+ added to Vcm is less than the upper limit of the common mode range and V- subtracted from Vcm is greater than the lower limit of the common mode range.

The ability of a differential input to ignore or reject this Common Mode voltage and only measure the voltage between the two inputs is referred to as the input's Common Mode Rejection Ratio (or CMRR). Graphical depictions of the Common Mode concepts are provided in **Figure 8**.

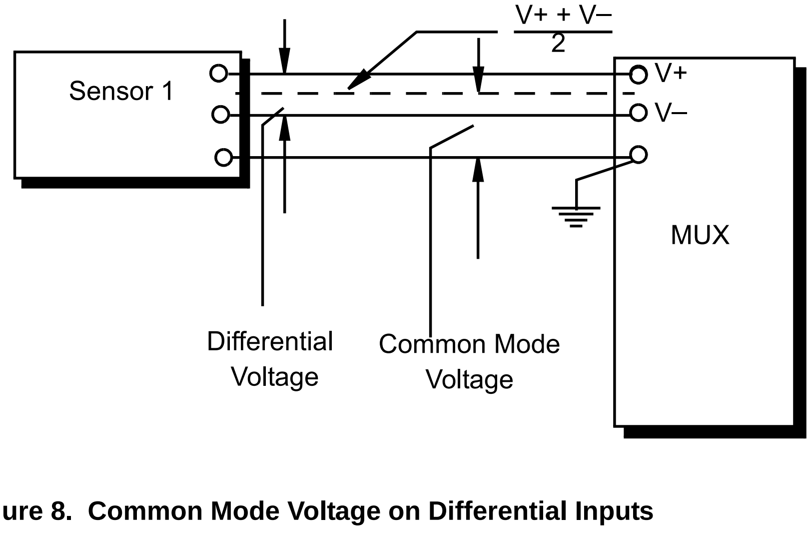


Figure 8. Common Mode Voltage on Differential Inputs

Figure 8 depicts a differential input and points out the key aspects of common mode voltage and the common mode range.

The Common Mode Rejection Ratio of modern input amplifiers is often 120 dB or greater. The standard formula for calculated CMRR in dB is:

$$\text{CMRR} = 20\ \text{Log}\ (\text{Vin} / \text{Vcmrr})$$

With a bit of algebra, we can solve for the error term Vcmrr:

$$\frac{V_{cmrr}}{V_{in}} = \frac{1}{10^{(cmrr)/20}}$$

In our example, with a CMRR of 120 dB, the ratio is one part in one million. For each volt of Common Mode on the input, there is a Common Mode Error of 1 Microvolt. As you can see, common mode can be ignored in all but the most sensitive applications.

1.15.3 Connecting to a Differential Input

The noise reducing and common mode rejecting capabilities of a differential input are not entirely free. To properly take advantage of a differential input, you must be careful to connect it properly to your signals. In particular, you must pay attention to whether your signal source is isolated from the analog input (at the signal source, at the A/D input or both) or whether the two systems share a common ground reference. (Note that when we say common ground reference, we do not mean they are at the same ground, only that a low impedance connection path exists between the two grounds.)

There are two common cases which we will consider separately. These are:

- Connecting to a signal that is isolated from the DAQ input
- Connecting to a signal that shares a common ground reference with the DAQ input

1.15.4 Isolated Inputs

Isolated signals are straightforward to connect to a DAQ analog input, but do require one additional connection to perform reliably. In addition to connecting the + and – terminals to the high and low side of the DAQ input, you also need to provide a connection between one of the inputs (typically the – input) and DAQ input ground. Without this connection to ground, there is no current path that keeps the inputs from floating outside of the valid Common Mode Range (see previous section). A simple 10k, 20k, or higher value resistor will solidly lock the inputs within the Common Mode Range without compromising the differential nature of the input. The DNA-STP-AI-U screw terminal panel provides a jumper that will automatically make this connection for users, without requiring an external resistor. Most other vendors offer a similar capability. A diagram of the proper connection of an isolated two-wire sensor to a differential input is shown below.

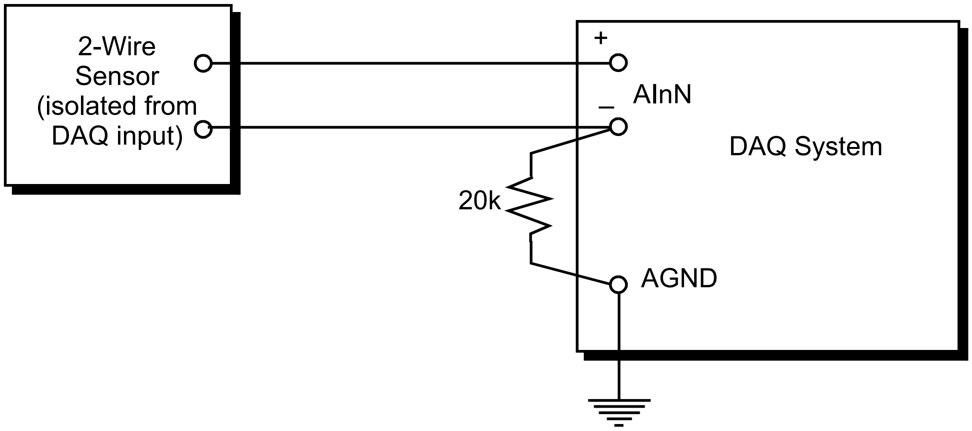


Figure 9. Differential Analog Input from a Two-wire Sensor

Figure 9 shows a preferred connection of an isolated two-wire sensor to a differential analog input.

NOTE: (If your thermocouple is not isolated from your DAQ input, you should not require this resistor.)

1.15.4.1 Input Signals with a Common Ground

Connecting a signal that shares a common ground with the DAQ input is a straightforward process, but there are two possible connections schemes. One is typically “preferred” while the alternate connection scheme may certainly be tried if the preferred configuration does not provide the performance desired. The preferred connection scheme is shown in **Figure 9**. The alternate connection is shown in **Figure 10**.

.

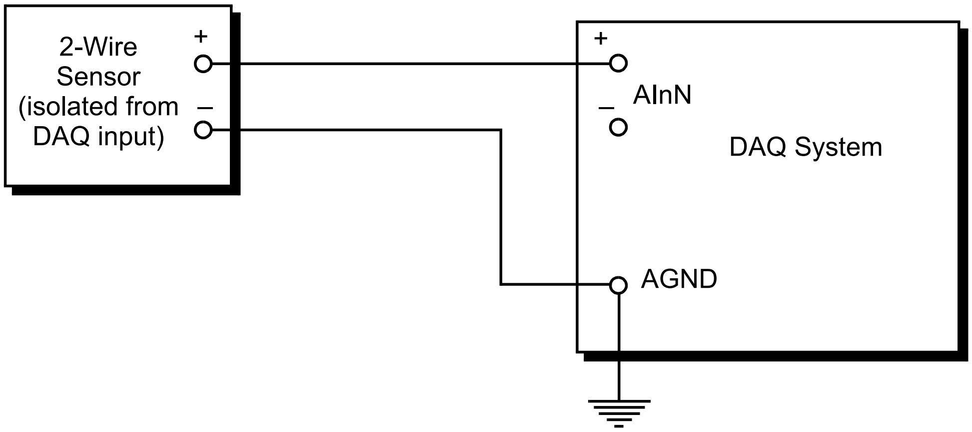


Figure 10. Alternate Connection for a 2-Wire Isolated Sensor

In both configurations, the signal source + terminal is connected to the Vin+ terminal of the analog input. The difference between the configurations is how the Vin- terminal is connected. In the preferred method, the ground connection from the signal source is connected to the Vin- terminal. Though this configuration typically leads to the best performance, in some systems it is advantageous to not connect the signal source ground to the Vin- terminal and instead connect the Vin- terminal directly to an analog ground on the DAQ input itself. **Figure 8** shows the alternate method of connecting a signal source with a common ground reference.

1.15.4.2 Ground Loops

What should be typically avoided, however, is connecting both the signal source AND the DAQ input analog ground to V-. This creates the dreaded “ground loop” that has created so much suffering in the measurement world.

1.16 Temperature Measurements

Temperature is almost certainly the most commonly measured phenomenon in data acquisition. Whether the application is deep beneath the sea, on the highway, in the air, or in deep outer space, temperature plays a key roll in many systems. The most common temperature sensors are the Thermocouple, the RTD (Resistance Temperature Detector), the Thermistor, and the Semiconductor temperature sensor.

Entire books have been written regarding temperature measurement and an in-depth coverage is beyond the scope of this article, but we will offer the following abbreviated discussion which should provide enough information for most users in most applications.

1.16.1 Which Sensor to Use?

In many cases, more than one of the temperature sensor types would provide the required results. However, considering only the following factors will almost always point to a clear favorite for a given application. These factors are:

- Accuracy / Sensitivity
- Temperature Measurement Range
- Cost
- System Simplicity

Table 2 provides a quick overview of the four most popular temperature sensors.

The Thermocouple (a.k.a. TC) is the workhorse of the temperature measurement world. It offers an excellent combination of reasonable accuracy, wide temperature measuring range, low cost, and can be measured with simple inputs. The RTD offers exceptional accuracy, repeatability, and a wide measurement range, but is fairly expensive and is somewhat complex to use. Interestingly, thermistors range from very inexpensive, low accuracy devices all the way to very expensive, high accuracy units. The thermistor measures temperature over a fairly limited range and is somewhat complex to use. Finally, the semiconductor sensor offers reasonable accuracy, a limited measurement range, and can be monitored with simple systems. Semiconductor sensors are also very inexpensive. A more detailed description of each of these sensors is provided in the following sections.

Table 2. **Comparison of Key Temperature Sensor Parameters**

	Accuracy	Temp Range (approximate)	Cost	Measurement Complexity	Notes
Thermocouple	low	-200 to 1800°C	Low	Medium	Rugged and Reliable
RTD	High	-200 to 850°C	High	Complex	Accurate but expensive
Low cost Thermistor	Very Low	-40 to 120°C	Very low	Complex	Often Fragile
High Accuracy Thermistor	High	-80 to 150°C	High	Complex	Fragile but highest acc
Semiconductor Temp Sensor	medium	-55 to 125°C	Low	Simple	Easy to use and low cost

Specifications are for “typical” sensors. Special order or function sensors may be available with substantially different characteristics.

1.16.2 Thermocouples

The thermocouple (a.k.a. TC) is the most commonly used temperature sensor. It offers a combination of good accuracy, wide measurement range, low cost, and simple usage. The thermocouple is also available in a wide variety of formats including bare wire, stainless steel probes, hypodermic probes, and bolt-on units. The wide variety of thermocouple configurations makes it easy to mechanically adapt the sensor to the application.

The thermocouple’s ability to sense temperature is based on the so-called “Seebeck Effect”, discovered by Thomas Seebeck in 1821. The Seebeck Effect, also known as the thermoelectric effect, states that any electrical conductor will produce a voltage when subjected to a thermal gradient. The magnitude and polarity of the voltage produced varies with the type of metal used for the conductor and the magnitude of the thermal gradient.

A thermocouple is constructed by connecting two conductors, composed of dissimilar metals. Since the second conductor senses the same thermal gradient as the first, it also produces a voltage. This voltage, however, is different from that of the first conductor because it is made from a different metal. The small difference between the two voltages, which is typically in the millivolt range, is used for measurement of the thermal gradient. A simple way[1] to think of a thermocouple is to think of it as any connection of two dissimilar metals that produces an output voltage related to the junction temperature.

---

[1] Though not strictly true from a physics point of view, this description of a thermocouple is commonly used and induces no error or confusion in the application of the device

1.16.2.1 Thermocouple Types

Any combination of dissimilar metals creates a thermocouple when they are connected. Wrap a copper wire around a steel coat hanger and you have created a thermocouple. However, you would have little idea what the characteristics of your coat hanger TC would be. In order to use it to make any sort of reasonably accurate temperature measurement, you would have to characterize it by measuring the output voltage at various temperatures. This would be a long and laborious task, and since you could not be assured the next coat hanger you wanted to use had an identical chemical composition, you could not use the same characteristics for the next TC you created.

To simplify the use of TCs, the industry has created a number of standard thermocouples, with standard chemistries that provide predictable and repeatable measurements. These TCs are typically designated by a single letter. Popular thermocouple types include: J, K, T, E, R, S, and N. There are also a number of less popular, but still well defined types available. Though many of the TCs offer similar capabilities, they each offer a different combination of scale factor, overall accuracy, measurement temperature range, and cost. The table below describes the more popular types of TCs and their defining characteristics.

Table 3. Thermocouple Characteristics[1]

Type	Metals (+)/(-)	Temp (°C) (approximate)	Scale Factor @ 25 °C	Accuracy* (Greater of)	Notes
J	Iron/Constantan	-210 to 760°C	52 µV / °C	1.1 °C or 0.4%	Wide range, general purpose
K	Chromel/Alumel	-270 to 1370°C	41 µV / °C	1.1 °C or 0.4%	Wide range, general purpose
T	Copper/ Constantan	-270 to 400°C	40 µV / °C	0.5 °C or 0.4%	High accuracy, narrow range
E	Chromel/ Constantan	-200 to 1000°C	61 µV / °C	1.0 °C or 0.4%	High output per degree
R	Pt/Pt with 13% Rh	0 to 1700°C	6 µV / °C	0.6 °C or 0.1%	High Temp
S	Pt/Pt with 10% Rh	0 to 1700°C	6 µV / °C	0.6 °C or 0.1%	High Temp
N	Ni Cr SI/Ni Si Mg	-270 to 1300°C	52µV / °C	1.1 °C or 0.4%	Stable at high temps

1. Various manufacturers offer TCs in both “Standard” and “Special” grades. The “Special” grade TCs have been selected with tighter tolerances and will provide higher accuracy than “Standard”. We have used “Special” grade specifications, except where noted with a (std).

TC characteristics are now standardized and under the general control of NIST. For additional information on each of these thermocouple types please visit: http://srdata.nist.gov/its90/main/

1.16.2.2 Cold Junction Compensation

One unfortunate, but nonetheless important, complication of thermocouple-based temperature measurement is that each system consists of more than one thermocouple. The primary measurement thermocouple is created at the point where the two dissimilar metals are intentionally connected. However, two additional thermocouples are created where the main thermocouple wire is attached to the DAQ system. These “undesirable” thermocouples are referred to as the “cold junctions” or “reference junctions.” **Figure 11** depicts a typical thermocouple input configuration showing both the “measurement” and “cold” thermocouple junctions. The cold-junction also creates an output voltage which, if not compensated for, becomes an error signal. The elimination of the cold junction error is referred to as cold junction compensation.

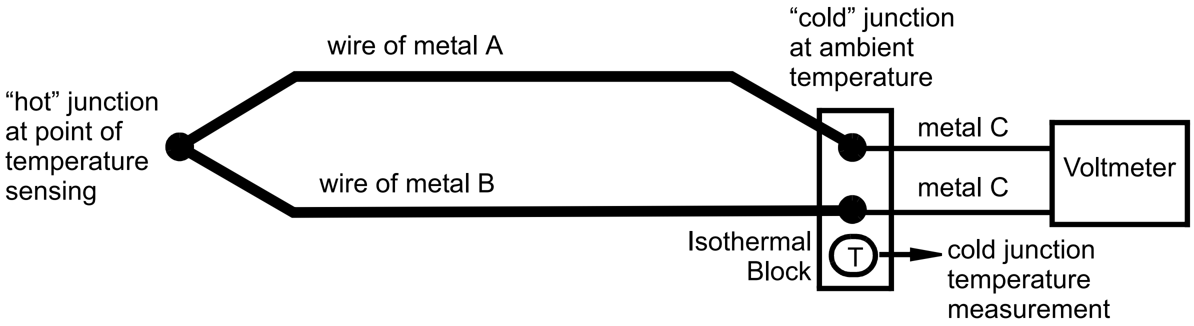


Figure 11. Thermocouple Connection with Cold Junction Compensation

Each thermocouple has a well-defined scale function that allows the user to convert the thermocouple output voltage into temperature. However, if the cold junction voltage is not somehow removed, a substantial error will be induced. This is typically performed by first measuring the temperature of the cold junction. Since there is a well defined mathematical conversion between voltage and temperature, it is possible to for the system software to convert the temperature measured at the cold junction into a voltage. This error voltage may then be subtracted from the analog input’s thermocouple voltage measurement before performing the voltage to temperature conversion.

The inquisitive reader may be wondering at this point what thermocouple transfer function should be used at the cold junction since we do not know the chemical content of the screw terminals. Mercifully, this is not an issue and the calculations are based on the transfer function of the primary thermocouple. The law of intermediate metals[1]states that a third metal, inserted between the two dissimilar metals of a thermocouple junction, will have no effect provided that the two junctions are at the same temperature. This allows us to ignore the chemistry of the screw terminals as well as all the intermediate metals between the screw terminals and the A/D converter.

---

[1] Based on this law, it is quite acceptable to make a thermocouple junction by soldering the two metals together as the solder will not affect the reading. In practice, however, thermocouple junctions are more frequently made by welding the two metals together (usually by capacitive discharge) because this ensures that the performance is not limited by the melting point of solder.

It is important to keep this in mind when we construct screw terminal panels and other interconnection devices that will be used for thermocouples. If the two terminals where the TC is connected are not at the same temperature, the cold junction compensation algorithm will not provide an accurate result. Also, if the interconnection terminals are not at the same temperature as the cold junction temperature sensor, an error will be generated. UEI’s DNA-STP-AI-U and DNA-STP-207TC screw terminal panels make use of a device called an isothermal bar to minimize these errors. The isothermal bar is typically a solid bar made of a good thermal conductor (usually aluminum or copper) placed in close proximity to the cold junction temperature sensor and the thermocouple connection terminals. Its high thermal conductivity, combined with its large thermal mass, combine to help keep the entire system at a single uniform temperature. The thermal mass of the bar also has the added advantage that it reduces the system’s susceptibility to fast, localized temperature changes caused by air currents as well as other phenomena.

1.16.2.3 Linearization

Once the cold junction error voltage is removed, it is possible to accurately convert the TC output voltage into temperature. Though the output voltage over the temperature range of each TC type is well characterized, the relationship is not linear and requires that some mathematics be applied. The TC to voltage conversions are available at the NIST website at: http://srdata.nist.gov/its90/main/

NIST provides the conversion in a look-up table which can be referred to for manual conversions or may be copied into a program or database to be referred to when required. The conversion is also provided as a high order (typically 5th or greater) polynomial. For example, to convert the voltage measured from a J-type TC into temperature (over the range of -210 to 760 °C) o, you may use the following equation.

$$T = a + bVin + cVin^2 + dVin^3 + eVin^4 + fVin^5 + gVin^6 + hVin^7 + iVin^8$$

where:

a = 0.000000000000E+00
b = 0.503811878150E-01
c = 0.304758369300E-04
d = -0.856810657200E-07
e = 0.132281952950E-09
f = -0.170529583370E-12
g = 0.209480906970E-15
h = -0.125383953360E-18
j = 0.156317256970E-22

Either the lookup table or the polynomial approximation may be used within the accuracy of the TC itself. Most computer programs will perform the conversion based on the polynomial approximation. (NIST also provides the reverse equation, which returns a voltage output for a temperature input. This equation is required to determine the Cold Junction error based upon the measurement of the screw terminal temperatures).

The good news in all of this is that most manufacturers provide the software required to automatically convert the TC input voltage into temperature. One final note regarding the TC is that most systems will require users to dedicate an input channel to monitor the cold junction temperature sensor that can then be used by the program for CJC.

1.16.3 The RTD (or Resistance Temperature Detector)

The RTD is most frequently used in applications where more temperature measurement accuracy is required than possible (or at least easily achievable) using thermocouples. In addition to excellent accuracy, the Platinum RTD (often referred to as a PRT or PRTD) offers exceptional stability, provides accurate measurements over the range of -200 °C to +850 °C, and is very resistant to corrosion or other chemical degradation.

Credit for the invention of the RTD is often given to Sir William Siemens (one of the brothers who founded the very successful Siemens Corporation) though it should more rightfully go to Sir Humphrey Davy. It was Davy who in 1821[1] discovered that the resistance of metals varied with temperature. However, 50 years later, Siemens began using Platinum as the basis of his RTDs and Pt remains overwhelmingly the element of choice in RTDs to this day.

By far the most common of all the RTDs are the 100-ohm Platinum units. These devices offer a resistance of precisely 100 ohms at 0°C. The two popular versions of the 100 ohm Pt RTD offer slightly different scale factors of 0.385 and 0.392 ohms per °C. The 0.385 version is typically called the “DIN” standard while the 0.392 version is commonly called the “American” standard. The various RTD types are typically designated by an $\alpha = 0.00385$ and 0.00392. “α” actually represents the average scale factor in ohms per ohm per °C. Since the base resistance is 100 ohms, the overall scale factors are 100 times, or 0.385 and 0.392 ohms per °C respectively.

Although the resistance of a platinum RTD varies directly with temperature over a wide range, it is not a perfectly linear relationship. To correct for inherent linearity errors, the industry uses the Callendar-Van Dusen equation and coefficients, developed many years ago by a British physicist named Hugh Callendar and later extended to temperatures below 0°C by Van Dusen. The Callendar-VanDusen equation for temperatures above 0°C is:

$$R_1 = R_0[1 + AT + BT^2 + C(T - 100)T^3]$$

where R_0 is the resistance in ohms at 0°C
R_1 is the resistance in ohms at 100°C
T is the temperature in °C

A, B, and C are the Callendar-Van Dusen coefficients, which are determined from resistance measurements at 0°C, 100°C, and 260°C and other empirical constants called α, δ, and β. The relationships between these constants and the Callendar-Van Dusen coefficients are as follows:

$$A= \alpha + (\alpha+\beta)/100 \qquad B= (-\alpha+\delta)/100^2 C = (-\alpha + \beta)/100^4$$

$$\alpha = (R_{100} - R_0)/(100*R_0) \qquad \beta = \text{Constant for } T < 0°C$$

$$\delta = [R_0 (1 + \alpha\ 260) - R_{260}]/4.16\ R_0\ \alpha$$

The international standard for defining Class A and B performance of platinum RTDs is the IEC 751. The DIN 43760, BS-1904, JIS-C1604 all match the IEC 751 standard. Only platinum RTDs have an international standard.

---

[1] Interestingly, this is the same year Seebeck made his discovery of the thermoelectric effect that is the basis for today’s thermocouples.

The following chart shows the characteristic curves of resistance vs. temperature for two types of RTDs and, for comparison, the curve for a typical thermistor. The near linear curve for the platinum RTD is a major reason this type of RTD is the most used in industry.

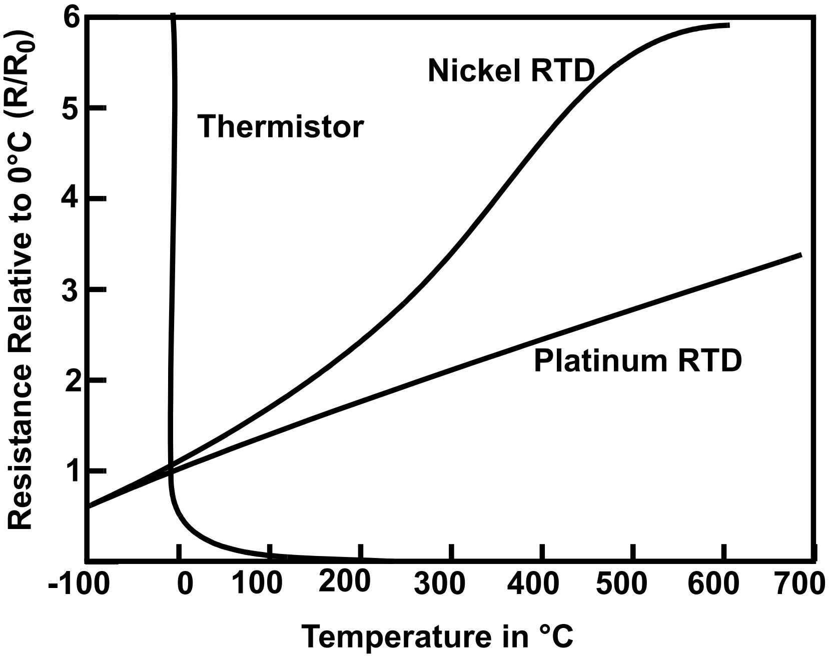


Figure 12. Characteristic Curves of RTDs and Thermistors

The classical method of measuring change in resistance with respect to temperature is to use a Wheatstone bridge, such as shown below:

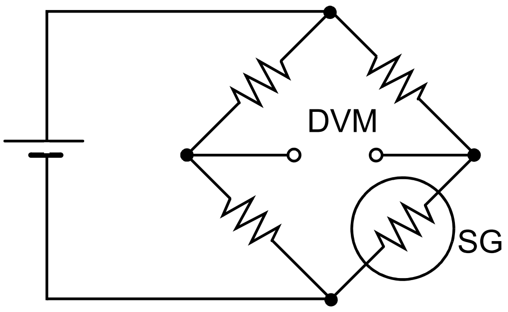


Figure 13. Wheatstone Bridge

Because the Pt RTD has such a low value of resistance (100 ohms at 0°C), however, care must be taken to avoid errors caused by lead resistances.

The bridge output voltage is an indirect indication of the RTD resistance. The bridge requires four connection wires, an external source, and three resistors that have a zero temperature coefficient. To avoid subjecting the three bridge-completion resistors to the same temperature as the RTD, the RTD is separated from the bridge by a pair of extension wires, as shown below:

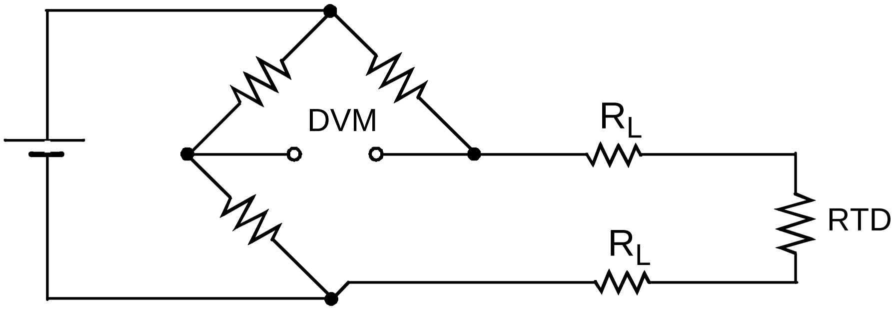


Figure 14. Two-Wire RTD Circuit

The impedance of the extension wires, however, affects the temperature reading, introducing errors that may be significant. One way this effect can be minimized is by using a *three-wire bridge* configuration, as shown below:

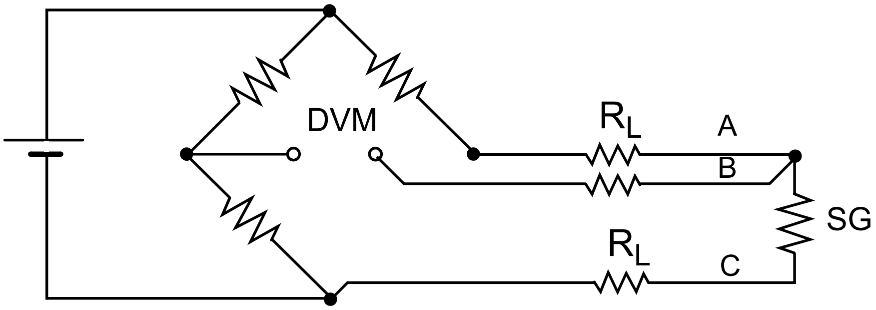


Figure 15. 3-wire RTD Circuit

In this configuration, if wires A and C are perfectly matched in length, their impedance effects cancel because each is in an opposite leg of the bridge. The third wire, B, acts as a sense lead, carries no current, and thus does not introduce any error in the measurement.

A better technique for eliminating errors caused by lead resistances is to use a four-wire RTD circuit with a current source and digital voltmeter, as shown below. This method avoids many problems associated with bridge circuits.

.

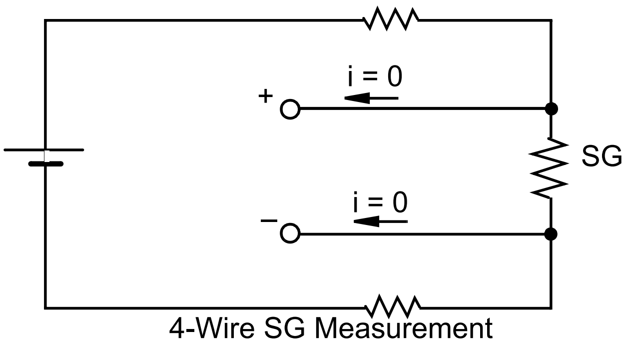


Figure 16. 4-wire RTD Circuit

In a four-wire RTD circuit, the output voltage read by the DVM is directly proportional to RTD resistance, so only one conversion calculation is necessary. The three bridge-completion resistors are replaced by one reference resistor. The digital voltmeter measures only the voltage drop across the RTD and is insensitive to the length of the lead wires.

The one disadvantage of using a 4-wire RTD circuit is that it requires one more extension wire than the 3-wire bridge circuit.

1.16.3.1 Metal Film RTDs

Another type of RTD, the metal film RTD, has become more popular recently. It is constructed by depositing or screening a metal-glass slurry film on a small ceramic substrate, trimming the RTD with a laser beam, and then sealing the unit. The film RTD has higher values of resistance than other RTDs, smaller size, fast response, and low production cost. Although different expansion rates between the substrate and the deposited RTD material, such as platinum, can cause strain errors and stability problems, the advantages of the metal-film concept make the device increasingly popular.

1.16.3.2 Accessory Equipment for RTDs

RTDs are frequently used with metal thermowells and other mechanical or chemical protection sheaths for harsh environments. Sometimes, however, more design effort is required for choosing the optimum protection device than for selecting the RTD sensor element.

1.16.4 The Thermistor

A thermistor is a highly non-linear temperature sensing device whose resistance varies with temperature. They are available in two basic types: PTC (positive temperature coefficient) or NTC (negative temperature coefficient).The PTC thermistor is sometimes called a Posistor, but the NTC is always called a thermistor. Figure 12 on page 29 showing characteristic curves of RTDs and thermistors illustrates how non-linear a thermistor curve is. The thermistor was invented and patented by Samuel Ruben in 1930.

An NTC thermistor is composed of a sintered semiconductor material with a resistance that decreases significantly with a small change in temperature. A thermistor is one of the most accurate types of temperature sensors, with typical accuracies in the ± 0.1 or ± 0.2 °C range, but its application is typically limited to the range of 0°C to 100°C.

Because of the highly non-linear characteristic of the thermistor, the instrument or DAQ system used to read the temperature must linearize the measurement. Some manufacturers, however, do produce thermistors (at higher cost) with built-in linearization components that make the unit act as a linear device.

1.16.4.1 Linearization — Steinhart-Hart Equation

For accurate temperature measurements, the Steinhart-Hart equation is used as a third-order approximation, as follows:

$$\frac{1}{T} = A + B(\ln R) + C(\ln R)^3$$

where

T is the absolute temperature in Kelvin and A, B, and C are constants that can be determined by measuring three sets of resistance and temperature values during calibration.

Most thermistors are NTC units in which resistance decreases with an increase in temperature. They are specified according to their nominal resistance at 25°C, which is typically in the range from 250 ohms to 100 kohms.

1.16.4.2 Self-Heating Effects

When current flows through a thermistor, it generates heat that raises the temperature of the themistor above the surrounding environment. If the thermistor is being used to measure temperature, the self-heating of the unit can introduce an error if not corrected.

1.17 Strain (& Stress) Measurements

1.17.1 Introduction to the Strain Gauge

The Strain Gauge (a.k.a. Strain Gage) is one of the most commonly measured devices in data acquisition and DAQ systems. Strain is often measured as the actual parameter of interest. If the application is actually interested in how much an object expands, contracts, or twists, the desired measurement is strain.

Strain is also frequently measured as an intermediate means to measure stress, where stress is the force required to induce a strain. Perhaps the most common examples of this translated measurement are load cells, where the strain of a known, well characterized metallic bar is measured, though the actual output scale factor of the cell is in units of force (e.g. pounds or newtons). The stress/strain relationship is well defined in many materials in certain configurations, making the conversion from strain into stress a straightforward mathematical calculation. Making matters easier still is that for many materials, including virtually all metals, the relationship between stress and strain when the stress is applied in pure tension or compression is linear. The linearity of the relationship is referred to as Hooke’s law, while the actual coefficient that describes the relationship is commonly referred to as either the modulus of elasticity or Young’s modulus.

Whether stress or strain is the actual measurement of interest, the mechanics of the strain gauge and the electronics required to make the measurement are virtually identical. To create a simple strain gauge, you need only firmly attach a length of wire to the object being strained. If attached in-line with the strain as the object lengthens under tension, the wire too is lengthened. As the wire length increases so does its resistance. On the other hand, if the strained object is compressed, the length of the wire decreases, and there is a corresponding change in the wire’s resistance. Measure the resistance change and you have an indication of the strain changes of your object. Of course, the scale factor needed to convert the resistance change into strain would have to be determined some how, and it would not be a trivial process. Also, the resistance change for a small strain change would be miniscule, making the measurement a difficult one.

Today's strain gauge manufacturers have solved both the scale factor and, to a certain extent, the magnitude of resistance change issues. To increase the output (resistance change) per unit of strain, today's strain gauges are typically created by placing multiple "wires" in a zig-zag configuration (see **Figure 17**) A strain gauge with 10 zigs and 10 zags would effectively increase the output scale factor by a factor of 20 over the single wire example. For a simple application, all you need to do is align the strain gauge so the "long" elements are in parallel to the direction of strain to measure, and affix the gauge with an appropriate adhesive.

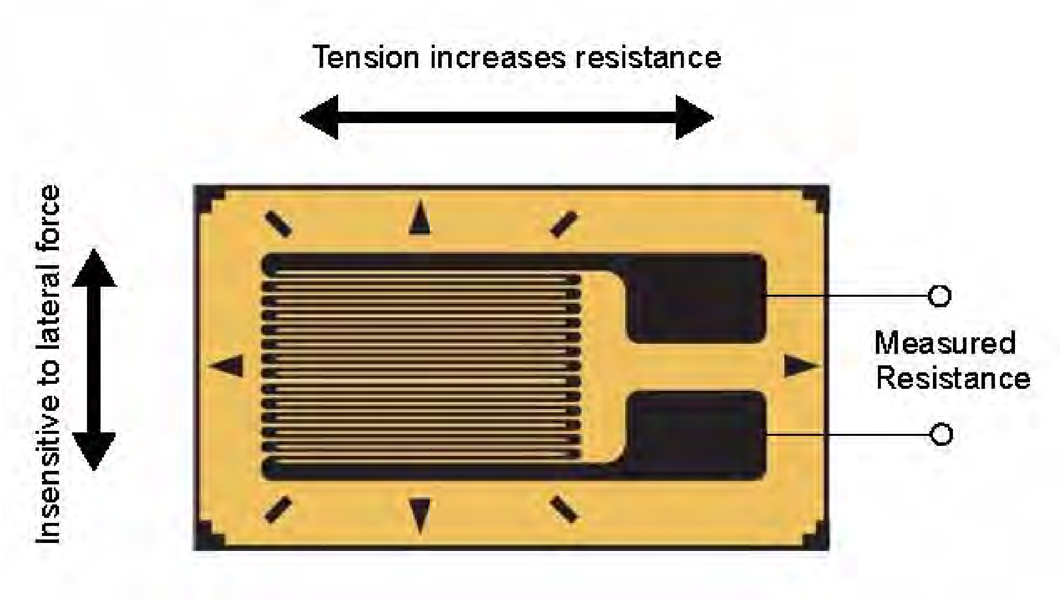


Figure 17. Typical Strain Gauge Layout

A typical strain gauge layout is shown above. As the substrate is stretched in tension, the resistance of the gauge increases. As the gauge is relaxed, the resistance decreases. Strain can thus be deduced by measuring the resistance of the strain gauge.

The strain gauge manufacturers also provide gauges with very accurate scale factors. This allows users to convert the resistance measurement into strain, with a simple, linear equation (not including temperature effects…more on this later). The scale factor of a strain gauge is referred to as its Gauge Factor, which depending on the source is commonly abbreviated as GF, F_g, or even K. For our purposes, we will use the generic acronym GF. The GF is defined by the following equation.

$$GF = \frac{R}{R_0}$$

where

R = the change in resistance induced by the strain

R_o = the resistance of the "unstrained" gauge

= the induced strain (in units of length per unit of length (e.g. inches/inch)

Of course, we don't want an equation solving for GF. We wish to solve for as shown below.

$$= (R / R_o) / GF$$

To further simplify our discussion, we'll note that by far the most common form of strain gauge is the metal foil gauge. These are available in a wide variety of sizes and configurations, but most have two common threads. First, most are offered with R_o=120, 350 or 1000 ohms. The second commonality is that most metal foil gauges have a GF of approximately 2. Note that even with today's amazing ability to manufacture consistent devices, if is impossible (or at least economically not feasible) to produce these strain gauges with a predetermined scale factor accurate enough not to compromise the ability of the gauge to measure strain. However, gauges within a specific manufacturing batch are extremely consistent. Current strain gauge manufacturers take advantage of this fact and build a "batch" of strain gauges. A number of these are tested and the GF for the entire batch is determined experimentally. Each of the gauges in a batch is then labeled with the appropriate GF.

1.17.2 Strain Gauge Measurements

Strain gauge outputs are relatively small in terms of the resistance change for strains in the area of interest. For example, many, if not most strain applications require a full scale output in the region of ±1% strain. If we assume GF=2, and R_o = 120 ohm, the full scale output in terms of resistance is ±2.4 ohms. This is a relatively small output. As such, most strain gauge measurement systems are based upon an electrical circuit referred to as a Wheatstone Bridge. This is shown in **Figure 18.**

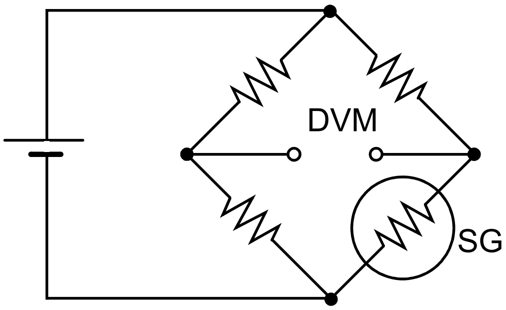


Figure 18. Typical Wheatstone Bridge

The Wheatstone Bridge configuration is ideal for measuring small changes in resistance without requiring ultra-high performance analog inputs.

The classic Wheatstone bridge equation is shown below.

$$V_0 = V_x\left(\frac{R_d}{R_c + R_d}\right) - \frac{R_b}{(R_a + R_b)}$$

This equation is provided as part of every tutorial on the Wheatstone bridge, and while accurate, it is not particularly useful in an automated data acquisition application. It is unknown why so few texts take the next step, but what is really required is to solve this equation for the resistance value of the unknown. In this case, we will assume the strain gauge is inserted in place of Rc, and thus we need to solve for Rc.

With a bit of careful algebra we arrive at the following equation where Rc is the calculated unknown. Vx is the excitation voltage. Vo is the output voltage measured by the DAQ system and Ra, Rb, and Rd are bridge completion resistors.

$$R_c = \frac{R_d - ((R_d V_0)/V_x - (R_d R_b)/(R_a + R_b))}{V_0/V_x + R_b/(R_a + R_b)}$$

It’s not a pretty equation, and is a bear to calculate repetitively by hand, but who calculates things by hand these days? It really is quite easy to implement this equation in Excel, LabView, or any programming language. And once you have it, you won’t need to write it again. Also note that there is no linear approximation in this equation, so it will work for larger changes, if the bridge is out of balance.

In many cases, you may also simplify the equation because Ra = Rb. In this case the equation for Rc drops to:

$$R_c = \frac{R_d(V_x - 2V_0)}{2V_0 + V_x}$$

We may know the static/unloaded value of Rc from *a prioi* knowledge, or we may need to calculate Rc based upon the measurement. However, once we have the initial value of Rc (we’ll call it Rc_o), we can then go on to solve our equation for the strain gauge. Recalling our original strain equation,

$$R_c = \frac{R/R_0}{GF}$$

We can now insert the values for Rc. Once again assuming Rc_o is the unstrained condition, and based upon our measurement and the calculations above, Rc_1 is the resistance measurement of Rc when strained. GF again is the Gain Factor constant (typically ~ 2.0).

$$= (R_{c1} - R_{c0})/R_{c0}/(GF)$$

or

$$R_c = R_{c1}/(R_{c0} - 1)/(GF)$$

If the variable of interest is Stress rather than Strain, we can now proceed to calculate it. In the simplest case, with a sample in tension and/or compression and with a known Modulus of elasticity (a.k.a. Young's Modulus), you may calculate Stress with the following equation:

$$\sigma = \frac{\varepsilon}{E}$$

where τ is the stress, ε is the strain, and E is the Modulus of Elasticity. Recognizing that stress is defined as the Force per unit of area, we can generalize the above equation into:

$$F / A = \varepsilon / E$$

where F is the force applied and A is the cross sectional area of the object under test.

If the Modulus of Elasticity is known and the cross sectional area of the object under test is known, we can use the strain gauge to determine the force applied.

$$F = \frac{\varepsilon A}{E}$$

The ability to determine force with a strain gauge is the basis for almost all electronic scales and load cells.

1.17.3 Temperature Effects in Strain Measurement

Though complete coverage of the impact of temperature on strain measurements is beyond the scope of this document, it is too important a topic not to be addressed. Temperature adversely impacts strain measurements in many ways, though three are of primary concern.

- The device or object studied will almost always have a non-zero coefficient of thermal expansion. Unless compensated for, changes in temperature will cause the item to which the strain gauge is attached to expand or contract, which is then indicated as a change in strain.
- The materials of the strain gauge itself have a non-zero coefficient of thermal expansion. Changes in temperature will cause the strain gauge itself to expand or contract, independent of any strain in the part to which it is attached.
- The wiring and the strain gauge itself will have a non-zero Temperature Coefficient of Resistance. That is, as the temperature changes, the resistance of the strain gauge and connecting wires will change independently of any change in strain. (For example, copper wire resistance changes at approximately 3900 ppm per °C (.393% /°C).)

1.17.4 Thermal Expansion/Contraction Issues

Some texts treat the first two items as the same effect. After all, if the coefficients of expansion of the gauge and the item under test are the same, they will contract or expand at the same rates in response to a temperature change. In this case, a change in system temperature would not cause any change in the indicated strain, except that based on the gauge's temperature coefficient of resistance.

It's important to note that in some applications, it may be desirable or even critical that strain induced by temperature changes be noted. Envision an application where a "hot section" turbine blade is being tested to ensure proper clearance between the blade tip and the surrounding shroud. It's important to know how much the blade has elongated based upon temperature in addition to the centrifugal force of rotation.

On the other hand, if the parameter of interest is really stress, or its close relative, force, any strain caused by temperature changes would induce a true error in the result. A strain gauge used to measure the "g" forces on a supersonic aircraft wing skin might see temperatures from -45°C to 200 °C. If the g-force information was critical to not overstressing the wing, you'd certainly not want significant temperature-induced error. In a more simple case, the load cell used to measure the force placed on a postal scale should not induce errors simply because the scale is next to the window on a sunny summer day!

Most applications fall into the second category, where the key measurement parameter is really stress, and the ideal system would be not to recognize any changes caused by thermal expansion or contraction. Like most engineering challenges, there is more than one way to skin this proverbial cat. They are: (1) Calculate the error and eliminate it mathematically, (2) Match the strain gauge to the part, (3) Use an identical strain gauge in another leg of the bridge.

1.17.5 Calculate the Error and Eliminate It Mathematically

If you know the actual difference in coefficients of thermal expansion between the strain gauge and the part being tested, it's theoretically possible to mathematically eliminate the error caused by changes in temperature. Of course, to do this, you also need to measure the temperature accurately at the strain gauge installation. The strain gauge expansion coefficients however, are not generally available from the manufacturer as they tend to change from batch to batch. Though possible, compensating for temperature effects using only this "calculated" method is seldom done.

More common, but still not very common, a pseudo-calculated method is performed. Rather than use *a priori* predicted coefficients to calculate the differential strain induced by temperature changes, it is possible to determine the function experimentally. The word function is used intentionally, as the actual strain versus temperature curve is infrequently linear, especially over large temperature changes. However, if the application allows the system's strain vs. temperature curve to be determined experimentally, it becomes fairly straightforward to remove the error mathematically.

1.17.6 Match the Strain Gauge to the Part Tested

The use of different alloys/metals allows manufacturers to provide strain gauges designed to match the thermal expansion/contraction behavior of a wide variety of materials commonly subject to strain (and stress) testing. This type of gauge is referred to as a "Self Temperature Compensated" (or STC) strain gauge. These STC gauges are available from a variety of manufacturers and are specified for use on a wide assortment of part materials. As you might imagine, the more common a metal, the better the chances are there is an STC gage that matches. However, you may count on being able to find a good match for such materials as aluminum, brass, cast iron, copper, carbon steel, stainless steel, titanium and many more.

Though the match between the STC gauge and the part under test may not be perfect, it will typically be accurate enough from freezing to well past the boiling point of water. For more details on the precise accuracy to expect, you should contact your strain gauge manufacturer.

1.17.7 Use an Identical Strain Gauge in Another Leg of the Bridge

Due to the ratiometric nature of the Wheatstone bridge, a second, unstrained gauge (often referred to as a "dummy" gauge) placed in another leg of the bridge will compensate for temperature induced strain. Note that the dummy gauge should be identical to the "measuring" gauge and should be subject to the same environment.

Strain gauges tend to be small, and have short thermal time constants (i.e., their temperature changes very quickly in response to a temperature change around them), while the part under test may have substantial thermal mass and may change temperature slowly. For this reason, it is good practice to mount the dummy gauge adjacent to gauge being measured. However, it should be attached in such a way as not to be subjected to the induced strain of the tested part.

In some cases, with relatively thin subjects and when measuring bending strain (as opposed to pure tensile or compressive strain), it may be possible to mount the dummy gauge on the opposite side of a bar or beam. In this case, the temperature impact of the gauges is eliminated and the scale factor of the output is effectively doubled.

1.17.8 Quarter, Half and Full Bridges

Strain gauges and measurement devices based upon strain gauges (e.g., load cells) can be configured in three different configurations. These are referred to as Quarter, Half and Full Bridges.

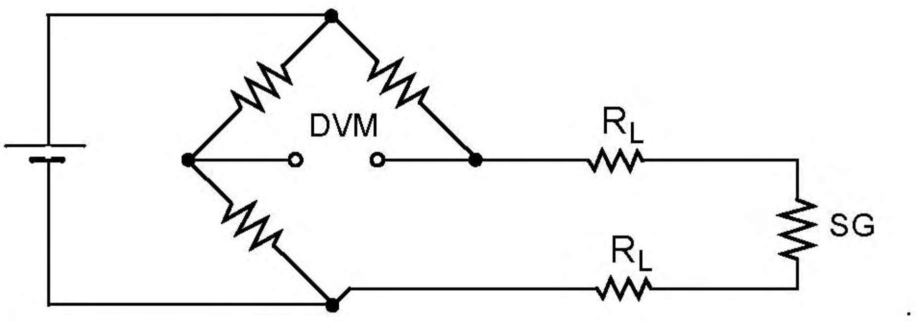


Figure 19. Quarter Bridge Strain Gauge

The quarter bridge gauge shown in **Figure 19** is the simplest and probably most common strain gauge configuration (though some devices “based” on strain gauges are more likely to be provided in half or full bridge). The name “quarter” comes from the fact that in this configuration, the strain gauge represents one out of four, or one quarter of the resistors in the Wheatstone bridge. In this configuration, the user must supply the other three resistors.

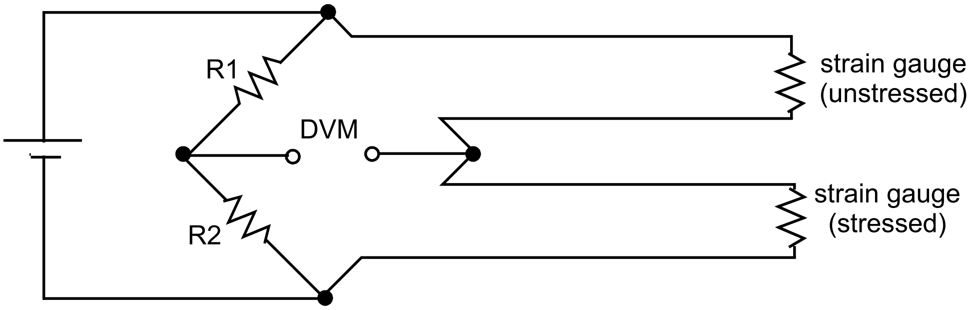


Figure 20. Half Bridge Strain Gauge

In the half-bridge configuration shown in **Figure 20,** two resistors or half of the bridge are provided in the strain gauge itself. Half Bridge configurations have two advantages over the single bridge. First, they simply require the user to provide one less resistor. Second and more important, however, is the fact that most half bridge sensors automatically provide temperature compensation, made possible by having two identical gauges in the same side of the bridge.

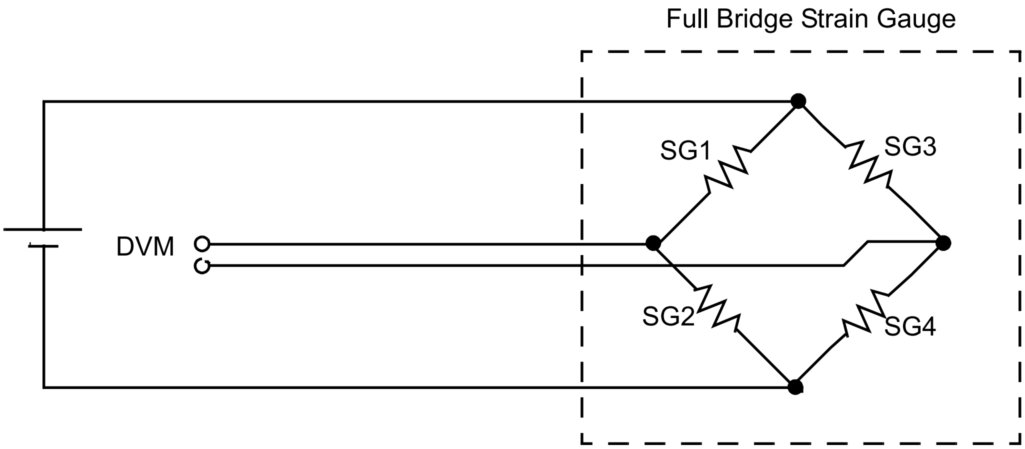


Figure 21. Full Bridge Strain Gauge

As you might expect, the full bridge sensor shown in **Figure 21** provides all four resistors, in effect, providing the entire bridge. All the measurement system needs to provide is an excitation voltage and a differential analog input. Like the half-bridge configuration, most full bridge gauges are temperature compensated.

Index