

When selecting a *sensor*, the first thing to do is to outline **the requirements** for the particular application.

When knowing what is needed, one is ready to evaluate what is available.

Emerging applications



Industrial applications: for detecting noncontact temperature, thermal imaging, humidity, air flow, ionizing radiation, smell, dielectric constant of objects, material composition, range (distance), air pressure, produce freshness, etc. Medical applications: for the inner (core) and skin body temperatures, thermal imaging, arterial blood pressure, EKG, blood factors (glucose, cholesterol, haemoglobin oxygen saturation), deep body imaging, smell (e-nose), behaviour modification, etc. Military: for night vision, detecting poisonous gases, proximity, ionizing radiation, explosives, chemical and biological agents, etc. **Consumer**: for the body core temperature, heart rate, radon gas, pregnancy detection, breathalyser for alcohol and hydrogen sulphide, food composition, behaviour modification, proximity, UV level, electromagnetic pollution, surface temperature, etc.

In the past decade, a major market for sensors has emerged:

the Mobile Communication Devices (MCD)

such as

- smartphones,
- smartwatches,
- tablets



No longer a telephone is just for far (tele) transmission of speech voice (phone), it has been evolving into our personal **cyber-valet** that can perform a multitude of service.



Basic applications



Today a generic MCD contains a rather limited number of sensors, yet they support thousands of the apps for many industrial, scientific, consumer, and medical purposes.

These currently built-in sensors are:

Imaging camera: it takes photos and videos.

Microphone: it detects sound mostly in the audible frequency range.

Accelerometer: it detects motion of the MCD and direction of the gravity force.

<u>Gyroscope</u>: it measures spatial orientation of the MCD.

Magnetometer (compass): it detects strength and direction of magnetic fields.

<u>GPS</u>: an RF receiver and processor for identifying global coordinates.

Proximity detector: detects closeness of the MCD to the user's body.

10 COMMANDMENTS OF MOBILE SENSOR DESIGN

List

Intelligent sensor: built-in signal conditioner and DSP

Built-in communication circuit (I2C, NFC, Bluetooth, etc.)

Integrated supporting components (optics, thermostat, blower, etc.)

High selectivity of the sensed signal (reject interferences)

Fast response

Miniature size to fit a mobile device

Low power consumption

High stability in changing environment

Lifetime stability: no periodic recalibration or replacement

Low cost at sufficiently high volumes







Accuracy





(In)Accuracy

(In)accuracy rating may be represented in several forms, some of which are:

Directly in terms of measured value of a stimulus (Δ). This form is used when error is independent on the input signal magnitude. Often, it relates to additive noise or systematic bias, but also combines with all other conceivable error sources, like calibration, manufacturer's tolerances, etc. For example, it can be stated as 0.15 °C for a temperature sensor or 10 fpm (foot-per-minute) for a flow sensor. Usually, a specific range of the stimulus accompanies this statement, then the accuracy specification may read:

10 fpm in the range from below 100 ft/min, and 20 fpm in the range over 100 ft/min

- In percentage of the input span (full scale). This form is useful for a sensor with a linear transfer function and closely relates to the above form. It is just another way of stating the same thing because the input range must be specified for nearly any sensor. This form is not useful for a sensor with a nonlinear transfer function, unless a small quasilinear range is specified. For example, a thermo-anemometer has a response that can be modelled by a square root function: it is more sensitive at low-flow rates and less sensitive at high flows. Let us assume that the sensor has a span of 3000 fpm and its accuracy is stated as 3 % of the full scale, which is the other way to say 90 fpm. However, for measuring low-flow rates, say from 30 to 100 fpm, this full-scale error of 90 fpm looks huge and in fact is misleading due to a nonlinearity.
- In percentage of the measured signal. This is a multiplicative way of expressing error because the error magnitude is shown as fraction of the signal magnitude. It is useful for a sensor with a highly nonlinear transfer function. Considering the previous thermo-anemometer example, the 3 % of the measured signal is more practical for low-flow rates because it will be just few fpm, while for the high-flow rate range it will be in tens of fpm and is also reasonable and realistic. Still, using this form is not generally recommended because typically the error varies with a stimulus. It makes more sense to break up the total nonlinear span into smaller quasilinear sections. Then, the in percentage of the input span should be used instead for each individual section.
- I terms of the output signal. This is useful for sensors with a digital output format so the error can be expressed, for example, in units of LSB.

Calibration error



Calibration error is the *inaccuracy* permitted when a sensor is calibrated in the factory. This error has a systematic nature, meaning that it is added to all possible real transfer functions.

No accurate calibration is possible if one uses a not-so-accurate reference.



A hysteresis error is a deviation of the sensor's output at a specified point of the input signal when it is approached from the opposite directions. The typical causes for hysteresis are the design geometry, the friction, and the structural changes in the materials, especially in plastics and epoxy.

Non-linearity

- **line 1**: here, near the terminal points, the nonlinearity error is the smallest and it is higher somewhere in between.
- line 2: the *independent linearity* is best straight line (L₂=2·L₁). It often is used when all stimuli in the range are equally important.
- **line** 3: the sensor is calibrated in the region where the highest accuracy is desirable. Then, the approximation line may be drawn through the calibration point c, as a result, nonlinearity has the smallest value near the calibration point and increases toward the ends of the span (e.g. thermometer).

At some level of the input stimuli, the sensor output signal no longer will be responsive. Further increase in stimulus does not produce a desirable output. It is said that the sensor exhibits a span-end nonlinearity or saturation

Repeatability

Repeatability (reproducibility) error is caused by the inability of a sensor to represent the same value under presumably identical conditions. The repeatability is expressed as a maximum difference between the output readings as determined by two run cycles, unless otherwise specified.

It is usually represented as percentage of FS:

$$\delta_r = \frac{\Delta}{FS} 100\%$$

The possible sources of a repeatability error may be thermal noise, build up charge, material plasticity, etc.

stimulus (s)

Dead bands

Dead band is the insensitivity of a sensor in a specific range of the input signals. In that range, the output may remain near a certain value (often zero) over an entire dead-band zone.

Resolution describes the smallest increment of a stimulus that can be sensed.

When a stimulus continuously varies over the range, the output signals of some sensors will not be perfectly smooth, even under the no-noise conditions. The output may change in small steps. This is typical for potentiometric transducers, occupancy infrared detectors with grid masks, and other sensors where the output signal change is enabled only upon a certain degree of stimulus variation.

Besides, any signal that is converted into a digital format is broken into small steps where a number is assigned to each step. The magnitude of the input variation, which results in the output smallest step, is specified as a resolution under specified conditions (if any).

For a wire-wound potentiometric angular sensor, resolution may be specified as "a minimum angle of 0.5°". Sometimes, it may be specified as percent of a full scale (FS). For instance, for the angular sensor having 270° FS, the 0.5° resolution may be specified as 0.18 % of FS. It should be noted that the step size may vary over the range, hence, the resolution may be specified as typical, average, or worst.

The resolution of a sensor with a digital output format is given by the number of bits. For instance, resolution may be specified as "8-bit resolution". This statement to make sense must be accomplished with either of FS value or the value of LSB (least significant bit). When there are no measurable steps in the output signal, it is said that the sensor has continuous or infinitesimal resolution.

Output impedence

Sensor with current output

Sensor with **voltage** output

To minimize the signal distortions, a current generating sensor should have an output impedance Z_{out} as high as possible, while the interface circuit's input impedance should be low. Contrary, for the voltage connection, a sensor shall have a lower Z_{out} , while the interface circuit should have Z_{in} as high as practical.

*

FM

time

time

time

AS

PWM

E

E

Output format

Output Format is a set of the output electrical characteristics that are produced by the sensor alone or by its integrated excitation circuit and signal conditioner.

The characteristics may include *voltage*, *current*, *frequency*, *amplitude*, *phase*, *polarity*, *charge*, *shape of a signal*, *time delay*, and *digital code*. A sensor manufacturer should provide sufficient information on the output format to allow for efficient applications.

The most popular digital communication between an integrated sensor and peripheral device is a serial link. As the name implies, a serial link sends and receives bytes of information in a serial fashion: **one bit at a time**. These bytes are transmitted using either a binary format or a text (ASCII) format. For communicating an integrated sensor with a digital output format, the most popular formats are PWM (pulse-width modulation) and I2C and its variations.

The I2C (pronounced I-squared-C) protocol was developed by Philips Semiconductors for sending data between the I2C devices over two wires. It sends information from a sensor to a peripheral device serially using two lines: one line for data (SDA) and one for clock (SCL). The protocol is based on a concept of the master and slave devices. A master device is a controller (often a microprocessor) that is in charge of the bus at the present time and controls the clock. It also generates START and STOP signals. Slaves simply listen to the bus and act on controls or data that they sent. The master can send data to a slave or receive data from a slave—slaves do not transfer data among themselves. The basic communication speed is selected between 0 and 100 kHz. Some sensors are relatively slow (contact temperature sensors, for example), thus a slow slave module may need to stop the bus while it collects and processes data. It can do this while holding the clock line (SCL) low forcing the master into the wait state. The master must then wait until SCL is released before proceeding.

I2C protocol

 ACK/NACK Bit. If an address frame or data frame was successfully received, an ACK bit → 0 is returned to the sender from the receiving device.

- Start condition: SDA: HIGH \rightarrow LOW while SCL is HIGH
- Stop condition: SDA: LOW \rightarrow HIGH while SCL is HIGH

The start and the stop condition are the only two times in the whole I2C communication, where the SDA line changes when the SCL line is HIGH. In all other conditions the SDA line only changes state when the SCL line is LOW.

Read/Write Bit

- Write: Master is sending data to the slave: 0
- Read: Master is requesting data from the slave: 1

I2C protocol

Advantages and Disadvantages of I2C communication

Advantages 💽	Disadvantages		
Only uses two wires	Slower data transfer rate than SPI		
Supports multiple masters and multiple slaves	The size of the data frame is limited to 8 bits		
Well known and widely used protocol			

Excitation signal

Excitation is the signal needed to enable operation of an active sensor. Excitation is specified as a range of voltage and/or current, or in some cases it may be light, magnetic field and any other type of a signal. For some sensors, the frequency and shape of the excitation signal, and its stability, must also be specified. Spurious variations in the excitation may alter the sensor transfer function, produce noise, and cause output errors.

An example of an excitation signal is electric current passing through a thermistor to measure its temperature-dependent resistance.

An example of the excitation signal specification is:

Maximum current through a thermistor	
	in still air 50 μA
	in water 1 mA

Dynamic characteristics

The warm-up time characteristic is a time delay between applying power to the sensor or the excitation signal and the moment when the sensor can operate within its specified accuracy. Many sensors have a negligibly short warm-up time. However, some sensors, especially those that operate in a thermally controlled environment (a thermostat, e.g.) and many chemical sensors employing heaters, may require seconds and even minutes of a warm-up time before they are fully operational within the specified accuracy limits. In the control theory, it is common to describe the input-output relationship through a constant-coefficient linear differential equation. Then, the sensor's dynamic (time-dependent) characteristics can be studied by evaluating such an equation. Depending on the sensor design, differential equations can be of several orders.

Dynamic characteristics

- A Zero-Order sensor is characterized by a transfer function that is time-independent. Such a sensor does not incorporate any energy storage devices, like a capacitor. A zero-order sensor responds instantaneously. In other words, such a sensors does not need any dynamic characteristics to be specified. Naturally, nearly any sensor still has a finite time to respond, but such time is negligibly short and thus can be ignored.
- A First-Order differential equation describes a sensor that incorporates one energy storage component. The relationship between the input s(t) and output E(t) is a first-order differential equation:

$$b_1 \frac{dE(t)}{dt} + b_0 E(t) = s(t)$$

A dynamic characteristic of a first-order sensor may be specified by a manufacturer in various ways. A typical is a frequency response, which specifies how fast the sensor can react to a change in the input stimulus. The frequency response is expressed in *Hz* or *rad/s* to specify the relative reduction in the output signal at a certain frequency of the stimulus.

• A **Second-Order differential equation** describes a sensor that incorporates two energy storage components. The relationship between the input s(t) and output E(t) is a differential equation:

$$b_2 \frac{d^2 E(t)}{dt^2} + b_1 \frac{dE(t)}{dt} + b_0 E(t) = s(t)$$

An example of a second-order sensor is an accelerometer that incorporates an inertial mass and a spring. A second-order response is specific for a sensor whose response includes a periodic signal. Such a periodic response may be very brief and we say that the sensor is damped, or it may be of a prolonged time and even may oscillate continuously. For many sensors such a continuous oscillation is a malfunction and must be avoided. Any second-order sensor may be characterized by a resonant (natural) frequency, which is a number expressed in *Hz* or *rad/s*. The natural frequency shows where the sensor's output signal increases considerably while the stimulus does not. When the sensor behaves as if the output conforms to the standard curve of a second-order response, the manufacturer will state the <u>natural frequency</u> and <u>the damping ratio of the sensor</u>. The resonant frequency may be related to mechanical, thermal, or electrical properties of the detector. Generally, the operating frequency range for a sensor should be selected well below (at least by 60 %) or above the resonant frequency. However, in some sensors, a resonant frequency is the operating point.

Dynamic response

ENVIRONMENTAL FACTORS

Storage conditions are the nonoperating environmental limits to which a sensor may be subjected during a specified period without permanently altering its performance when operating under normal conditions. Usually, storage conditions include **the highest and lowest storage temperatures** and **maximum relative humidities** at these temperatures. Word "noncondensing" may be added to the relative humidity number. Depending on the sensor's nature, some specific limitation for storage need to be considered. For instance, maximum pressure, presence of some gases, or contaminating fumes.

A short-term stability is manifested as changes in the sensor's performance within minutes, hours, or even days. Eventually, it is another way to express repeatability as drift may be bidirectional. That is, the sensor's output signal may increase or decrease.

A long-term stability (aging) may be related to degrading of the sensor materials, which is an irreversible change in the material's electrical, mechanical, chemical, or thermal properties. A long-term drift is usually unidirectional. It happens over a relatively long time span, such as months and years. A long-term stability is very important for sensors that are used for precision measurements. Aging greatly depends on the environmental storage and operating conditions, how well the sensor components are isolated from the environment, and what materials are used for their fabrication. Aging phenomenon is typical for sensors having organic components and, in general, is not an issue for a sensor made with only nonorganic materials.

Environmental stability is a very important requirement. Both the sensor designer and application engineer should consider all possible external factors that may affect the sensor's performance. Take for example a piezoelectric accelerometer that may generate spurious signals if affected by a sudden change in ambient temperature, electrostatic discharge, triboelectric effect, vibration of a connecting cable, electromagnetic interferences. Even if a manufacturer does not specify such effects, an application engineer should simulate them while prototyping the product. If, indeed, the environmental factors degrade the sensor's performance, additional corrective measures may be required. Examples are placing the sensor in a protective enclosure, electrical shielding, using a thermal insulation or thermostat, and employing a differential design.

Reliability

Reliability is ability of a product (sensor, e.g.) to perform a required function under stated conditions for a stated period of time. Reliability can be expressed in statistical terms as the probability that the device will function without failure over a specified time or a number of uses. Reliability specifies a failure, that is, a temporary or permanent malfunction of a sensor. While reliability is an important requirement, it is seldom specified however by the sensor manufacturers. The reason for that is. perhaps, the absence of a commonly accepted measure(s) for sensor reliability.

For many repairable electronic devices, the procedure for predicting the in-service reliability is the **MTBF** (*mean-time-between-failures*) whose calculation is described in the MIL-HDBK-217 standard*. Since sensors are nonrepairable devices and so after a failure they should be replaced, not repaired. Thus sensors are more conveniently characterized by the **MTTF** (*mean-time to-failure*), an average time of operation before the device fails. The MTTF determines the dependability of the device and is computed as:

$$MTTF = \frac{1}{n} \sum_{i} (t_{fi} - t_{oi})$$

where t_0 is time of the test start, t_f is time of failure, n is total devices tested, and i is the number of a device.

* U.S. Dept. of Defense. (1991, December 2). Military handbook. Reliability prediction of electronic equipment (Mil-HDBK-217F)

Extreme testing

A device reliability can be inferred after tests at extreme conditions. One approach (suggested by MIL-STD-883*) is 1000 h, loaded at maximum temperature. This test does not qualify, however, for such important impacts as rapid temperature changes and many other factors, like humidity, ionizing radiation, shock and vibrations, etc.

Extreme tests are especially helpful in the sensor design phase to uncover hidden problems. During the extreme tests, a sensor may be subjected to some strong environmental factors, which potentially can alter its performance or uncover hidden defects. Among additional tests that may reveal such issues are:

- high temperature/high humidity, while being fully electrically powered (e.g. 85-85 test);
- mechanical shocks and vibrations may be used to simulate adverse environmental conditions, especially in evaluating wire bonds, adhesion of epoxy, etc. ;
- extreme storage conditions may be simulated, for instance at +100 and -40 °C while maintaining a sensor for at least 1000 h under these conditions;
- thermal shock or temperature cycling (TC) is subjecting a sensor to alternate extreme conditions;
- to simulate sea conditions, sensors may be subjected to a salt spray atmosphere for a specified time, for example 24 h.

*Department of Defense. (1996, December 31). Test method standard. Microcircuits (MIL-STD-883E)

Accelerated life testing

Another important method reliability of testing would be an accelerated life (AL) qualification. It is a procedure that emulates the sensor's operation, providing the real-world stresses, but compressing years into weeks. Three goals are behind such a test:

- to establish MTTF;
- identify first failure points that can then be remedied by the design changes;
- identify the overall system practical lifetime.

One possible way of compressing time is to use the same profile as the actual operating cycle, including maximum loading and power-on, power-off cycles, but **expanded environmental highest and lowest ranges** (temperature, humidity, and pressure). The highest and lowest limits should be substantially broader than normal operating conditions. Performance characteristics may be outside specifications, but must return to those when the device is brought back to the specified operating range. For example, if a sensor is specified to operate up to 50 °C at the highest relative humidity (RH) of 85 % at maximum supply voltage of +15 V, it may be cycled up to 100 °C a 99 % RH and at +18 V power supply (still lower than the maximum permissible voltages).

Accelerated life testing: Quantum dot laser (Agilent Technologies)

* Journal of Lightwave Technology Vol.24 N.1 January 2006 pgg. 143-149

Accelerated life testing

To estimate number of the test cycles (n), the following empirical formula developed by Sundstrand Corporation, Rockford, IL and Interpoint Corp., Redmond, WA* may be useful:

$$n = N \left(\frac{\Delta T_{max}}{\Delta T_{test}}\right)^{2.5}$$

where N is the estimated number of cycles per lifetime, ΔT_{max} is the maximum specified temperature fluctuation, and ΔT_{test} is the maximum cycled temperature fluctuation during the test. For instance, if the normal temperature is 25 °C, the maximum specified temperature is 50 °C, cycling was up to 100 °C, and over the lifetime (say, 10 years) the sensor was estimated being subjected to 20,000 cycles, then the number of test cycles is calculated as:

$$n = 20000 \left(\frac{50 - 25}{100 - 25}\right)^{2.5} = 1283$$

As a result, the accelerated life test requires about 1300 cycles instead of 20,000.

*E. Suhir, Y. C. Lee, C. P. Wong. *How to make a device into a product. Accelerated life testing (ALT), its role attributes, challenges, pitfalls and interaction with qualification tests*, Chapter 8. E. Suhir, Y. C. Lee, & C. P. Wong (Eds.) 2007.

When taking individual measurements under real conditions we expect that stimulus (s) is represented by the sensor as having a somewhat different value s', so that error of measurement is expressed as

$\delta = s - s'$

An error can be compensated to a certain degree by correcting its systematic component. The result of such a correction can unknowably be very close to the unknown true value of the stimulus and thus it will have a very small error. Yet, in spite of a small error, the uncertainty of measurement may be very large since we cannot really trust that the error is indeed that small. In other words, an error is what we unknowably get when we measure, while uncertainty is what we think how large that error might be.

The International Committee for Weight and Measures (CIPM) considers that uncertainty consists of many factors that can be grouped into two classes or types*:

A: those, which are evaluated by statistical methods;

B: those, which are evaluated by other means.

Generally, A components of uncertainty arise from random effects, while the B components arise from systematic effects.

*International Organization for Standardization (1993). ISO Guide to the expression of uncertainty in measurements. Geneva, Switzerland.

We simply never can be 100% sure of the measured value

Type A uncertainty is generally specified by a standard deviation σ_i , equal to the positive square root of the statistically estimated variance σ_i^2 , and the associated number of degrees of freedom v_i . For such a component the standard uncertainty is $u_i = \sigma_i$. Standard uncertainty represents each component of uncertainty that contributes to the total uncertainty of the measurement result.

Evaluation of a Type A standard uncertainty may be based on any valid statistical method for treating data. Examples are calculating standard deviation of the mean of a series of independent observations, using the method of least squares. If the measurement situation is especially complicated, one should consider obtaining the guidance of a statistician.

Evaluation of a Type B of standard uncertainty is usually based on scientific judgment using all the relevant information available, such may include:

- Previous measurement data.
- Experience with or general knowledge of the behaviour and property of relevant sensors, materials, and instruments.
- Manufacturer's specifications.
- Data obtained during calibration and other reports.
- Uncertainties assigned to reference data taken from handbooks and manuals.

When both A and B uncertainties are evaluated, they should be combined to represent the combined standard uncertainty. This can be done by using a conventional method for combining standard deviations. This method is often called the law of propagation of uncertainty (root-sum-of-squares or RSS) and it combines the uncertainty components estimated as standard deviations:

$$u_c = \sqrt{u_1^2 + u_2^2 + u_3^2 + u_4^2 + \dots + u_n^2}$$

Souce of uncertainty	Standard uncertainty [°C]	Туре
Calibration of sensor		
Reference temperature source	0.03	А
Coupling between reference and sensor	0.02	А
Measured errors		
Repeated observations	0.02	А
Sensor inherent noise	0.01	А
Amplifier noise	0.005	А
DVM error	0.005	А
Sensor aging	0.025	В
Thermal loss through connecting wires	0.015	А
Dynamic error due to sensor's inertia	0.005	В
Transmitted noise	0.02	A
Misfit of transfer function	0.02	В
Ambient drifts		
Voltage reference	0.01	А
Bridge resistors	0.01	A
Dielectric absorption in A/D capacitor	0.005	В
Digital resolution	0.01	А

To illustrate the concept, the table shows an example of the uncertainty budget for an electronic thermometer with a thermistor sensor, which measures temperature of a water bath. While compiling such a table one shall be very careful not to oversee any standard uncertainty not only in a sensor, but also in the interface instrument, experimental setup, and the object of measurement. This shall be done for various environmental conditions, which may include temperature, humidity, atmospheric pressure, power supply variations, transmitted noise, aging, and many other factors.

No matter how accurately any individual measurement is made, that is, how close the measured temperature is to true temperature of an object, one never can be sure that it is indeed accurate. The combined standard uncertainty of 0.062 °C does not mean that error of measurement is no greater than 0.062 °C.