

THE SELF-VALIDATING SENSOR: RATIONALE, DEFINITIONS AND EXAMPLES

M. P. HENRY and D. W. CLARKE

Department of Engineering Science, Parks Road, Oxford OX1 3PJ

Abstract. Traditionally, the industrial sensor has been viewed as a simple signal generator. The application of microprocessor technology, digital communications and fault detection techniques, coupled with increasing demands for measurement quality assurance, have rendered inadequate such a simplistic view. In this paper a new sensor model is proposed which encompasses new demands and capabilities. This self-validating sensor performs self-diagnostics and generates a variety of data types, including the on-line uncertainty of each measurement. A demonstration system is described, based upon a coriolis mass flow meter.

Key Words. Self-Validating Sensors; Uncertainty; Fault Detection and Diagnosis; Fieldbus; Coriolis Meter.

1. INTRODUCTION

Although the sensor provides the on-line data which is a pre-requisite to monitoring and controlling any technical process, it is fair to say that in the past it has been one of the more neglected elements of the process plant. Far more attention has been paid to, say, the theory and implementation of feedback control, and it is still frequently assumed that the control system should be able to compensate for measurement limitations. Under-instrumentation remains a common practice in the process industries.

As until recently the only continuous contact between the sensor and the control system has been the unidirectional flow of measurement data (usually based upon the 4-20mA standard), it is understandable that the sensor has been viewed as a simple signal generator, and that its data has been assumed to be 'correct' unless over-ridden by extraordinary action (e.g. by the operator). This single stream of information has been used in a variety of ways: for monitoring the process, for feedback control, and also for ensuring safety through the use of hardwired trips.

As far as the control system is concerned, it is irrelevant whether the data has been generated by a simple thermocouple or an on-line analyser, and in this paper the term 'sensor' will be used to describe any on-line system generating measurement information.

More recently the sensor has been receiving greater attention, due in part to the greater demands placed on all aspects of process operation:

- Competition has resulted in higher goals for product quality, and plant efficiency and availability. The crucial role of measurement in achieving these aims is becoming more widely acknowledged.
- Safety standards are constantly rising. Measurements are the primary means of identifying potentially hazardous circumstances. In addition, environmental protection legislation is being enacted and enforced in many countries (for example Bigg (1990) describes the impact of the Environmental Protection Bill in the U.K.). Breaking environmental constraints can lead to heavy fines or loss of license to operate, and sensor data provides both the first sign of warning to the operator and demonstration of compliance to the authorities. In enhancing either safety or environmental protection aspects of plant operation, it is not sufficient merely to add more sensors: measurement quality must be ensured.
- There is increasing recognition that, if measurement data is to be used on-line, particularly for feedback control, then safety cannot be ensured without some form of on-line fault detection.

Matching these demands are improvements in technology which offer enhanced functionality:

- New sensor technologies are becoming available, extending both the set of properties that can be measured and the environments in which they can be sampled. Developments range from micro-sensors to ever more sophisticated on-line process

analysers.

- Microprocessor-based instrumentation and digital communications in the field are having a profound effect on the capabilities of the sensor.
- The techniques of fault detection are maturing and are slowly finding more widespread application.

In this paper it is argued that the extended role and capabilities of the sensor can no longer be described adequately by a signal generator model. A new conceptual model is required which encompasses the extra capabilities and accommodates the additional demands placed upon the sensor.

There are three cornerstones to the model of the self-validating or SEVA sensor (Henry and Clarke 1991, Henry and Wood 1992) which is described in this paper: the use of fault detection techniques, the application of digital technology, and the use of uncertainty analysis provided by the science of metrology. Their interaction is shown in Fig. 1.

Metrology provides a means of analysing all factors affecting accuracy and calculating a single uncertainty value as a quality index for each measurement. When a sensor has an in-built microprocessor, such calculations can take place on-line. Similarly, fault detection techniques can be applied within the device to provide internal diagnostics. Finally, by applying uncertainty analysis to a faulty sensor, it is possible to assess the *impact* of a fault upon measurement quality, thereby retaining the availability of the measurement despite the presence of a sensor fault.

Each of the three key areas are examined in more detail in subsequent sections. There follows an analysis of why a new sensor model is required, and then a set of definitions describing the SEVA sensor. The final section describes a system which demonstrates the SEVA concepts, based upon a Coriolis mass flow meter.

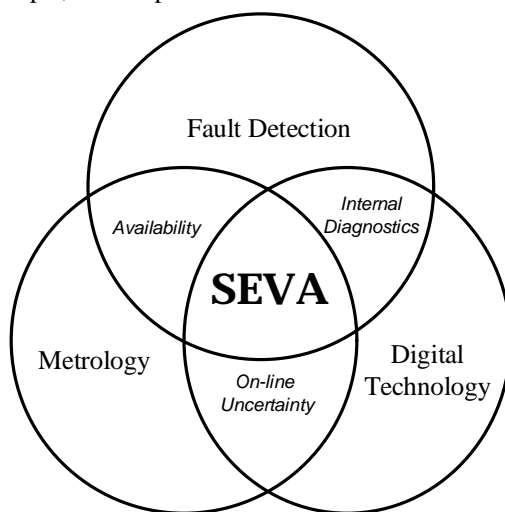


Fig 1. Cornerstones of SEVA Model

2. SENSOR AND PROCESS FAULT DETECTION

The first attempts to widen the naive signal generator model of the sensor have been made in the fault detection domain. This work was stimulated in part by the evident fact that acting on data from faulty sensors can lead to disaster. The resulting model of the sensor can be described as follows:

A sensor is designed to generate measurement data (i.e. to transmit an estimate of a process parameter). In practice the measurement is not a perfect representation of the process parameter: the effects of the sensor (including faults) and other process parameters or plant components (including those attributable to “faulty” behaviour) are also present in the measure-

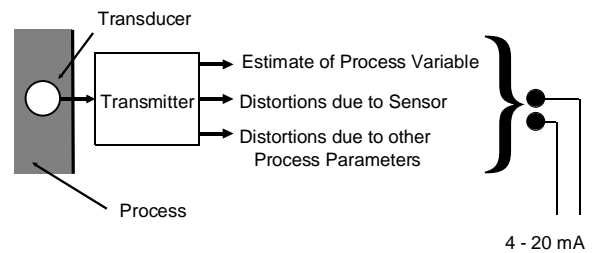


Fig. 2. Extended Model of the Analogue Sensor

ment signal (see Fig. 2).

With the limitations of analogue communication, measurement data has been the only source of information about the process *and* sensor. Ingeniously, workers in the fault detection domain have attempted to extract sensor, plant and process fault information from measurement data, but the inherent difficulties of distinguishing the separate strands from within a single analogue signal are universally acknowledged.

For example, Fig. 3 shows the output of a mass flow meter placed on an experimental rig. The rise in value which begins at $t = 10$ s is due to the onset of a sensor fault, while the true mass flow rate remains approxi-

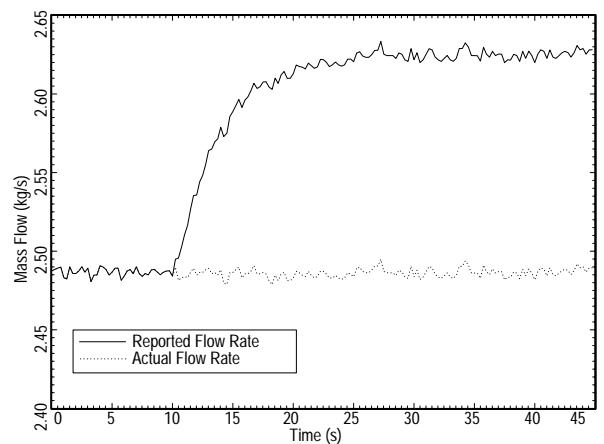


Fig. 3. Step change in measurement due to sensor fault

mately constant at about 2.5 kg/s. However, in the absence of other information, it might equally be inferred from the reported flowrate that the sensor is functioning properly and that the process mass flow has in fact risen. It is the task of fault detection to distinguish between plant or sensor faults and legitimate changes in a measurement signal.

This task is made more difficult by the impact of feedback control in the presence of a fault, as described in the next subsection.

2.1. Feedback Control

Feedback control adds to the complexity of fault detection in process plant by masking measurement deviations that might indicate a fault, and by making it difficult to distinguish between sensor, actuator and plant failure. For example, Fig. 4 shows the results of an experiment performed on a test rig using a mass flow meter and a valve. A PID controller is used to maintain a constant flow rate. Initially there are regular deviations about the set point of 1.0 due to slight stiction in the valve. At $t = 30\text{s}$ there is a disturbance, to which the controller responds, and after $t = 60\text{s}$ the reported mass flow rate resumes its characteristic behaviour. In fact the disturbance at $t = 30\text{s}$ is due to a sensor fault of the same type as is shown in Fig. 3. However, without the benefit of knowing the nature of the fault and the actual flow rate through the rig, it is possible to construct a number of hypotheses to explain the given pattern of measurement disturbance and control response, as follows:

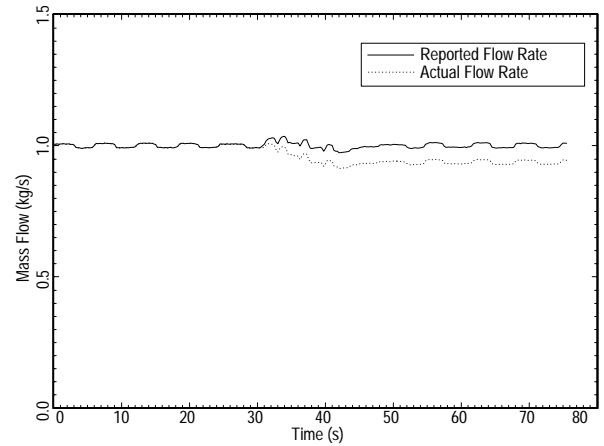
Process disturbance. A variation in the process (such as a rise in fluid density or pressure) increases the mass flow rate, and the controller responds by reducing the flow rate back towards the set point.

Valve fault. A fault in the valve (e.g. a deviation in the supplied air pressure) results in the valve sliding further open (without a corresponding increase in the control signal) thus causing a higher flow rate. The controller acts to correct the mass flow. A disparity has been introduced between the control signal and valve position.

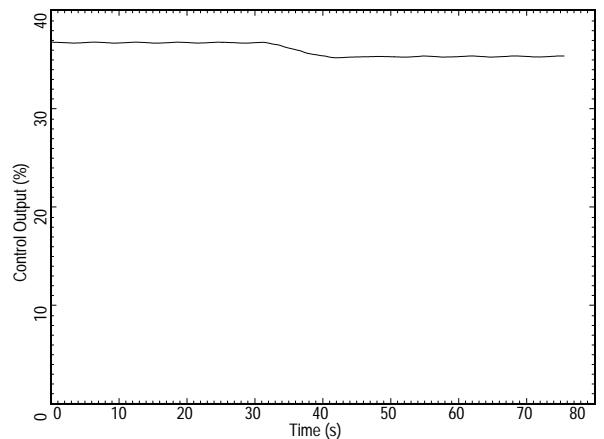
Sensor fault. A fault in the sensor introduces a positive bias into the measurement. The controller responds to the apparent increase in the flow rate, thereby causing the true mass flow to be maintained at a level below the set point. As stated above, this is what actually happened. Comparing the reported and actual flow rates in graph 3a, a 6% discrepancy has been introduced.

Several conclusions can be drawn from this example:

- Feedback control makes it difficult to distinguish between sensor, actuator and process faults.
- The effects of sensor faults are particularly serious, as feedback acts to suppress any measure-



(a)



(b)

Fig. 4. Sensor fault and feedback response: (a) mass flow (b) valve position

ment deviation. Thus after any initial disturbance the measurement may appear deceptively normal.

- There may be a limited time window for observing the gross effects of a sensor, actuator or process fault.
- There may be a 'knock on' effect. As the true flow rate stabilises after $t = 50\text{s}$ at 6% lower than the set point, further problems may occur downstream of the valve. Similarly, had the process disturbance hypothesis been correct, the rise in flow rate may have been caused by a fault occurring elsewhere in the plant upstream of the valve and sensor. In either case, after the initial disturbance the reported flow rate returns to the set point, making process 'trouble-shooting' difficult.

The consequences of a fault depend very much upon the particular application. At worst it could be disastrous, at best it will be detrimental to product quality, plant efficiency, or both. A vigilant operator may notice certain faults, particularly those affecting open loop measurements, such as in Fig. 3, where there may be a permanent discrepancy between the reported value and its usual level. However, faults occurring in closed

loops such as in Fig. 4 are far more difficult to notice, particularly as the operator is likely to be monitoring measurement data alone. Furthermore, the ‘knock on’ effect may have the result that when symptoms are observed, they may be at some distance from the source of the problem.

To achieve the goals of improved efficiency and safety, it is therefore highly desirable to have an automated system monitoring plant behaviour for faults.

2.2. Fault Detection, Identification and Accommodation

Given the difficulties of detecting faults in individual plant components, the most common strategy for developing an automated fault detection system is to pool data from several sensors, actuators and/or controllers. This approach is usually termed Fault Detection, Identification and Accommodation (FDIA) (Frank, 1990), as there are several stages involved in the procedure.

Firstly some model of the process is created. A software system is then implemented which monitors for inconsistencies between the model and on-line data (fault detection). There follows the task of identifying which plant component (whether sensor, actuator or other item) is responsible for the inconsistency. Finally, the system must respond to the presence of the fault in a safe and efficient manner (accommodation), although this is such a plant- and device-specific activity that it is scarcely more than mentioned in the literature, and in practice most papers are restricted to FDI. There are two principal approaches of the FDI type:

Analytical Redundancy (AR). This approach exploits the implicit redundancy in the static and dynamic relationships between measurements and actuator inputs, using a mathematical model (Fig. 5).

A vector of functions, called the *residual*, is generated, which is examined by the decision-making process, resulting in a fault decision. Each component of the residual has the property that (ideally) it is zero as long as plant operation remains normal, but some or all of these components become non-zero when an inconsistency occurs. Thus the residual components verify the continuing redundancy relationship between the measurement outputs and actuator inputs.

Fault detection entails no more than monitoring the magnitude of the residual vector. The identification of the source of the fault is more difficult, and a number of strategies have been devised for the design of the residual, each reflecting a different approach to fault identification. In the *parity space* approach (Patton and Chen, 1991), for example, the space formed by possible residual values is spanned by a set of *fault signature directions*. The residual is designed to move in a particular direction when the corresponding fault occurs.

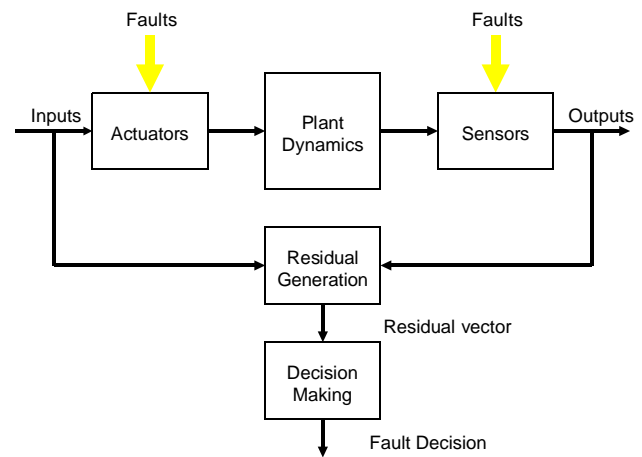


Fig. 5. FDI scheme using Analytical Redundancy (Patton and Chen 1991)

Detailed descriptions of the various strategies, implementation issues and the latest developments in analytical redundancy can be found in the survey papers of Frank (1991), Gertler (1991), Patton and Chen (1991) and Isermann (1991).

Knowledge-based Methods (KBM). In this approach, qualitative models of the process are built and manipulated using heuristic reasoning. Such models might include knowledge concerning operational conditions and associated fault modes, patterns of signal behaviour characteristic of particular faults, or historical fault statistics. Reflecting a less rigorous mathematical analysis of the plant, there is a considerable variety of techniques that can be employed. These include expert systems (Tzafestas, 1989 and 1991), neural nets (Hoskins *et al.*, 1988; Himmelblau, 1992), petri nets (Maßberg and Seifert, 1991) and fuzzy logic (Vachkov and Matsuyama, 1992).

Measurement Aberration Detection (MAD). A third approach is to examine the output of the single sensor for indications of faults. The techniques are described by Yung and Clarke (1989) although in their paper they use the term ‘sensor validation’. When applying MAD it is assumed that the ‘true’ measurement signal has certain time and/or frequency domain properties and that ‘faults’ (defined as deviations from normal behaviour) may occur randomly at any time. For example, Fig. 6 shows the frequency spectrum of a sensor output, in which the lower frequencies are considered to be ‘measurement’ and the high frequencies ‘sensor’ related. The presence of a high frequency harmonic may be interpreted as implying a sensor fault. Faults are classified in terms of how they change the behaviour of the signal (e.g. bias, noisy, saturation), rather than suggesting any underlying, sensor-specific cause for the fault (e.g. corroded transducer).

MAD techniques have also been used to detect plant faults. As shown in Fig. 2, the analogue measurement carries a component that is ‘process related’, i.e. it is

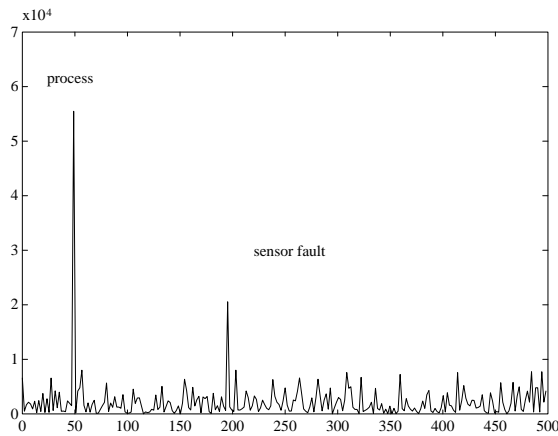


Fig. 6. MAD scheme using frequency spectrum

affected by aspects of the process (or plant) other than the measurand (or sensor). In practice such data has been used by operators who, by observing minor fluctuations or characteristic patterns in a measurement signal, are able to gauge the 'health' of plant components or the process near to the transducer. In effect the sensor is providing an auxiliary measurement related to health. Of course such practice is far from systematic or widespread, depending upon the good fortune of useful features being recognizable within the measurement signal. However, MAD-type schemes have been developed to automate the detection of plant-related phenomena. For example, in lab trials, worn or broken impellor blades on a nearby pump have been detected using measurement data from an orifice-plate flow meter (KB MUSICA 1992). Such work has strong parallels with acoustic monitoring, which can be used to provide health data on plant and/or process (Belchamber and Collins 1990). However, in the latter case signals are observed for their health content only, and a measurement signal is not used as a 'carrier'.

2.3. Implementation Issues

When selecting one or more techniques to use in a particular application, a number of criteria must be borne in mind, such as the consequences of an undiagnosed fault, and available expertise and resources. However, it is the degree of detailed knowledge of the plant which is the major determinant of which techniques can be applied.

Analytical Redundancy (AR) is theoretically the most sensitive method of fault detection, but its effectiveness is critically dependent upon the availability and quality of a mathematical model of the process. In mechanical systems such as aircraft, AR has been successfully applied, but in chemical plant where only poor or imprecise models are available, AR remains difficult. The creation of AR schemes which are robust to plant disturbances and modelling errors is a major research topic (Patton *et al.*, 1992). In addition, Frank (1990) points out that the quality of fault isolation is heavily dependent upon the number of measurements that are available in the plant. Even when AR is appli-

cable, Frank (1990) suggests that AR and KBM are complementary and can (and should) be combined in a single software package.

As there are a variety of techniques described under the general heading of Knowledge Based Methods, the key task is to match available knowledge to the appropriate technique. Himmelblau (1992) outlines the strengths and weaknesses of various methods. Lacking the rigour of AR it is possible to express far more knowledge of the process than can be described in a mathematical model; conversely, there is no guarantee that the reasoning employed will be comprehensive or fault-free. There is also the problem of capturing and encoding 'process knowledge' using the selected paradigm(s). Tzafestas (1991) describes some of the difficulties encountered in applying 'first generation' expert systems to process fault diagnosis, and points to improvements likely to feature in future systems.

MAD has advantages over a pure FDI scheme in that there is no need to 'identify' which of a number of sensors has malfunctioned - its own monitor will flag the guilty party. In addition MAD can be applied selectively to sensitive or valuable measurements, and so can be much more cost effective. However, in practice, detection usually requires a considerable amount of process- and device-specific tuning, and there remains the problem of legitimate changes in the statistical properties of device or process. In addition, as there is no attempt to correlate data from multiple sensors, MAD cannot detect faults which are manifest as inconsistencies between well-behaved signals. Thus in comprehensive implementations MAD can be used as a form of diagnostic prefiltering (spotting gross deviations in individual sensors) to an FDI scheme (see Fig. 7). Frank (1990), Patton *et al* (1989) and Isermann (1991) discuss in more detail the architecture of fault detection schemes which employ a combination of methods.

Irrespective of the techniques applied, for any FDI scheme it is in general difficult to develop a model that

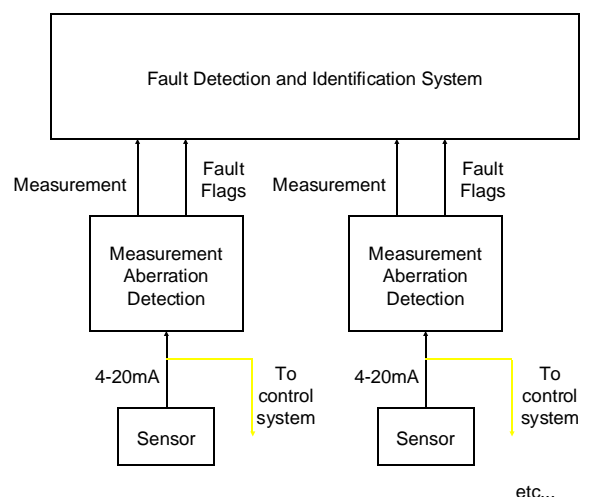


Fig. 7. MAD combined with FDI techniques

is robust to process disturbances and modifications to the plant, and to test thoroughly the fault detection scheme when it is rarely feasible to induce actual faults into the plant for 'practice'.

There are also economic difficulties in transferring FDI techniques to real applications. As sensor data is pooled to detect inconsistencies, the smallest viable application must use data from a number of sensors, and therefore a model of some unique piece of plant is normally required. A large investment of modelling expertise is needed for each plant, as there is little carry-over of effort between applications, and key personnel must be available to maintain and update the system as the plant changes. Potential benefits must be weighed against this investment, losses in production while the system is installed, and the perceived risk that the system may not be successful. When the total investment is large, it is difficult to sanction such a project, but without previous successful examples it is hard to quantify potential benefits. It is not easy when dealing with cost-conscious management to argue for funding a project on the grounds that theoretically it can prevent a major incident once every ten years.

2.4. Integrating Fault Detection and Plant Operation

Given these difficulties, it is perhaps not surprising that despite the desirability of fault detection, and the maturity of its techniques, it has yet to find widespread application within the process industries (Halme and Selkainaho 1991). A reasonable conclusion might be that either the costs must be lowered, or the perceived benefits increased, or both.

Fault detection is expensive because:

- hundreds of hours of expert labour are required for each application;
- a software system must be constructed in parallel with, and interfacing to, the control system.

The perceived benefits are insufficient because:

- Fault detection is seen as preventative rather than as enhancing production;
- The performance of a fault detection system is in general difficult to guarantee.

Many of these problems are due to the fact that fault detection is viewed, and described by its practitioners, as an activity in isolation. This is particularly true of AR which is described in purely mathematical terms. It is up to the implementor to map the reality of the plant into a mathematical form, code the diagnostic procedures and map the results of the on-line significance tests back into the plant context. AR offers at best an indication of some inconsistency stemming from a particular plant component. Any further reasoning

about the true nature of the fault, its potential consequences, and action that must be taken, cannot be described or manipulated within the rigorous mathematical framework of AR. Of course there is little practical benefit in performing fault detection unless these additional aspects are addressed.

If the extra tasks of fault interpretation and accommodation are to be automated, then knowledge-based techniques are better suited to the task, providing another reason for adopting a mixed architecture in an FDI system. In any case, much of the effort of creating an on-line FDI system will be taken up not so much with fault detection as such, but with knowledge acquisition, interfacing to data acquisition and control systems, creating a user interface, real-time graphics, an advisory system, testing, and so on.

One reason for the use of expert system shells in diagnostic systems is precisely because many of these facilities are provided within the package, thus reducing the need for recoding between applications. More importantly, however, within such shells fault detection can be combined with other tasks such as production scheduling, which can be more easily shown to be making a positive impact on the balance sheet (Rowan 1992). Typically, payback is demonstrated by improved first pass yield and waste reduction. Similarly, because of the broader capabilities of such systems it is economically viable to keep in-house expertise of a particular shell.

Thus by seeing fault detection as just one facet of the more general goal of enhancing safety and efficiency, the add-on costs are reduced and the potential benefits are seen in terms of extra productivity rather than mere prevention. Looking to the future, Milne (1991) argues that integration should be taken to the point of largely eliminating the use of 'experts' to encode diagnostic knowledge. Rather, diagnostic information should be drawn directly from existing CAD and process data describing the state of the plant and its organisation: thus the plant is designed to be diagnosed, and is diagnosed from its design.

2.5. The SEVA Approach

In the SEVA approach, the reliability and ease of integration of fault detection is increased, while the costs are reduced, by 'factorising out' the need for sensor or actuator validation at the system level. Instead, these tasks are performed within the devices themselves, as extra functionality provided by the manufacturer: sensors are designed to be self-diagnosing. Within each sensor, fault detection is combined with measurement validation, and fault information is generated in a standard format which can be readily used by the control system. This approach is made possible by the application of digital technology as described in the next section.

3. DIGITAL TECHNOLOGY

There are two principal areas in which digital technology is having a profound impact on process instrumentation. These are the use of microprocessors within sensors and actuators, and the development of digital communications.

3.1. Digital sensors

The recent development of cheap and powerful microprocessors has lead to a new generation of digital sensors and actuators with considerable in-built processing power. The advantages conferred on instrumentation by the introduction of microprocessors include the following (de Sa 1988):

- Improved accuracy - typically doubled;
- Corrections for sensor non-linearity and temperature effects;
- Output in engineering units;
- Configurability (e.g. output units, ranging);
- Internal storage of location and servicing details;
- Internal diagnostics.

For example, Table 1 compares the characteristics of typical analogue and digital differential pressure sensors:

Table 1 Characteristics of Analogue and Intelligent D.P. Sensors
(% is of full span)

Characteristic	Analogue	Digital
Accuracy	$\pm 0.25\%$	$\pm 0.1\%$
Repeatability	$\pm 0.05\%$	$\pm 0.02\%$
Hysteresis	$\pm 0.1\%$	$\pm 0.02\%$
Drift (6 months)	$\pm 0.15\%$	$\pm 0.1\%$
Ambient Temperature Effect	$\pm 0.75\%$	$\pm 0.1\%$
Static Pressure Effect	$\pm 1.0\%$	$\pm 0.1\%$

Blickley (1991) describes not only similar developments with actuators, but also some of the environmental legislation which has required manufacturers to use digital technology to enhance and validate the performance of their devices.

3.2. The Coriolis Mass Flow Meter

Several of the examples described in this paper are based upon work performed using a commercial Coriolis mass flow meter, which is a typical 'intelligent' instrument. This microprocessor-based sensor is used in the petrochemical, chemical and food industries. It measures three process parameters: mass flow, density

and temperature, and can additionally calculate volumetric and total flow. The principle of Coriolis acceleration is used to measure the mass flow of fluids directly, without the need for external pressure, temperature or specific gravity measurements.

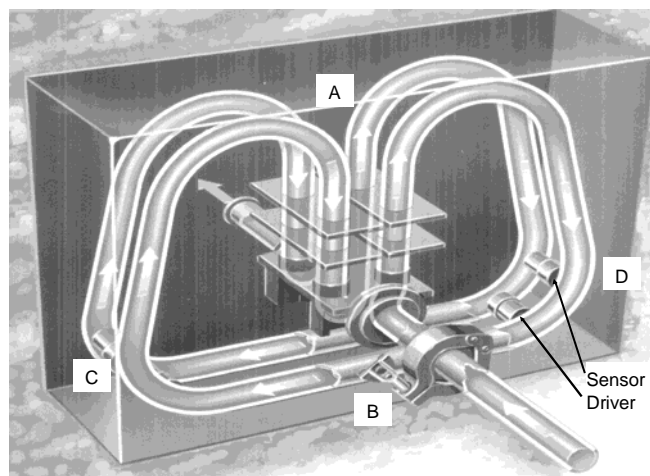


Fig. 8. Coriolis Meter Flowtube

The meter consists of two separate units connected by 12-core cable: the flowtube, which is placed in line with the process piping, and the transmitter, which houses the drive circuitry and the microprocessor which performs the measurement calculations. A simplified diagram of the flowtube is shown in Fig. 8. In practice the flowtube is encased in a protective covering.

Strictly speaking there are two separate measurement subsystems: the temperature subsystem consisting of an RTD buried in the central mass or manifold, and the mass flow/density subsystem which consists of the two vibrating, serially connected, large bore tube loops positioned in parallel. The tube is anchored to the rigid central mass. The process fluid is pumped through the central mass, through one tube loop back into the central mass, and then through the second tube loop such that the fluid flow is parallel to that in the first loop.

Two electromagnetic drivers and two electromagnetic velocity sensors bridge both loops at opposite extremities, equidistant from the centre. The former drive the tubes with two 180° out of phase sinusoidal waveforms at the damped resonant frequency of the vibrating system, so that both loops rotate about AB. A feedback system in the transmitter modifies the drive currents in order to maintain a constant sensor voltage amplitude.

Coriolis forces act on the pipe sections parallel with CD, causing tube deflections. The instantaneous velocity between the two loops is measured by the sensors adjacent to the drivers. The measured phase angle between the sensor voltages is proportional to the mass flow rate of the fluid in the flowtube, while the density of the process fluid can be obtained from the frequency of oscillation of the system.

3.3. Sensor Self-Diagnosis (SSD)

An increasingly common feature of digital sensors is a self-diagnostic capability. In-built processing power can exploit all the device-specific knowledge available to the instrument designer. While MAD detects faults based solely on observing the behaviour of the measurement signal from outside of the sensor, Sensor Self-Diagnosis (SSD) takes place within the sensor, where there are typically many signals other than process measurements, and indeed other sources of knowledge, which can be used to monitor the health of the device. These other sources may include device-specific tests that typically stimulate the transducer and obtain diagnostic information from the response.

For example, within the Coriolis meter there are many parameters of potential use for fault detection, including internal signals, the amplitudes of harmonics, and physical and electrical properties. Only the measurement signals themselves are available outside the device, and these are far from ideal for detecting sensor faults precisely because they are designed to convey information about the process, not the sensor. The graphs in Fig. 9 illustrate this point.

A certain condition can arise within a process which can have a severe but temporary impact on the measurement capability of a Coriolis meter. Fig. 9 show the results in lab trials of this process-induced fault upon the mass flow rate and L , a physical property calculated using the output of a device-specific hardware test. The L test is designed to show proper mechanical working of the meter, and is relatively immune to most changes in process conditions, including the mass flow. These minor process effects are easily modelled, resulting in a narrow range of permitted values of L shown by the dashed region on the graph in Fig. 9a. However, the test is extremely sensitive to even small levels of the fault, which can thus be detected easily.

The effect of the fault on the mass flow rate reported by the sensor can be seen in Fig. 9b. The true mass flow rate in the rig was constant throughout the series of experiments, and so we may deduce that the fault causes a gradual lowering of output value which becomes more severe as the level of fault increases. However, in a real plant, this drifting behaviour might easily be mimicked or masked by genuine changes in the process mass flow. If the measurement was a controlled variable, then any drop in sensor output value due to the fault would quickly be masked by control action.

A measurement signal is designed to convey information about the process, and therefore any method of sensor fault detection based solely on measurement information is inherently process-dependent. Application-specific schemes must be devised, involving process modelling which is expensive and which cannot guarantee results. However, assuming in-built processing power to perform self-diagnostics, it is reasonable

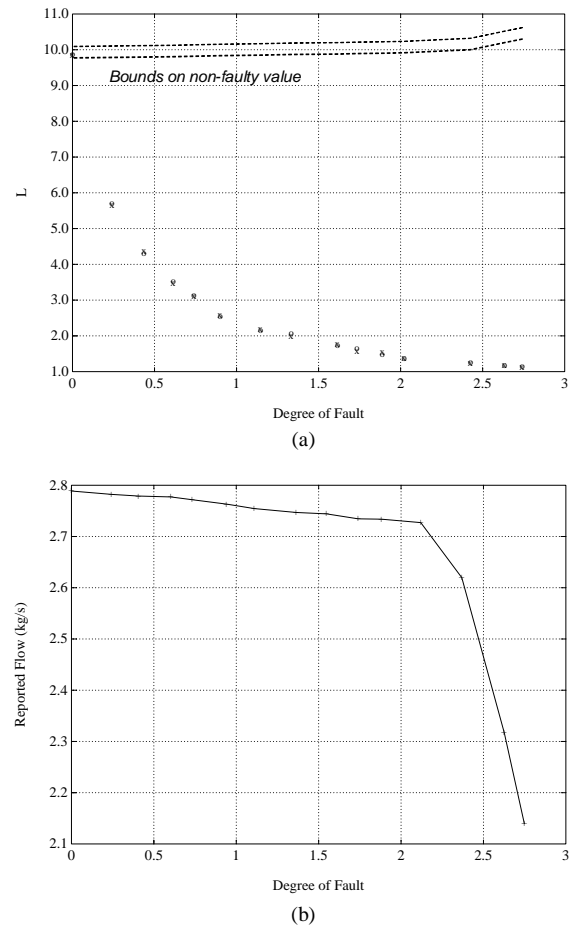


Fig. 9. Effect of fault on (a) L (b) mass flow rate

to assert that additional signals (such as L) can be identified (or be specifically designed and generated) within the sensor that can give indications of faults *irrespective* of the behaviour of the measurand. Indeed L is an esoteric example; trivial threshold logic on signal levels within the sensor can detect faults which may be difficult or impossible to trace within the measurement, as is demonstrated later. If tests or signals such as these can be found or designed, then the same fault detection algorithms can be used in all identical sensors, without regard to the processes in which they are installed.

It would be naive to suggest that comprehensive internal diagnostics can be generated with equal ease over the entire range of sensing devices, from the thermocouple to the process analyser. However, it is surely the case that in every sensor simple checks can be implemented in order to reduce the possibility of an undetected fault affecting an on-line measurement.

Economically this approach is appealing. The cost of developing internal diagnostics for a single product can be spread over the thousands of devices sold to industry. Indeed as many of the requirements (test rigs, skilled personnel) are retained by instrument vendors in order to develop new products, it would appear that the most cost-effective way to incorporate diagnostics within a sensor is as a natural extension of the design and development process. It is economically viable to

emulate all known fault modes, even if the sensors under test are destroyed, because the unit cost is normally small. Furthermore customer feedback can provide an excellent means of reviewing and improving fault detection strategies, particularly if sensors with unusual or unexpected faults are returned to the vendor for detailed analysis.

A self-diagnosing capability is of course just one of the enhancements achievable in a microprocessor-based instrument, but it is becoming increasingly common in new instruments. This trend is set to continue. The fault detection community has, however, been slow to recognise and adapt to these developments.

SSD should be seen as a more powerful version of MAD. It is more reliable because it is based upon manufacturer's knowledge and takes place within the device where faults can be observed in a variety of ways; with MAD one must merely hope that faults will manifest themselves in the form of statistical anomalies in the measurement signal. In theory self-diagnosing devices should be a welcome advance, as all the hard work has been done by the manufacturer.

The problem lies in integrating this local fault detection capability into a more general scheme. The only sensor data required by a monolithic FDI system is the analogue measurements, which are already provided by the control system (see Fig. 7). The transfer of diagnostic information between the MAD monitor for each sensor and the central fault detection system can therefore be specified and implemented by the FDI system designers themselves. While this is expensive, it is at least feasible. When diagnostics are performed within the sensor itself, how is fault information to be conveyed to higher levels? The solution to this problem is offered by the second major development within process instrumentation.

3.4. Digital Communications

In-built processing power facilitates separate estimation of the measurand and of the sensor fault state. However, these data cannot be transmitted separately over a single 4-20mA analogue channel. Operator inspection of the transmitter, perhaps via a hand-held terminal, is barely adequate. The transmission of multiple parameters over a single pair of wires, for routine use by automatic control/alarm/safety systems requires a digital communication link (see Fig. 10).

Digital communications offer many other advantages such as two-way data/command flow and remote re-configuration of devices. One particular benefit is the reduction of measurement error due to data transmission. Figure 11 illustrates the progression from analogue to digital processing and transmission of sensor data. At step (b), the incorporation of a microprocessor (labelled uP) results in overall improvements in accuracy, as described in the previous sub-section. At step

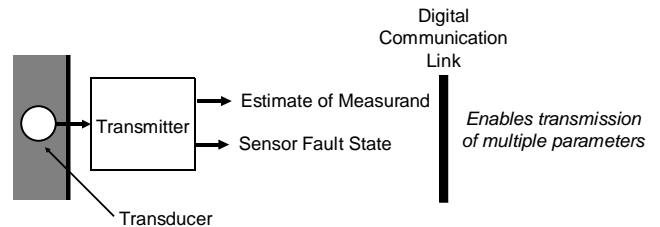


Fig. 10. The intelligent sensor using digital communications

(c), direct digital communication replaces the re-conversion of data into an analogue signal. This eliminates the errors (noise, non-linearities, discretisation) introduced by the conversion process at each end. In particular, the precision limit imposed by the maximum resolution of the D/A (and A/D at the other end) is removed.

A number of proprietary digital communication protocols are currently available, many operating over existing 4-20mA links, but digital communications will become more widely adopted by industry with the completion of the Fieldbus standard (Wood 1991). All major suppliers and several leading users are co-operating to complete and demonstrate the IEC standard as soon as possible.

The vendor-independent Fieldbus standard provides a platform for rich message-passing between intelligent components of a plant fault detection scheme. By defining suitable data and command standards, the full potential of intelligent devices can be realised. Sensors and other devices can send fault information along with measurement data, and higher-level fault-detection software can return commands to reconfigure, perform additional self-checking, re-calibrate, or to log calls to maintenance (see Fig. 12).

However, the Fieldbus committees have given little consideration to the standardisation of fault information messages. Existing fault detection schemes using digital communications typically employ a single bit

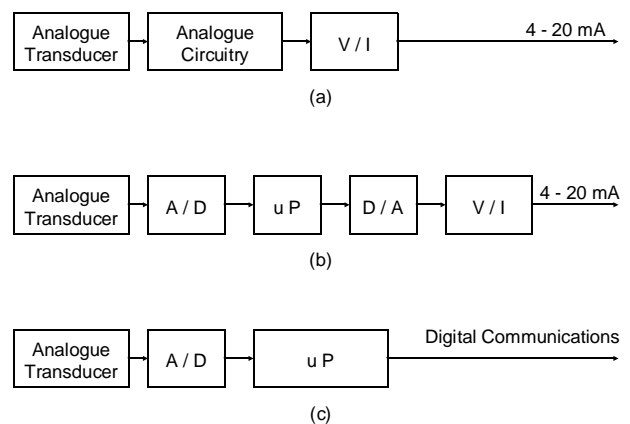


Fig. 11. Data transformations within the sensor:

- a) purely analogue
- b) microprocessor-based with analogue communications
- c) microprocessor-based with digital communications

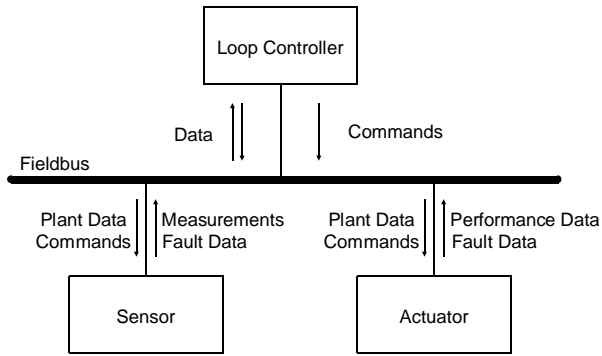


Fig. 12. Control using Fieldbus and intelligent devices

to indicate whether a measurement datum is 'valid' or not. Alternatively, the device-specific error code (e.g. Fault 43 - detached transducer), generated by a self-diagnosing sensor, is passed to the higher level system.

The chief objection to using a single bit is that the sensor must judge, on behalf of the control system, whether a measurement is acceptable. The only choices are yes or no. However, a fault may have only a marginal effect on the measurement. A yet more difficult situation occurs when a number of measurements are combined in some way to generate a result. If one of the input measurements is flagged as bad, should the result be?

For the purposes of monitoring and controlling the process, a 'bad' measurement cannot be used, in which case there is no point in generating any measurement data at all while the fault persists. Experience suggests, however, that operators at least would rather have 'bad' data than none at all. On the other hand, if a bad measurement is used anyway, there is little point having the fault flag. It is therefore likely that a one-bit scheme will cause an unwarranted reduction in the availability of the measurement. Some scale of measurement degradation might be preferable.

If error codes are employed, then the problem of judging the utility of the data is passed upwards to the control system: it is told what is wrong with the sensor, and it must interpret the consequences of this particular fault for the viability of continued operation. The major problem here is the device-specific nature of such error codes. The control system must be able to respond to any of tens of fault modes for each sensor type in the plant. This is unlikely to please the control engineer, for whom it raises the bleak prospect of updating a massive database of individual device characteristics, or substantially reconfiguring the system software, whenever a sensor is installed or replaced.

Unless some device-independent standard for describing sensor faults is agreed and adopted, the bottleneck of the 4-20mA analogue channel could be replaced by a cacophony of device-specific fault codes and commands.

A more philosophical issue is also raised by the potential of digital systems, concerning physical and functional identity. The new capabilities of localised computational power and digital communications mean that the computational functions taking place in physical objects are no longer fixed. Theoretically, almost any function can be performed anywhere in the plant, the relevant data being sent over the Fieldbus. For example, Fig. 12 shows the components of a typical loop, consisting of sensor, actuator and controller, communicating via a Fieldbus. The only functions which cannot be located anywhere are those associated with physical components: transducer data must be generated from within the sensor, and the positioning of the valve (or other plant component) must take place within the actuator. The location of any data processing is theoretically arbitrary. We could, for example, perform measurement calculation within the sensor, but carry out sensor fault detection within the loop controller (by sending all internal signals to the controller over the Fieldbus). Similarly we could perform the control algorithm within the actuator and dispose of the loop controller completely. In such circumstances, what do we mean by sensor, actuator or controller?

In practice, certain constraints remain. The Fieldbus has a limited bandwidth. The self-diagnosing sensor processes a large quantity of data which is summarised in the measurement and diagnostic parameters available to other system components. Processing this data outside of the sensor would place a heavy load on the Fieldbus. In addition a large amount of contextual knowledge will be required to interpret the data, which will include proprietary knowledge that the sensor manufacturer is unlikely to make freely available. There is therefore no obvious advantage, and several apparent problems, with exporting sensor data processing to other network components.

The possibility of performing control calculations within the valve is more attractive. A simple PID law could easily be incorporated within a valve thus eliminating the need for a loop controller. In applications where more sophisticated control is required, a controller which contains a large amount of proprietary knowledge, perhaps incorporating such features as predictive or adaptive control, would still be viable as a stand-alone device.

3.5. The SEVA Approach

The application of digital sensors and communications offer enormous advances in sensor functionality. Potential enhancements are so powerful, however, that they threaten to distort the current understanding of the roles and functions of devices, thereby making system integration far more difficult. Fieldbus notwithstanding, there is as yet no description of the generic sensor (or actuator) that encompasses the extra capabilities offered by digital technology. The SEVA approach offers such a description.

One aspect required of any generic description of the sensor is a standard, device-independent means of communicating sensor faults. Indeed a single sensor fault must be described in a number of ways: a fault code or detailed description is required for the purposes of maintenance, but for monitoring and controlling the process, the important issue is the impact a fault has upon the quality of the measurement. The most appropriate method of evaluating this impact is provided by the techniques of uncertainty analysis, as described in the next section.

4. METROLOGY AND UNCERTAINTY ANALYSIS

In the discussion thus far, the emphasis has been upon fault detection and improving sensor performance. As yet there has been no consideration of measurement validation, that is assessing the ‘quality’ of measurement data. The science of metrology covers the domain of measurement and its accuracy, and the principal tool for examining error is uncertainty analysis.

The concept of Uncertainty was originally defined by Kline and McClintock (1953). Every measurement has an associated error, which is of course unknown. However a single uncertainty number is often needed to express a reasonable limit on that error (ANSI 1985).

For any observed measurement M , the uncertainty in M , w_M , can be defined as follows: we assert that

$$M_{true} \in [M - w_M, M + w_M] \quad (1)$$

with a certain level of confidence (typically 95%). This uncertainty is readily expressed in a relative form as a proportion of the measurement (i.e. $\frac{w_M}{M}$).

A propagation rule exists for obtaining the uncertainties of arbitrary functions of primary measurements (Kline and McClintock 1953, ANSI 1983, 1985). For example, for an arbitrary function R of variables x , y , and z ,

$$R = R(x, y, z) \quad (2)$$

the uncertainty of R is given by

$$\begin{aligned} \left[\frac{w_R}{R} \right]^2 &= \left[\frac{\partial R}{\partial x} \right]^2 \left[\frac{w_x}{R} \right]^2 + \left[\frac{\partial R}{\partial y} \right]^2 \left[\frac{w_y}{R} \right]^2 \\ &+ \left[\frac{\partial R}{\partial z} \right]^2 \left[\frac{w_z}{R} \right]^2 \end{aligned} \quad (3)$$

This sum of squares form is derived from the Taylor series, and assumes the independence of x , y , and z , that their relative uncertainties are ‘small’, and that all uncertainties are expressed at the same probability level. For the remainder of this paper all uncertainties will be quoted at 95% probability.

One of the uses of the uncertainty propagation formula is to reveal particular circumstances which can result in a higher or lower than expected level of uncertainty. For example (Kline, 1985), if R is calculated from x and y using the equation

$$R = x - y \quad (4)$$

The uncertainty in R , w_R , is given by

$$\frac{w_R}{R} = \left[\left(\frac{x}{x - y} \frac{w_x}{x} \right)^2 + \left(\frac{y}{x - y} \frac{w_y}{y} \right)^2 \right]^{1/2} \quad (5)$$

Suppose $x = 1.00$ and $y = 0.98$, and the uncertainty in both x and y is 1%; then the uncertainty in R is:

$$\begin{aligned} \frac{w_R}{R} &= \left[\left(\frac{1}{0.02} 0.01 \right)^2 + \left(\frac{0.98}{0.02} 0.01 \right)^2 \right]^{1/2} \\ &= 0.700 = 70\% \end{aligned} \quad (6)$$

Consider now a result R found from a variable z through the formula

$$R = \left(\frac{1}{1 + z} \right)^{1/2} \quad (7)$$

The corresponding uncertainty is given by

$$\frac{w_R}{R} = \frac{w_z}{2(1 + z)} \quad (8)$$

If $z = 0.1$ and the uncertainty in z is 20%, then the uncertainty in R is only 0.91%.

On reflection these results should not be surprising. In the first example two quantities of similar magnitude are being subtracted: clearly this will increase the relative error in the result. In the second, although z has a large uncertainty, its influence on R is relatively small. Of course for different values of x , y , and z the impact of their uncertainties will vary. The usefulness of uncertainty analysis is that it can quantify these effects.

In the original paper by Kline and McClintock (1953), uncertainty was conceived in the context of the single sample experiment. Since then uncertainty analysis has been widely applied, and has achieved the status of an international standard for calibration and for the reporting of experimental results (ISA 1980; ANSI 1983; ANSI 1985). Indeed, certain ASME journals require an uncertainty analysis to be submitted with all experimental data (Dean 1975).

In the domains of calibration and experiment, however, repeated measurements are normally possible and usually necessary. As a result, the standard documents usually partition uncertainty into *bias* and *precision*. Bias is a systematic error assumed constant for a given set of readings under the experimental conditions. Precision error is a random component which is assumed to give a sequence of independent errors in the readings, and subject to statistical estimation via repeated measurements. Two uncertainty terms estimating these components are usually combined to present a single overall uncertainty value associated with a particular result or calibration.

It must be stressed that it is not possible to associate uncertainty with a rigorous confidence interval because bias is often based on judgement and/or past experience, while precision is calculated using statistics. Therefore no function of these two numbers will have a rigorous statistical basis (ANSI 1985).

4.1. Sources of Uncertainty

Measurement error sources can be categorized into three groups (ANSI 1985):

- calibration errors;
- data acquisition errors;
- data reduction errors.

Calibration Error The calibration process exchanges the large bias error of an uncalibrated or poorly calibrated instrument for the smaller combination of the bias error of the standard instrument and the precision error of the comparison. This exchange of errors is fundamental to all calibration processes. Clearly the measurement technology employed in the instrument undergoing calibration will have a fundamental influence upon the magnitude of the calibration error, as well as any subsequent calibration drift.

Calibration is also used to provide *traceability* to known reference standards, through a hierarchy of standards laboratories (see Fig. 13). Each calibration in the hierarchy constitutes an error source, and so uncertainty increases down the chain of calibration, but at the same time traceability is established from the standard to the individual instrument.

Data acquisition errors. Inevitably the sampling process itself has a fundamental influence on the uncertainty. Other error sources which may be present and included in this category are environmental effects (e.g. changes in ambient temperature), and installation effects.

Data reduction errors. Calculations are performed on raw data to produce measurement data in engineering units. Other parameters may be calculated using these measurement data. Errors in these calculations stem from

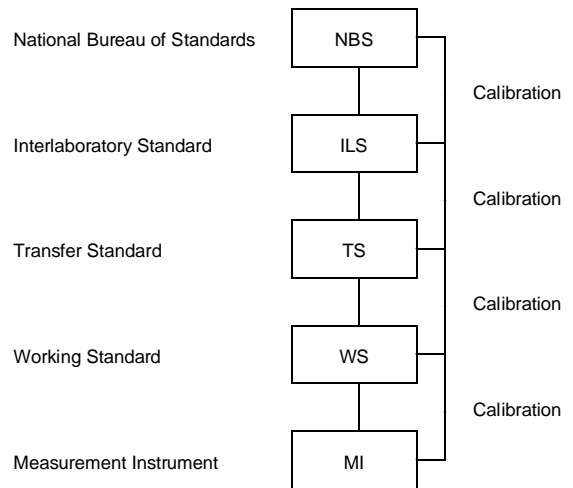


Fig. 13. Typical calibration hierarchy (from ANSI 1983)

the use of curve fits, correlations and/or approximate numerical methods, as well as loss of resolution (as demonstrated in the first worked example above).

4.2. Uses of Uncertainty Analysis

Other than its primary role in the calibration of instrumentation, uncertainty analysis has been put to good effect in the design and analysis of experiments. Kline (1985) describes how uncertainty can be used to particular advantage in the design and operation of experimental rigs, by:

- minimising instrument cost for a given output uncertainty;
- identifying which instruments or procedures control the overall uncertainty, thus focusing attention on points where care is particularly important;
- designing instrumentation to minimise uncertainty;
- checking against unknowingly entering a region of the data hyperspace where there are large uncertainties due to data reduction effects.

Moffat (1985) describes how uncertainty can be used to perform simple static analytical redundancy checks: two estimates of the same parameter 'agree' when they do not differ by more than the root-sum-square of their uncertainties. This principle can be extended to any number of estimates. ANSI (1985) provides a method for weighting multiple estimates of a parameter in order to obtain a 'best estimate'.

4.3. Uncertainty and Process Instrumentation

Uncertainty has in the past been used as a static analysis, providing a single uncertainty estimate for an instrument. In such a calculation, 'average' or 'typical'

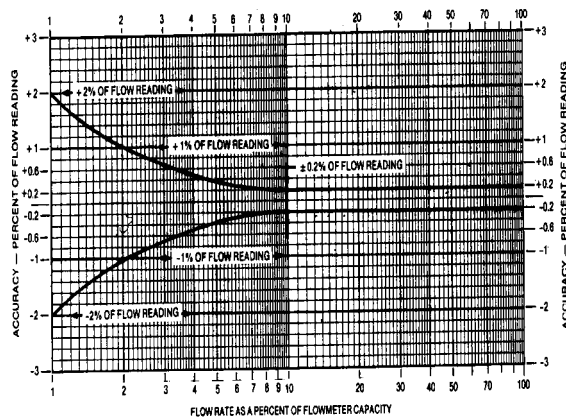


Fig. 14. Typical accuracy graph supplied with a commercial instrument

values are substituted for parameters which vary. However, for process instrumentation there can be considerable variation in uncertainty over the ranges of the measurand, process temperature and pressure, and other influencing factors. Instrument manufacturers often provide 'accuracy' charts (see for example Fig. 14) showing how 'accuracy' varies with the measurand, but these are of limited use. They usually cite performance data under reference conditions in a laboratory, and in doing so underestimate the true uncertainties of their instruments under process conditions. This is partly due to the difficulties of gauging and representing the impact of multi-dimensional process conditions in graphical form, but also perhaps due to the desire to cite favourable figures in order to gain market advantage (a practice known as 'spec-manship'). For example, Cork (1989) discusses the shortcomings of manufacturer's accuracy data for Coriolis meters.

Where it is particularly important for the uncertainty of a measurement to be kept within specified limits (for example in legal metrology applications), the only way to do this at present is to have an instrument certified by a registered testing laboratory. In this process the uncertainty of the instrument is assessed on a test rig while varying the measurand and process conditions. The certification of the instrument only remains valid as long as the process conditions remain within the limits under which the instrument was tested. This is both costly and restrictive.

However, the demand for measurement quality assurance is increasing, stimulated by environmental legislation as well as the drive for product quality assurance, and a means of assessing uncertainty on-line under all process conditions would be a major step forward.

4.4. The SEVA approach

The SEVA approach envisages each instrument providing an on-line assessment of its own uncertainty using in-built processing power and the experience and expertise of the manufacturer. Factors involved in this calculation include process noise, the A-D or V-F

converter, calibration (including time since last calibration to estimate drift) and the measurement calculation, as well as the presence of sensor faults. This approach, like sensor self-diagnosis, is now both technically and economically feasible.

Returning to the concept of the single sample (Kline and McClintock 1953), only a single measure of uncertainty is proposed, as opposed to separate calculation of bias and precision, for in the reality of on-line industrial measurements the process variable and error sequence both vary in a complex time-dependent manner. Outside of a calibration procedure, measurements cannot be repeated, and therefore the concepts of bias and precision have limited merit.

In order to realise this approach it is necessary to combine uncertainty analysis with fault detection within the sensor. This is, however, a natural extension of uncertainty analysis, as faults are simply another source of error. The manufacturer must expand the measurement model of the sensor to include the impact of each potential fault upon the measurement value and its uncertainty. How this is to be achieved is outlined later.

It may be argued that if 'spec-manship' is common amongst competing manufacturers now, then the uncertainty values generated by self-validating devices are likely to be somewhat optimistic. The counter argument is that the SEVA concept provides a framework in which uncertainty data can actually be used. Spec-manship remains possible when it is expensive to verify accuracy claims directly in the laboratory and difficult to spot greater than claimed uncertainty in an installed instrument. Indeed, under such circumstances, asserted accuracy has only limited use as a criterion for selecting instrumentation.

If, on the other hand, as a result of using on-line uncertainty data, worse than claimed performance from a sensor is demonstrated, then the manufacturer is likely to suffer dire consequences. Rather than risk such an eventuality, manufacturers are more likely to redouble their efforts to improve the actual performance of their devices, confident that improvements can be demonstrated and therefore rewarded.

Parallels can be drawn between potential uses of uncertainty data in process plant and existing uses of uncertainty analysis in experimental rigs as described by Kline (1985), and summarised above. Potential uses range from design through to control and maintenance of the plant:

- During plant design a maximum uncertainty is specified for each measurement in the plant. This figure will take into account how each measurement is to be used: whether for monitoring of product quality, demonstration of compliance for

legal metrology or environmental protection, feedback control or alarm purposes.

- Purchasing decisions can be based on the lowest cost for the required uncertainty performance.
- When the plant is operational, sensor maintenance and calibration can be scheduled to keep on-line uncertainty within the specified limits. Statistical process control can be implemented using the same criterion. When a sensor fault occurs, the impact on measurement quality is indicated by an increase in uncertainty, which can be used to decide the appropriate control system response. Simple, static redundancy calculations based on uncertainty intervals can be used to verify plant performance. Finally, recorded data can be used to demonstrate legal or environmental compliance, as each measurement and uncertainty value can be traced back to a known standard.

The role of on-line uncertainty in the SEVA framework is described more fully later.

5. THE NEED FOR A NEW SENSOR MODEL

Previous sections have described how the output of the analogue sensor has been used in increasingly ingenious ways:- for plant monitoring, feedback control, safety trips, and for process, plant and sensor fault detection. The limitations of the analogue sensor are apparent, but these very limitations have enabled all of these activities to co-exist, and the 4-20mA standard is easily understood and implemented.

The balance between these activities is now, however, in jeopardy. The current piecemeal attempts to enhance one aspect of sensor functionality using digital technology can be at the expense of others.

For example, the advantages of embedding microprocessor technology within the sensor have been described in sections 3.1 and 3.3. The resulting improvements are, however, at the expense of attempts to extract process and plant health indicators from the measurement signal. Returning to Fig. 11, the analogue circuitry shown in (a) may allow the typically high-frequency health-related components of the transducer signal to pass through into the measurement signal. However, the incorporation of a microprocessor shown in (b) decouples the measurement signal from the transducer signal. In this configuration, high frequency components will only be passed on if the microprocessor samples the transducer output at a higher rate than is strictly necessary for measurement purposes alone. In practice, the sensor designer wishes to eliminate every influence other than the measurand itself in order to improve the measurement performance of the sensor. Thus the health component in the measurement signal from a digital sensor may be distorted, attenuated or removed completely. In some instances plant

operators have been reluctant to accept digital instrumentation for this reason.

The fundamental problem is that workers in the various fields appear to have little awareness of each other's achievements and aspirations, and so have not given integration issues a sufficiently high priority.

- Manufacturers developing internal diagnostics within digital sensors have given insufficient thought as to how sensor fault information is to be utilised (by the control system or by higher level fault detection systems), resulting in their generation of single bit or device-specific error codes.
- Workers in the fault detection domain have been slow to respond to the innovations of the sensor manufacturers. They still view measurement information as the only source and carrier of fault information, and therefore rarely distinguish between a fault and its impact upon the measurement. There has also been insufficient consideration of how fault detection is to be integrated into process operation.
- The drafting of standards for digital field communications has provided an excellent opportunity to develop one or more generic models of extended sensor functionality, incorporating such features as the transmission of sensor fault or measurement validity data. No such models have emerged.
- Metrological standards have yet to respond to the potential of intelligent instrumentation for generating on-line uncertainty, and the need to include fault detection as one aspect of measurement validation.

It is clear, therefore, that the varied requirements on the sensor need to be integrated into a new and extended conceptual model.

Must this new conceptual model include all of the aspects proposed in this paper? It has been argued that all facets are required:

- Fault detection is required because safety, efficiency and productivity cannot be guaranteed if it is possible for the control system to act upon faulty data.
- Metrology is required because fault detection on its own is insufficient to guarantee measurement quality, as there are many other influencing factors such as measurement technology, calibration, and process conditions. There is no viable alternative to uncertainty as a comprehensive index of measurement quality, given, for example, the universality of instrument calibration and the central role of uncertainty in the calibration process.

- Digital systems are required to provide economic access to the manufacturer's expertise, which is essential for thorough fault detection and the calculation of uncertainty on-line. It is not conceivable, for example, that uncertainty could be adequately assigned if it were based solely on the (faulty or non-faulty) properties of the measurement, as provided by a conventional FDI and/or MAD scheme. The proper assessment of measurement uncertainty for the industrial sensor requires a detailed, device-specific diagnosis which in turn requires proprietary knowledge and access to signals within the device. This is only possible within digital sensors, which in turn require digital communications to transmit the results of local diagnostics, measurement and uncertainty calculations to the control system.

6. THE SELF-VALIDATING SENSOR

6.1. Introduction

In this section a new model of the industrial sensor is described. The model description consist of a number of definitions, with additional text providing explanation or illustration. This description is not prescriptive: implementation and computational issues will be the subject of future papers. The definitions proceed from the most general to more specific requirements.

6.2. General Concepts

D1: The Self-Validating (SEVA) Sensor The self-validating sensor is an information source. It has memory and the capability for internal computation and digital communications.

Comment For the rest of this section the term 'sensor' will imply a SEVA sensor unless otherwise stated.

D2: The Next Level Up (NLU) The sensor is at the bottom of a hierarchy of processing levels. The sensor interacts with higher level functions having particular attributes:

Control System This regulates the operation of the plant to deliver the required performance.

Fault Detection System This monitors plant operation for indications of actual or impending failure in any plant component. This is achieved both by detecting inconsistencies between measurement data and by interacting with self-diagnosing plant components.

Maintenance System This schedules maintenance, repair, replacement and calibration of plant components.

Compliance System This records plant data in order to demonstrate compliance with safety and environmental protection restrictions. It also flags any actual or impending condition likely to break these con-

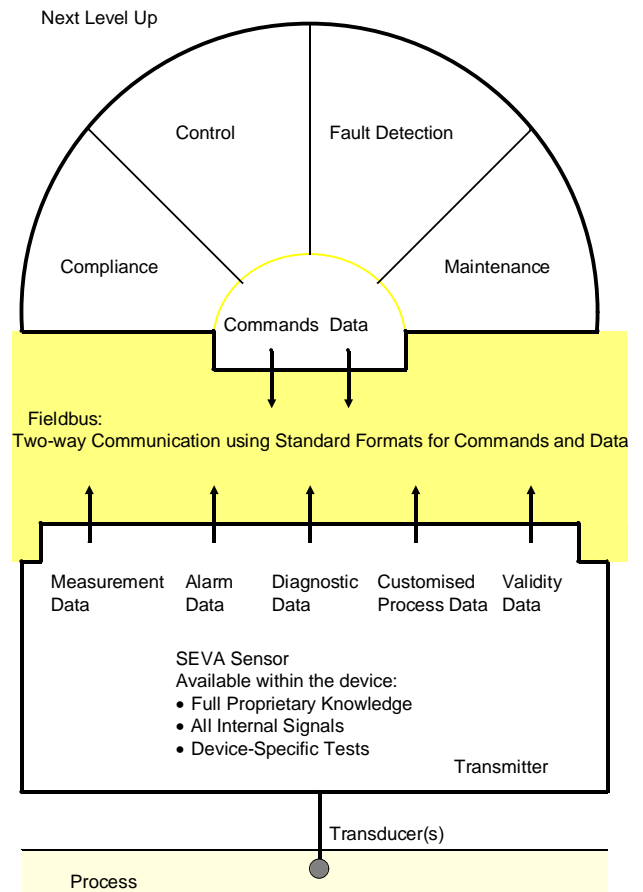


Fig 15. The self-validating sensor and the next level up

straints.

Collectively these systems are described as the Next Level Up (NLU). The sensor provides data to, and receives data and commands from, the NLU via a digital communications system (see Fig. 15).

Comment Each system may be manual (i.e. its function carried out by the operator) or automatic, implemented separately or in a single software framework. Where appropriate, reference to particular systems will be made.

D3: The Hierarchical Relationship The sensor assumes that all data it receives from the NLU is correct. It does not attempt to validate or reconcile this information. Conversely, the NLU may accept or reject sensor data. A sensor may, however, reject a command which it cannot carry out.

Comment The NLU has a wide range of data available to it from numerous sources, while the sensor's view of the process is localised. The sensor will therefore always attempt to reconcile itself with the NLU. The exception is for situations where the sensor is required to perform an impossible task - for example to reconfigure its measurement range beyond its capability. It is assumed that the systems comprising the NLU (see Definition 2 above) co-operate so that the sensor is dealt with consistently.

D4: The Role of the Sensor The sensor performs a variety of tasks. From each task derives one or more principals to guide the design and operation of the sensor:

Plant Operation The sensor provides measurement information that is used to monitor and control the process. It performs validation and provides validity data for each measurement.

- O1:** maximise the availability of each measurement
- O2:** maximise the accuracy of each measurement
- O3:** provide validity data with each measurement

Maintenance The sensor is a piece of plant that must be maintained and repaired.

- M1:** provide detailed diagnostic information to assist maintenance/repair
- M2:** provide timely warning of impending sensor failure

System Integration The sensor is a component of an information network: it must undergo installation, commissioning, reconfiguration, and must communicate with other devices in the network.

- I1:** conform to a standard interface between system components
- I2:** minimise the device-specific nature of communications

Process Fault Detection The sensor provides data which can be used to detect faults in the process or plant components.

- F1:** provide a maximum amount of process and plant health data

D5: The Sensor Boundary The sensor is bound by its physical interface with the process on one side, and its communication interface with the NLU on the other. It may contain several transducers; these may be physically separate from the transmitter which performs all data processing. Any computational function taking place within the sensor has access to all data within the sensor. Beyond the sensor, only certain data are available, to be obtained through standard request messages.

Comment This definition encompasses the simple thermocouple-transmitter pair as well as more elaborate configurations such as a multiple-transducer tank-gauge system or the multi-measurement Coriolis meter. Defining the computational boundary of the sensor in terms of data access and standard functions or services has obvious similarities with structured programming concepts in computer science.

6.3. Sensor Data Types

D6: Sensor-generated data types The sensor gener-

ates several types of data which are made available to the NLU (see Fig. 15):

Measurement Data providing an estimate of a process parameter.

Alarm Data indicating that a process parameter is exceeding pre-set bounds.

Diagnostic Data describing the health of the sensor.

Validity Data the validity of the associated measurement.

Customised Process Data user-designed parameters providing data on the health of the process or plant in the vicinity of the sensor.

In addition, the sensor allows read and restricted write access to device-specific configuration data such as calibration constants, ranges, and so on.

Comment Each of these data types is now described in more detail.

D7: Measurement Data A measurement is the best estimate of the current value of a process variable. A measurement must be generated at the required frequency, even in the presence of a sensor fault. A single sensor may generate multiple measurements.

Comment The calculation of a measurement in the presence of a sensor fault is described later.

D8: Alarm Data The sensor can indicate to the compliance system that a measurand has exceeded, or is anticipated to exceed, some pre-programmed limit. Different alarm levels can be used to indicate the severity of deviation. Limits are likely to be set for reasons of safety or environmental protection.

D9: Diagnostic Data The sensor generates detailed device-specific diagnostic information to assist in maintenance or repair. Additionally, a summary of the current or anticipated state of the sensor is generated to enable maintenance scheduling.

D10: Validity Data Each measurement has an associated validity. This indicates the degree of confidence or belief to be ascribed to the measurement. If a fault occurs in the sensor, the associated validity conveys the reduced confidence in each measurement. There are two complementary aspects of measurement validity:

Accuracy Taking all relevant factors into account, how close to the true process value can the measurement be asserted to lie?

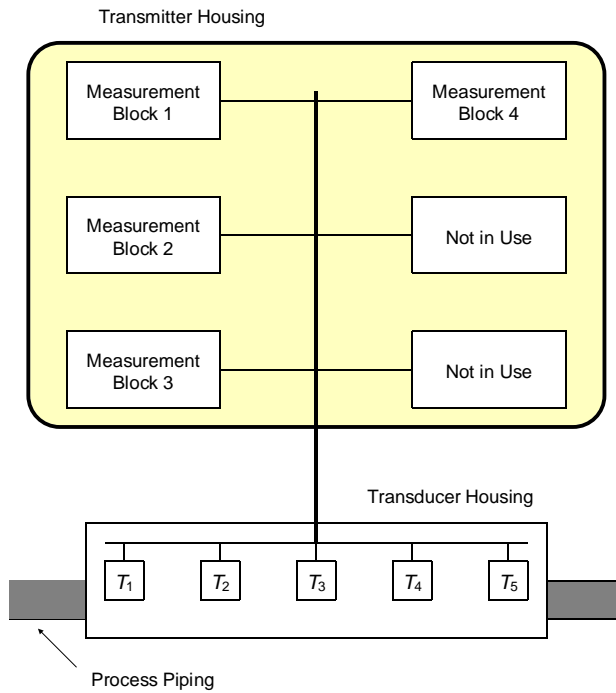


Fig. 16. The modular sensor system

Normality It is required by Definition 7 that a sensor should provide measurement information even in the presence of a sensor fault. The NLU must be made aware of how ‘normal’ the circumstances were when the measurement was generated.

Comment For example, if the transducer is damaged and no longer responds to process changes in the nominal manner, then the measurement may no longer be suitable as a controlled variable, but may still be useful in some secondary capacity. The components of accuracy and normality are largely orthogonal. Clearly a sensor in less than nominal condition will have degraded accuracy. Equally, however, two sensors measuring the same process variable, but using different techniques, will have different accuracies even when both are in prime condition.

D10 (continued) Reflecting these dual aspects of validity, there are two indices of validity associated with each measurement - Uncertainty, and Measurement Value Status - representing the accuracy and normality of the measurement respectively.

Measurement Uncertainty The uncertainty of an individual measurement generated by a SEVA sensor combines bias and precision terms into a single value. It is calculated based upon the uncertainties of the raw data, the calibration process, data processing, and any other relevant factors. It is recalculated each sample to take into account time-varying influences.

Measurement Value Status (MV Status) The Measurement Value Status indicates the circumstances in which the measurement was generated, and takes one of

the following set of values : SECURE, CLEAR, DAZZLED, BLURRED, BLIND, UNVALIDATED.

Comment The interpretation and assignment of MV Status values is described later.

D11: Customised Process Data Optionally, the sensor can provide additional computational facilities enabling the user to process raw transducer data to generate additional parameters related to process or plant health. The responsibility for validating such parameters lies entirely with the user.

Comment Current trends, such as the increasing costs of installation and maintenance, the decreasing cost of computational power, and the widening application of microsensors, suggest that the optimal design for process instrumentation may change radically. Flexible, modular multi-measurement sensor systems, designed to extract as much data as possible from a single breach of the process piping could become both technically feasible and economically attractive. Such a system (Fig. 16) might contain many transducers, measuring both physical and analytical parameters, users installing only those software modules needed to generate the measurements they require. In addition users might generate their own process or plant health-related parameters based on raw transducer data via programmable filters, neural nets or other data reduction processing units (see Fig 17). In this way the measurement and health data could be generated from the same transducers without interfering with each other. Certain transducers (e.g. acoustic) could be provided solely for generating customised parameters. These parameters could be discrete (e.g. flagging some plant or process change), or continuous (e.g. estimation of degree of wear).

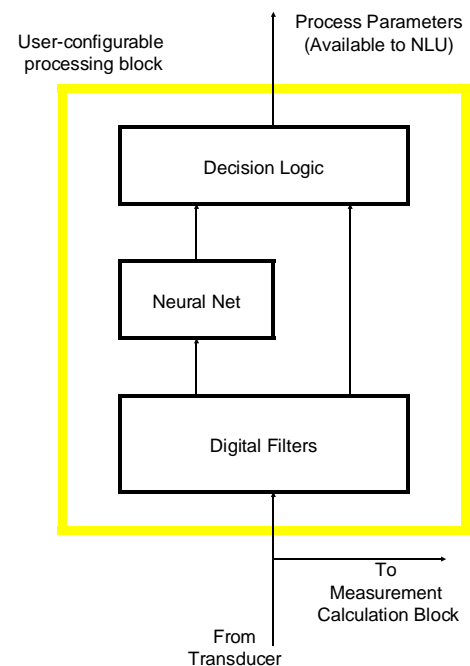


Fig. 17. Example of processing unit generating one or more customised process parameters

6.4. Sensor Diagnostics

D12: Sensor Fault A sensor fault is an aberration that impairs any aspect of sensor functionality. A fault is device-specific: it refers to the underlying cause of the problem rather than any symptoms (such as might be observed on the measurement signal). Faults may be either discrete (i.e. either present or not), or have an associated degree of seriousness.

D13: The Diagnostic State The sensor constantly updates its internal assessment of its condition and performance. The resulting diagnostic state lists all faults that are judged to have occurred or to be imminent, with their estimated degrees where appropriate. The diagnostic state of the sensor determines how all output parameters are to be generated for the current sample.

Comment The diagnostic state is the sensor's best estimate of its condition, and is subject to correction by the NLU (e.g. in the form of a maintenance engineer). Competent self-diagnostics should normally ensure that the diagnostic state is reasonably accurate. Notification of the diagnostic state is supplied to different components of the NLU in different ways, reflecting different aspects of the sensor's functionality.

D14: Raw Measurement Value (RMV) A Raw Measurement Value (RMV) is a measurement value calculated on current raw data under the assumption that there are no faults present.

D15: Raw Uncertainty (RU) The Raw Uncertainty value associated with a RMV is the corresponding uncertainty of the RMV, taking all relevant factors into account other than the presence of faults.

Comment The RU is typically calculated using equations derived from the measurement calculations. Uncertainties of calibration constants are quantified during the calibration process.

D16: Validated Measurement Value (VMV) A Validated Measurement Value (VMV) is the measurement value representing the sensor's best estimate of the measurand, given the current diagnostic state.

D17: Validated Uncertainty (VU) The Validated Uncertainty is the uncertainty of the VMV.

D18: Diagnostic Information Sources There are a variety of diagnostic information sources available within the sensor.

Raw data The first signal available in the device, generally an electrical image of a transducer output, for example a frequency, voltage or resistance.

RMV data Signals which correspond to process pa-

rameters e.g. mass flow or temperature.

Auxiliary data Signals within the sensor which are not directly related to the measurand.

Device-specific tests Data from active self-testing by the sensor.

NLU inputs Inputs from the Next Level Up, generally discrete values and thresholds relating to expected process behaviour.

Application knowledge base Information relating process conditions to likely fault behaviour.

Comment Each component is described in more detail.

D19: Raw data Raw data contain the maximum information available on the transducer response and are therefore a very rich source of information for statistical tests.

D20: RMV data Most fault detection techniques in the literature are based largely upon detecting aberrations in measurement signals (either on an individual or a collective basis). Within the SEVA sensor such techniques can be applied to the RMV in order to provide one aspect of diagnostic information.

D21: Auxiliary Information Auxiliary information is provided by the potentially large number of other signals within the sensor which are not directly related to the raw data or RMV, but which can give useful information about the health or performance of a sensor. These might include the electrical properties of components within the sensor, signal levels at the input or output stages of power amplifiers, or hardware errors. Associated with the auxiliary information may be statistical or other tests to identify characteristic behaviour. Hardware errors are a special, preprocessed form of auxiliary information generated by digital components within the sensor, requiring little or no processing or interpretation. They generally map directly and immediately to a physical diagnosis (e.g. memory checksum error).

Comment All of this auxiliary data is highly device-specific and best defined by, or in close consultation with, the instrument designers.

D22: Device-specific tests In addition to monitoring auxiliary data, the sensor may be able to actively perform device-specific tests. These typically stimulate the transducer or other sensor component and judge health by observing the response.

Comment Again, such tests are best implemented by the sensor designers. Artificial stimulation of the transducer may, of course, result in a 'faulty' measurement: it is dealt with in the same way as any other fault as described later.

One difficulty associated with localised validation is that there may be situations where the sensor has insufficient information to be able to come to a valid conclusion, for example in distinguishing between certain types of sensor fault (e.g calibration drift) and legitimate process changes. The sensor must therefore be able to appeal to the NLU (which has access to additional information such as data from other sensors) for clarification.

D23: NLU Inputs The sensor can ask for information about process behaviour (e.g. anticipated process limits) through standard requests to the NLU. In addition the NLU may provide warnings of changes in process behaviour which can be used to minimise false alarms.

D24: Application knowledge base The manufacturer's accumulated knowledge of conditions likely to result in particular faults should be made available to the sensor's diagnostic considerations. The data may include factors such as time since last calibration and maintenance.

Comment In some applications sensors are known to degrade much more rapidly under certain conditions, e.g. at extremes of operating range. Such knowledge potentially allows a sensor to estimate a "wear and tear" effect on its performance. For example, if a normal range pH probe is exposed to more than about 12 pH for as little as an hour it may become alkali-conditioned and fail to respond when the solution becomes more acidic.

D25: The Diagnostic State Machine (DSM) The Diagnostic State Machine (DSM) is the software system responsible for co-ordinating and resolving all diagnostic information, as well as for scheduling self-tests and requests for data from the NLU.

Comment DSM complexity may vary widely depending upon the sophistication of the sensor. A variety of diagnostic techniques can be used within the DSM, including FDI, signal analysis, and neural nets. The highest level of the DSM is likely to contain a knowledge-based element in view of the heuristic nature of the decisions that must be taken.

D26: Steps Taken Each Sample Figure 18 illustrates the steps taken each sample within the sensor. Raw data is sampled from the transducer(s) and propagated through the measurement equations to generate the RMV(s) and the corresponding RU(s). Statistical and/or other feature extraction processing is performed upon the RMV(s) and the raw data. Other diagnostic data, as described in Definition 18 above, may also be subject to statistical or other processing, either every sample or on an occasional basis. All of this data is passed to the DSM which updates the current diagnostic state. Based on this the output parameters are calculated: each VMV with associated VU and MV Status, Alarm Outputs, Device Status and detailed Diagnostic Data.

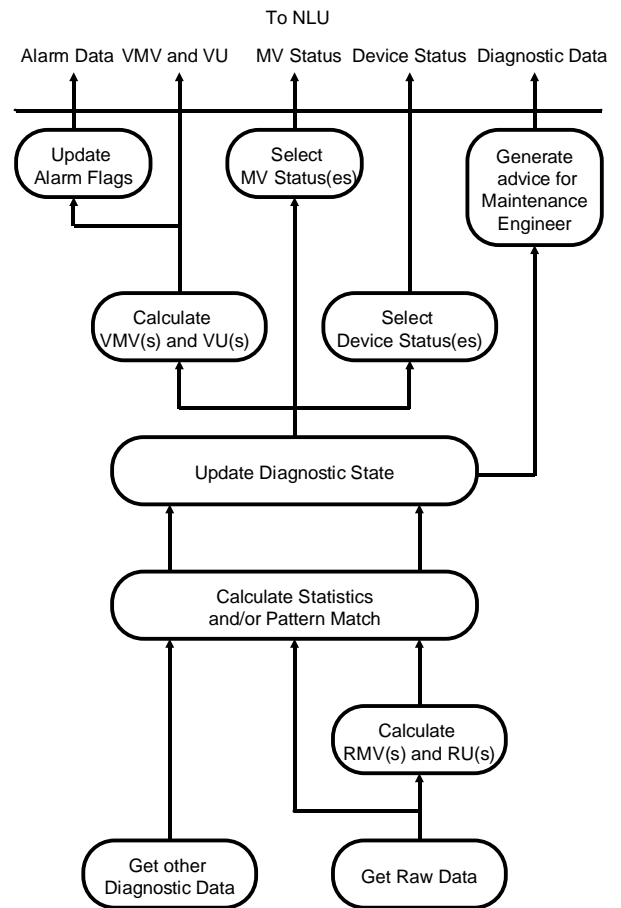


Fig 18. Steps taken each sample

D27: Calculation of VMV and VU Any of a number of strategies may be used to calculate the VMV and the VU, given the current Diagnostic State, as listed below in decreasing order of preference. Each sample one method is selected for each measurement generated by the sensor. Different strategies may be used for different measurements, reflecting the different impact the current Diagnostic State may have upon each measurement. A fault, diagnostic test or other condition (e.g. time since the last calibration) may have a detrimental effect upon a measurement: in the following descriptions, the underlying cause is less important than the resulting confidence in the RMV.

Nominal confidence in RMV If there is no fault or test condition which has any impact on confidence in the RMV, then the VMV is set equal to the RMV and the VU is set equal to the RU.

Reduced confidence in RMV A fault or test which has only a partial impact upon the measurement process should be accommodated. This means that the VMV and VU should be assigned corrected values. The method of correction depends upon the degree of understanding of the impact of the fault or test upon all aspects of the measurement process and calculation. Any introduction of bias into the (partial) measurement should be corrected; any increase in uncertainty should result in a corresponding increase in the appropriate uncertainty

variable in the measurement calculation. Such corrections should take place as far back as possible within the measurement calculation. If necessary the entire measurement calculation must be reperformed incorporating all corrections. The VMV and VU are then set equal to the corrected RMV and RU values.

Comment The following examples illustrate the range of fault accommodation strategies that may be applied.

- A fault has the effect of changing the property of the transducer, so that the 'true' value of one of the calibration constants changes. Calculation of the measurement based upon the normal calibration constant value will result in a biased measurement. Supposing this fault to be well-understood by the manufacturer, the sensor is able to estimate a correction factor for the calibration constant based upon diagnostic data (e.g. auxiliary data or a device-specific test). As this correction is more approximate than the normal calibration procedure, the uncertainty of the corrected calibration constant is increased accordingly. Recalculation of the measurement data will result in corrected values of the VMV and VU.
- Another fault, although detectable, has an impact upon the measurement which is poorly understood (due to its rarity and/or the expense of analysing the relationship between the fault and the measurement). Based upon a heuristic assessment, the RMV is still considered to be the best estimate of the measurement (so the VMV is set equal to the RMV), but the VU is set equal to the RU plus an extra uncertainty factor to accommodate the fault.

Clearly the first example is a far more satisfactory response to the presence of a fault than the second. This illustrates the possibility of assessing and/or improving instrument performance on the basis of measurement quality in faulty, as well as non-faulty conditions. In general a manufacturer might be expected to improve the detection and accommodation of faults and other conditions as experience accumulates over the lifetime of a particular product.

D27 (continued)

No confidence in the RMV A test or fault may have the effect that the RMV bears no relation to the process measurand. In such circumstances the VMV is calculated from historical data, and the VU includes the additional uncertainty entailed in extrapolating past behaviour to predict the present value. There are two types of historical data:

- Short-term history - the behaviour of the measurand immediately prior to the onset of the impairing condition;

- Long-term history - the process knowledge accumulated by the sensor such as limits on measurand value, maximum rates of change, and characteristic process behaviour.

This data is used to provide a best estimate of the current measurand value, with associated uncertainty.

Comment Faults of this type include those in which the transducer has entirely ceased to function or has become disconnected from the transmitter. Similarly, if a test stimulates the transducer, thereby completely obscuring

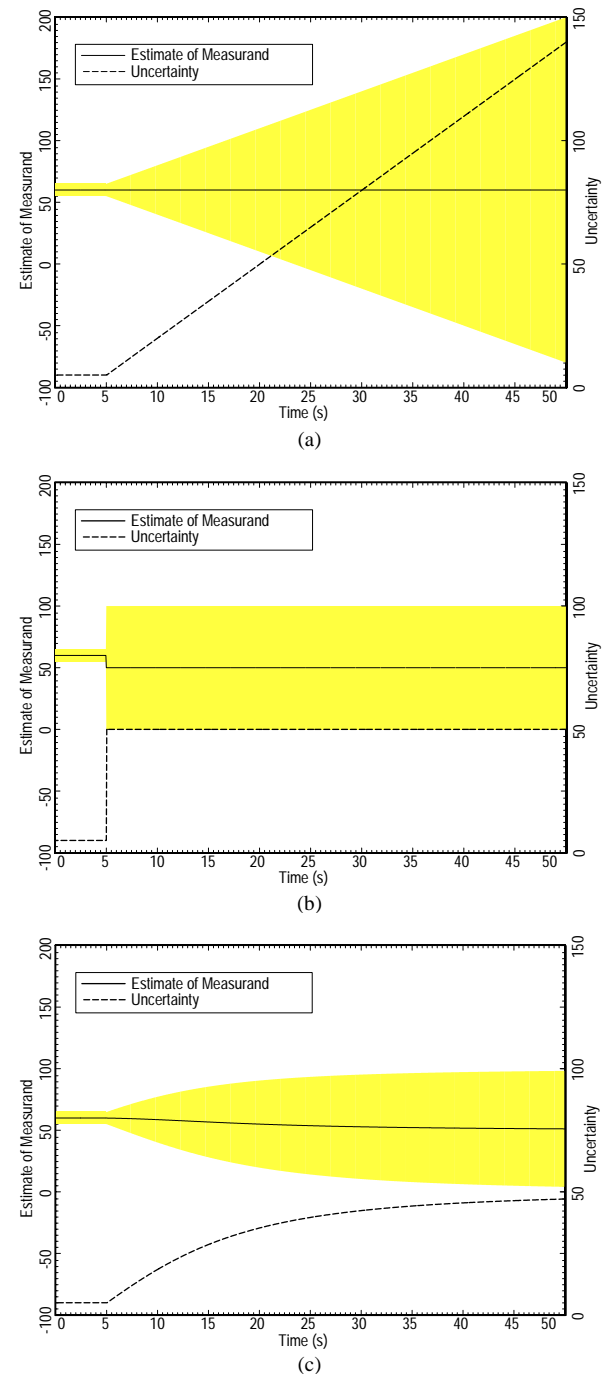


Fig. 19. Example of measurement and uncertainty estimation using historical data:
(a) Short Term Process History;
(b) Long Term Process History;
(c) combination of (a) and (b) using uncertainty calculus.

the process-related signal, then this strategy is appropriate for the duration of the test.

The calculation of the VMV and VU in these circumstances can be performed using varying degrees of sophistication for the process modelling. A simple strategy is presented here by way of example. This strategy combines two estimates of the measurand and its uncertainty, based on short-term and long term historical data, using the rules of uncertainty calculus. Figure 19 shows the resulting short-term and long-term estimates and their combination from a simple simulation. In Fig. 19, the unbroken line is the estimate of the measurand, as denoted on the left hand scale, the dashed line shows the uncertainty, denoted on the right-hand scale, and the shaded area shows the uncertainty interval around the estimated measurand.

In the simulation, the measurement is fault free, with a value of 60 units and an uncertainty of 5 units, until $t=5s$. At this point a fault condition occurs which reduces the confidence in the RMV to zero.

Figure 19(a) shows the strategy using short-term process history. The estimate is set equal to the last fault-free measurement, and the uncertainty is increased by the maximum observed rate of change of the process. In other words, this strategy asserts that the best estimate of the measurand is where it was last seen, and that it is unlikely to change from this value by more than the previously observed maximum rate of change of the process, which in the simulation is 3 units/s. This strategy is of course both heuristic and pessimistic, as probabilistically the measurement is likely to change more slowly.

Figure 19(b) shows the strategy using long-term process history. In the absence of a current estimate of the measurand, 95% confidence limits on its value can be calculated based upon long-term history. In Fig. 19(b) these limits are 0 and 100 units respectively. Once the fault has occurred, the estimate of the measurand is set to the average historical value (in this case 50 units), and the uncertainty is set to be half of the confidence interval. In other words, this strategy asserts that the measurement could take any value within previous experience.

Neither of these strategies is entirely satisfactory. The first works well in the immediate aftermath of the fault, but as time proceeds the uncertainty interval may exceed the physical limits of the sensor and/or process (e.g. include negative values). The second strategy is unsatisfactory initially, as there is a sudden change in the estimate of the measurand and an enormous increase in uncertainty.

A better strategy is one which combines the estimates generated by the first two. It is reasonable to assume that the two estimates are independent, and so using

uncertainty calculus it is possible to combine them to get a 'better' estimate (i.e. one with a smaller uncertainty). Given estimates e_1 and e_2 with uncertainties w_{e_1} and w_{e_2} , a better estimate is given by

$$e_{best} = \frac{w_{e_2}^2}{(w_{e_1}^2 + w_{e_2}^2)} e_1 + \frac{w_{e_1}^2}{(w_{e_1}^2 + w_{e_2}^2)} e_2 \quad (9)$$

Each estimate is thus weighted by the square of its uncertainty. The uncertainty of e_{best} is given by

$$w_{e_{best}} = \frac{w_{e_1} w_{e_2}}{(w_{e_1}^2 + w_{e_2}^2)^{1/2}} \quad (10)$$

Figure 19(c) shows the results of combining the short-term and long-term estimates. Immediately after the fault the short-term estimate is more prominent; subsequently the long-term estimate becomes dominant. The value of the estimate slowly moves from the last non-faulty value (60 units) towards the long-term average (50 units).

Of course there is no guarantee that the process measurand is actually within any of these uncertainty limits - particularly if the fault has been caused by some massive process change. However, the sensor is required to provide a best estimate of the measurand, with an associated uncertainty, and in the absence of on-line transducer data it can only assume that current process behaviour reflects historical trends. As the NLU has access to other process data, it is better able to determine whether this assumption is valid, and whether therefore the sensor's estimate can be trusted.

D27 (continued)

Combination of effects In cases where there is more than one condition affecting a measurement, then all effects are included, if possible:

- If any condition causes the confidence in the RMV to be zero, then the VMV and VU must be calculated using historical data only, as described above (i.e. this condition over-rides any causing only reduced confidence in the RMV).
- If two conditions cause reduced confidence in the RMV, then both sets of corrections should be applied. If the corrections are mutually exclusive then the larger correction is applied.

D28: Calculation of the Measurement Value Status

Definitions 7 and 10 state that the sensor should generate a best estimate of the measurand together with an associated uncertainty, even in the presence of faults. Definition 27 describes strategies for generating VMV and VU values in the presence of one or more faults. A change in uncertainty by itself gives no indication of the extent or likely evolution of the conditions giving rise to the change. In particular, the NLU may need to know

which strategy has been used to calculate the VMV and VU. It is also important to know whether an abnormal condition is expected to persist. This information is conveyed by the Measurement Value Status.

Given two conditions :

- 1) Expected persistence of any abnormal condition: long or short
 - 2) Confidence in latest RMV: nominal/reduced/zero
- The corresponding value of MV Status is shown in Table 2.

Table 2 MV Status Values

Expected persistence	Confidence in RMV	MV Status
not applicable	nominal	CLEAR
not applicable	reduced	BLURRED
short	zero	DAZZLED
long	zero	BLIND

Each measurement is assigned an MV Status value as determined from the current Diagnostic State. There are two additional status values:

- UNVALIDATED: validation is not taking place.
- SECURE: the VMV has been generated using redundant measurements. The confidence in each measurement is nominal.

Comment The status names were chosen by analogy with vision. Typical examples of each are given below:

CLEAR The RMV signal structure is within normal range for given process conditions.

DAZZLED The RMV signal structure is substantially abnormal but there is insufficient evidence to project future behaviour (typically spiky/outlier contaminated signal or initial response to a fault). Another use of DAZZLED is when the sensor is performing active self-testing which prevents reliable data from being generated by the transducer(s). The length of the test is likely to be known, and to be short. For the duration of the test, all measurements which are completely impaired are set to DAZZLED.

BLURRED The RMV signal structure is abnormal but still with reasonable correspondence to the measurand (typically noisy or with distorted spectrum).

BLIND The RMV signal structure is so abnormal that it bears no relation to the measurand (typically zero variance or outside physical limits for significant interval); the fault is expected to persist.

For example, Fig. 20 shows likely values of the MV Status in the scenario shown in Fig. 19. Prior to the fault the MV Status is CLEAR. At the onset of the fault, a serious aberration is detected immediately, but the

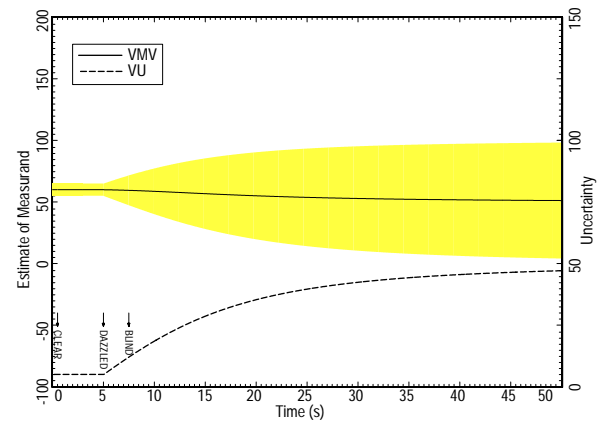


Fig. 20 MV Status behaviour corresponding to fault in fig. 19

cause, and therefore the persistence, of the fault cannot yet be determined, and so the Status is set to DAZZLED. A little later, the DSM determines the nature of the fault, which is both severe and permanent, and so the Status is set to BLIND.

The MV Status may be used in selecting the appropriate control strategy for a controller using the related measurement. There are many possibilities for how this can be done depending on the application. For a simple loop a possible strategy is shown in Table 3.

Table 3 Control Strategy using MV Status

Status	Operating Mode
CLEAR	Normal PID
BLURRED	Detuned PID
DAZZLED	Temporary freeze
BLIND	Forced manual or alternative measurement

The response to a BLIND measurement must almost inevitably be seriously disruptive, and therefore this Status value should only be adopted if there is strong evidence that the condition will persist. DAZZLED acts as an intermediary condition so that if a fault is severe but turns out to be temporary (e.g. a spike error), the MV Status sequence (CLEAR - DAZZLED - CLEAR) will be far less disruptive, than if DAZZLED was not defined (CLEAR - BLIND - CLEAR). Obviously if there is sufficient evidence within the sensor that the fault is permanent as soon as it occurs, then the Status is set to BLIND immediately.

D29: Calculation of Device Status The Device Status is a generic, discrete value summarising the health of the sensor. A single value is generated each sample. It is monitored by the Fault Detection and Maintenance systems. It takes one of the following values:

GOOD The sensor is in nominal condition.

TESTING The sensor is performing diagnostic tests

which may be responsible for any loss of measurement quality.

SUSPECT The sensor may be suffering from an aberration; the condition has not yet been diagnosed.

IMPAIRED The sensor is suffering from a diagnosed fault which has a minor impact on performance, warranting a low priority maintenance call.

BAD The sensor is suffering from a diagnosed fault which has a major impact on performance, warranting a high priority maintenance call.

CRITICAL The sensor is in a condition that is potentially dangerous, requiring immediate attention.

Comment It is stressed that that the (single) Device Status refers to the health of the *sensor*, whereas the MV Status refers to the quality of each (of one or more) *measurement*. Normally there will be some correlation between the Device Status and MV Status(es) - for example, if the principal measurement is BLURRED then the Device Status is likely to be IMPAIRED.

CRITICAL is used to indicate that the sensor is in a condition that may cause (or have caused) a hazard, such as a leak, of the process fluid or of a dangerous material (e.g. a reagent or a radioactive substance), fire or explosion. This status refers only to hazards generated by the sensor itself, rather than by the process. The following examples illustrate the different cases.

- If the process temperature and/or pressure is so high that the transducers are unable to operate, then the Device Status is set to BAD.
- If the process temperature and/or pressure is high enough to damage nearby plant, but not the sensor itself, then the Device Status may remain GOOD, although the sensor may have separate alarm functions programmed to flag such conditions.
- If the process temperature and/or pressure is so high that the sensor itself is likely to rupture, leak, explode or catch fire, then the Device Status is set to CRITICAL, even if the measurements themselves are perfectly functional.

Clearly only certain types of sensor are capable of being hazardous. Detailed, device-specific diagnostic data is also available for the instrument engineer.

7. APPLICATION: THE CORIOLIS METER

An on-line PC-based prototype of a self-validating Coriolis meter has been developed. This consists of a conventional, commercially available flowtube and transmitter, together with signal conditioning electronics and a 386 PC. The software package is called SIAMESE and is written in C++.

SIAMESE acts as a self-validating Coriolis meter:

- Signals relating to both measurements and sensor health are picked up from the transmitter.
- Measurement and uncertainty calculations are performed within the software. In non-faulty conditions, the resulting VMVs (one each for temperature, density and mass flow) show excellent agreement with the values generated by the transmitter.
- Uncertainty values are calculated based upon a detailed uncertainty analysis of the instrument. This analysis has been extended to model the impact of the most important fault modes of the sensor.
- A number of fault conditions can be detected. During the presence of a fault, the VMV is corrected and VU increased appropriately. On-site metered trials have demonstrated the superiority of the VMV values generated by the SIAMESE system over the faulty values generated by the transmitter.

A simple example is used to contrast the behaviour of the conventional and validated instrument in the presence of a sensor fault. The fault demonstrated is the loss of RTD input. Fig. 21 shows how the measurements generated by the conventional Coriolis meter behave when this fault occurs. Unsurprisingly, the response of the temperature measurement is drastic: it drops to about -110°C within a few seconds of losing the RTD input. However, as the mass flow and density measurements are calculated using the temperature input, there are corresponding shifts in these measurements, of about 7% and 20% respectively. At t=47s, the temperature input is reconnected, resulting in a return to fault-free operation.

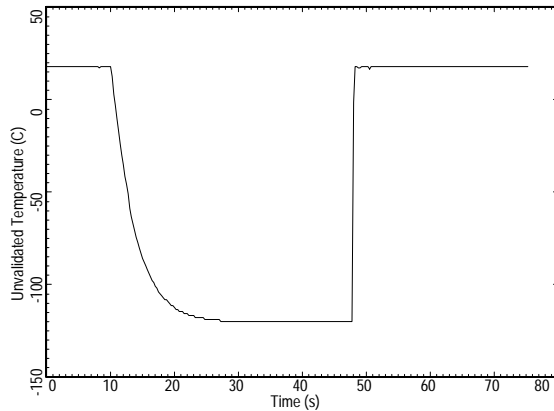
It is reasonable to assume that any fault detection system worthy of the name ought to be able to detect this fault simply by observing the temperature measurement. However, it is unusual for the Coriolis temperature output to be used by the control system (particularly in a 4-20mA regime, where normally only one measurement can be transmitted from the sensor). More typically only the mass flow or density values are used. The more marginal biases in these measurements are less clearly distinguishable from a genuine process change or disturbance. This has already been discussed at length in the introduction, as the fault shown in Figs. 3 and 4 is in fact the loss of RTD input (compare Fig. 21(c) with Fig. 3).

Even if the sensor has self-diagnostics, and is able to detect the fault, without the benefit of uncertainty analysis it is difficult to ascertain the impact this fault has upon the mass flow and density measurements, and therefore, assuming a choice of discrete flags, whether they are sufficiently impaired to be declared 'faulty' or

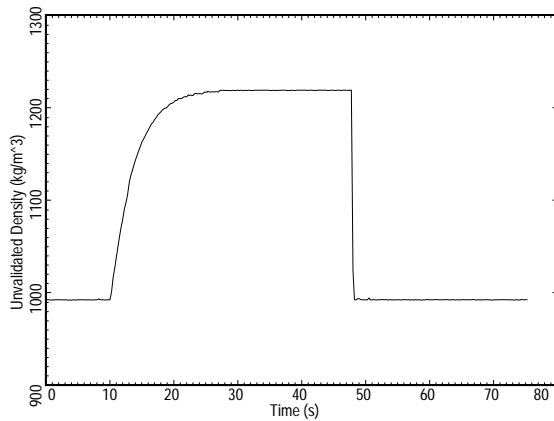
not.

Figure 22 shows how the SIAMESE program responds to the fault. Note that only the mass flow output has been drawn to the same scale as used in Fig 21: this enables features of the response to be displayed more clearly.

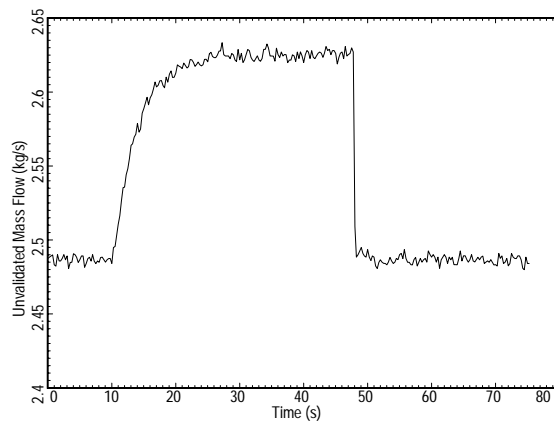
There is an initial warmup period terminated by the first change of status to CLEAR for each measurement.



(a)



(b)

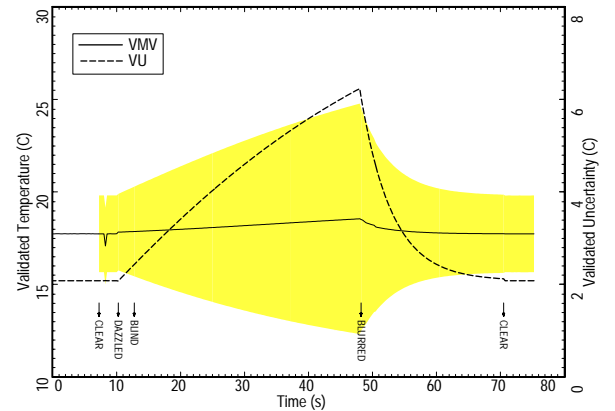


(c)

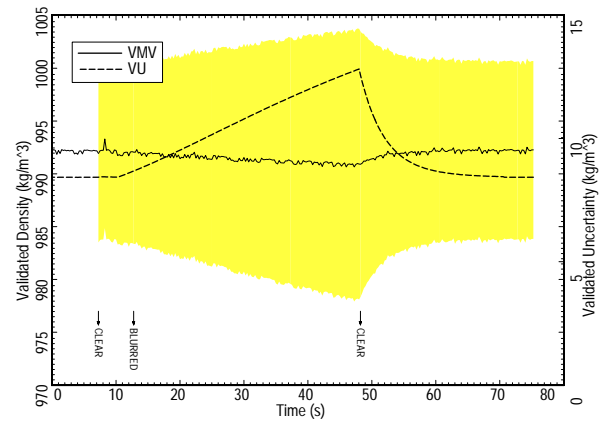
Fig 21. Response of unvalidated Coriolis meter to RTD input loss
(a) temperature
(b) density
(c) mass flow

The loss of RTD input is detected by observing an internal signal. The detection has two stages. Initially, there is a rapid drop in value of the signal, which is far faster than could be generated by normal RTD operation. A little later the signal settles at a constant value characteristic of RTD input loss, and hence the correct diagnosis is deduced.

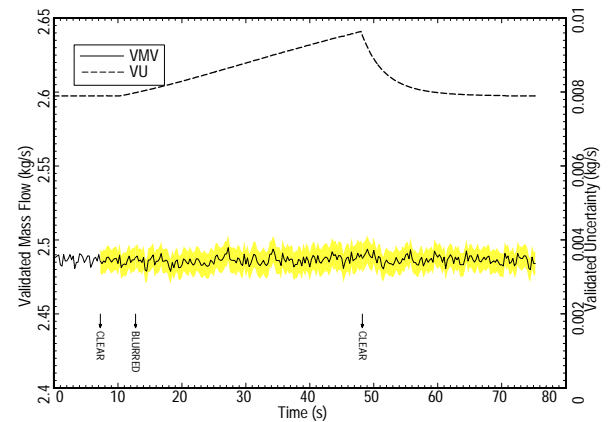
The measurements and their validities are calculated based upon this two-stage diagnosis. When the initial



(a)



(b)



(c)

Fig 22. Response of SIAMESE system to RTD input loss
(a) temperature
(b) density
(c) mass flow

drop is detected, the temperature is set to DAZZLED, and the VMV and VU are calculated using the method described in Definition 27 above. Subsequently, at $t=14s$, the loss of RTD diagnosis is confirmed and the status of temperature is set to BLIND. The VMV and VU are calculated in the same way as during the DAZZLED stage. When the RTD input is reconnected, a recovery mode is entered, during which the temperature is set to BLURRED.

The loss of the RTD input has an indirect impact upon the density and mass flow, as shown in Fig. 21. Within SIAMESE this is accounted for in the calculations of their VMV and VU values by the propagation of the temperature uncertainty through the measurement equations. For example, the mass flow is calculated using an equation of the form

$$M = (K_1 T + K_0) \cdot \frac{\varphi}{F} \quad (11)$$

where K_1 and K_0 are flowtube calibration constants, subject to uncertainties w_{K_1} and w_{K_0} respectively, φ is the phase difference between the two motion sensors, subject to uncertainty w_φ , F is the drive frequency, subject to uncertainty w_F , and T is the measured process temperature, subject to uncertainty w_T .

The corresponding equation for the uncertainty of the mass flow rate is given by the following:

$$\left(\frac{w_M}{M}\right)^2 = \frac{(T w_{K_1})^2 + w_{K_0}^2}{(K_1 T + K_0)^2} + \frac{K_1^2 w_T^2}{(K_1 T + K_0)^2} + \left(\frac{w_\varphi}{\varphi}\right)^2 + \left(\frac{w_F}{F}\right)^2 \quad (12)$$

When the fault in the temperature input is detected, the impact on the mass flow can be calculated by propagating the value of w_T into this equation. As the temperature uncertainty increases from about 2°C up to 7°C, the resulting uncertainty in the mass flow rate increases from 0.008 kg/s up to 0.0095 kg/s. Under different process conditions, the same fault may have greater or lesser consequences

The Status of the mass flow and density measurements are selected based upon the current diagnosis. When the reason for the temperature loss is unknown, these Status values remain CLEAR. However, once the fault is diagnosed, then it is judged that, as mass flow and density have a secondary dependence upon temperature, their Status values should be set to BLURRED. As soon as recovery begins, they are reset to CLEAR.

Assuming the temperature input is not so rapidly (and artificially) restored, how might the different systems comprising the Next Level Up respond to this fault?

- The Control System is informed that the temperature has gone BLIND, while the other two measurements have gone BLURRED with marginally increased uncertainties. Any feedback control based on temperature will have to be switched to manual or another measurement. Continued use of mass flow or density will depend upon application-specific criteria, which may be specified in terms of the generic parameters uncertainty and Status.
- The Fault Detection System notes the change of Device Status from GOOD to IMPAIRED; any local and temporary inconsistency between variable detected by an analytical redundancy scheme may be attributable to this fault.
- The Maintenance System notes the change of Device Status and schedules a low priority maintenance check. A detailed diagnosis is obtained from the instrument to enable the engineer to take appropriate action.
- The Compliance System is able to judge, based upon the increased levels of uncertainty, whether any of the measurements infringe safety, environmental protection, or legal metrology restrictions. If so, it will indicate to the Control System that further action is required.

The Siamese system is able to detect a growing number of faults within the Coriolis meter (currently some 15), occurring alone or in combination, the most important of which detect more subtle faults than a simple loss of input. However, even ‘trivial’ faults, like in the example above, may have an important impact on measurement output, and their detection can avert serious degradation of plant performance.

8. CONCLUSION

This paper has argued the need for an extended model of the generic sensor in order to accommodate increasing performance requirements and extended capabilities. A series of definitions have been provided to describe the self-validating sensor, and a prototype system has been presented. Clearly this is only a first step in describing a new methodology of plant operation.

Future papers will examine the following topics:

- Fault detection in the Coriolis meter;
- The self-validating actuator: rationale, definitions and examples;
- The use of uncertainty data in Analytical Redundancy schemes;

- The relationship between faults and uncertainty;
- The development of standardised strategies for exploiting sensor validation information at the loop and unit levels.

9. ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of DTI, SERC, Foxboro Great Britain Limited, and ICI PLC for the work described in this paper, carried out as the SEVA project under the LINK initiative. Particular thanks are given to our colleagues in the SEVA project: Paul Clarke, Doug de Sa, Milos Machacek, Steve Walsh, Martin Ward, Terry Wilkins, Graeme Wood and Janice Yang.

10. REFERENCES

- ANSI (1983). Measurement Uncertainty for Fluid Flow in Closed Circuits. ANSI/ASME MFC-2M-1983 (reaffirmed 1988)
- ANSI (1985). Measurement Uncertainty. ANSI/ASME PTC 19.1-1985
- Belchamber, R. M. and M. P. Collins (1990). Acoustic Monitoring of Processes. *IMC/IMA conference on Advances in Measurement*, Birmingham, U.K.
- Bigg, M. G. (1990). Integrated Pollution Monitoring. *IMC/IMA conference on Advances in Measurement*, Birmingham, U.K.
- Blickley, G. J. (1991). Valves and actuators changing to meet standards and regulations. *Control Engineering*, October 1991, 111-117.
- de Sá, D. (1988). The evolution of the intelligent measurement. *Meas. and Control*, 21(5) 142-144.
- Cork, S. P. (1989). Evaluation of Coriolis meters. SIRA Seminar on Coriolis Mass Flowmetering, December 1989, London.
- Dean, R. C. (1975). Editorial. *ASME J. Fluids Eng Trans.* ASME., 97, 141.
- Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy - a survey and some new results. *Automatica.*, 26, 459.
- Frank, P. M. (1991). Enhancements of robustness in observer-based fault detection. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Gertler, J. (1991). Analytical redundancy methods in fault detection and isolation. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Halme, A. and J. Selkainaho (1991). Advanced fault detection for sensors and actuators in process control. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Henry, M. P. and D. W. Clarke (1991). A standard interface for self-validating sensors. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Henry, M. P. and G. G. Wood (1992). The implications of digital communications on sensor validation. *IFAC Symposium on On-line Fault Detection and Supervision in the Chemical Process Industries*, Newark.
- Himmelblau, D. M. (1992). Use of artificial neural networks to monitor faults and for troubleshooting in the process industries. *IFAC Symposium on On-line Fault Detection and Supervision in the Chemical Process Industries*, Newark.
- Hoskins, J. C., K. M. Kaliyur, and D. M. Himmelblau (1991). Fault diagnosis in complex chemical plants using artificial neural networks. *AIChE J.*, 37, 137-142.
- ISA (1980). Measurement Uncertainty Handbook. ISBN:87664-483-3
- Isermann, R. (1991). Fault diagnosis of machines via parameter estimation and knowledge processing. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- KB MUSICA (1992) Enhanced Sensors: Report CHEM.ES.17.IC1.D1. *Knowledge-Based Multi-Sensors Systems in CIM Applications*, Project No 2671, ESPRIT II.
- Kline S. J. and F. A. McClintock (1953). Describing uncertainties in single sample experiments. *Mech. Eng.*, 3-8.
- Kline, S. J. (1985) The purposes of uncertainty analysis. *J. Fluid Eng.*, 107, 153-160.
- Maßberg, W. and H. -J. Seifert (1991) Petri net based system for monitoring, diagnosis and therapy of failures in complex manufacturing systems. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Milne, R. (1991) Integration: the key to second generation systems. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Moffat, R. J. (1985) Contributions to the theory of single-sample uncertainty analysis. *J. Fluid Eng.*, 107, 153-160.
- Patton, R. J., P. M. Frank and R. N. Clark (Ed) (1989). *Fault Diagnosis in Dynamic Systems, Theory and Applications*. Prentice-Hall, Englewood Cliffs, NJ.
- Patton, R. J. and J. Chen (1991) Parity space approach to model-based fault diagnosis - A tutorial survey and some new results. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Patton, R. J., J. Chen, and J. H. Millar (1992). Robust fault detection for a nuclear reactor system: a feasibility study. *IFAC Symposium on On-line Fault Detection and Supervision in the Chemical Process Industries*, Newark.
- Rowan, D. A. (1992). Beyond Falcon: industrial applications of knowledge-based systems. *IFAC Symposium on On-line Fault Detection and Supervision in the Chemical Process Industries*, Newark.
- Tzafestas S. (1989). System fault diagnosis using the knowledge-based methodology. In *Fault Diagnosis*

- in *Dynamic Systems: Theory and Applications* (R. Patton, P. Frank and R. Clark, editors). Prentice Hall. 509-572.
- Tzafestas S. (1991). Second generation expert systems: requirements, architectures and prospects. *IFAC/IMACS Symposium on Fault Detection, Supervision and Safety for Technical Processes*, Baden-Baden.
- Vachkov, G. and H. Matsuyama (1992). Identification of fuzzy rule based system for fault diagnosis in chemical plants. *IFAC Symposium on On-line Fault Detection and Supervision in the Chemical Process Industries*, Newark.
- Wood, G. G. (1991). Fieldbus for Integration of Sensors and Systems. *Sensors and Systems '91*, NEC Birmingham.
- Yung, S. K. and D. W. Clarke (1989). Sensor Validation. *Meas. and Control*, 22, 132-150.