Chapter 2
Fault Tolerant Control and Fault Detection and Isolation

When a fault occurs in a system, the main problem to be addressed is to raise an alarm, ideally diagnose what fault has occurred, and then decide how to deal with it. The problem of detecting a fault, finding the source/location and then taking appropriate action is the basis of fault tolerant control.

In this chapter, an introduction to fault tolerant control (FTC) and fault detection and isolation (FDI) will be presented. The chapter will start with some definitions and will describe different types of faults and failures which can occur in actuators and sensors. Later, different types of fault tolerant controllers and FDI schemes will be presented and discussed.

2.1 Faults and Failures

First, the terms fault and failure will be defined. The definition provided in this book is in compliance with the definitions given by the IFAC SAFEPROCESS technical committee (as given in [136]) which were developed to set a standard [51] in this area in order to avoid confusion among researchers. This will also enable the concept of fault tolerant control (FTC) to be specified in terms of faults and/or failures later in the chapter. The IFAC technical committee, as outlined in [136], makes the following definitions:

> ***fault***: an unpermitted deviation of at least one characteristic property or parameter of the system from the acceptable/usual/standard condition.
> ***failure***: a permanent interruption of a system's ability to perform a required function under specified operating conditions.

Clearly, a failure is a condition which is much more severe than a fault. When a fault occurs in an actuator for example, the actuator is still usable but may have a slower response or become less effective. But when a failure occurs, a totally different actuator is needed to be able to produce the desired effect.

In aircraft systems, there are some distinct types of actuator failure, the three most common are shown in Fig. 2.1. A *lock failure* is a failure condition when an actuator

H. Alwi et al., *Fault Detection and Fault-Tolerant Control Using Sliding Modes*, Advances in Industrial Control,
DOI 10.1007/978-0-85729-650-4_2, © Springer-Verlag London Limited 2011

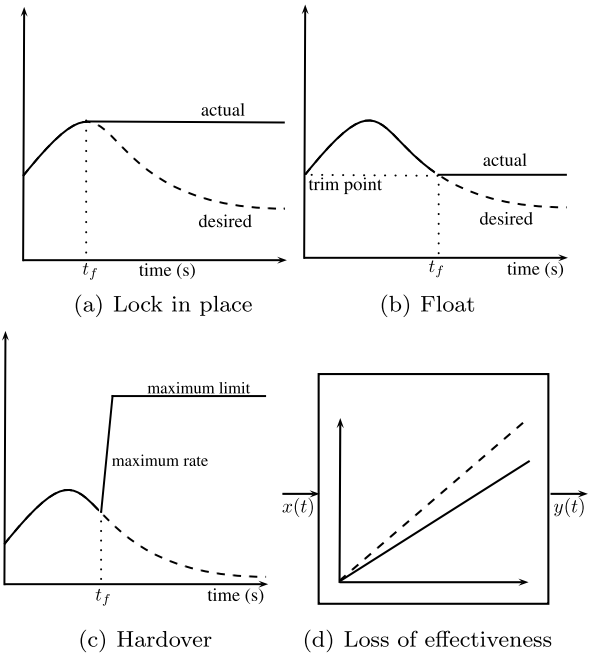

Fig. 2.1 Type of fault and failures on actuator (figure adapted from [32])

becomes stuck and immovable. This might be caused by a mechanical jam, due to lack of lubrication for example. This type of failure is considered in [53, 92, 105, 111, 292] and occurred in documented incidents such as flight 1080 (Lockheed L-1011, San Diego, 1977) [41] where one of the horizontal stabilisers jammed in the full trailing edge-up position; and flight 96 (DC-10, Windsor, Ontario, 1972) [41] where the rudder jammed with an offset.

A *float failure* is a failure condition whereby the control surface moves freely without providing any moment to the aircraft. An example of a float failure is the loss of hydraulic fluid. Examples of research considering float type failures are [42, 92, 105], and it has occurred in incidents such as Flight 123 (B-747, Japan, 1985) and DHL A300B4 (A300, Baghdad, 2003) [41] resulting from a total loss of hydraulics.

One of the most catastrophic types of failure is *runaway/hardover*. In a runaway situation the control surface will move at its maximum rate limit until it reaches its maximum position limit or its blowdown limit.[1] For example, a rudder runaway can occur when there is an electronic component failure which causes an uncommanded large signal to be sent to the actuators causing the rudder to be deflected at its maximum rate to its maximum deflection at low speed (or its blowdown limit at high

[1] A blowdown limit is an aerodynamic limit of the control surface deflection at a specified speed which overpowers the movement of the actuator. The blowdown limits might not be the maximum physical deflection of the control surface. Any deflection above the blowdown limit can cause structural damage [240] as it imposes the maximum physical and structural limit the control surface and the surrounding structure can have.

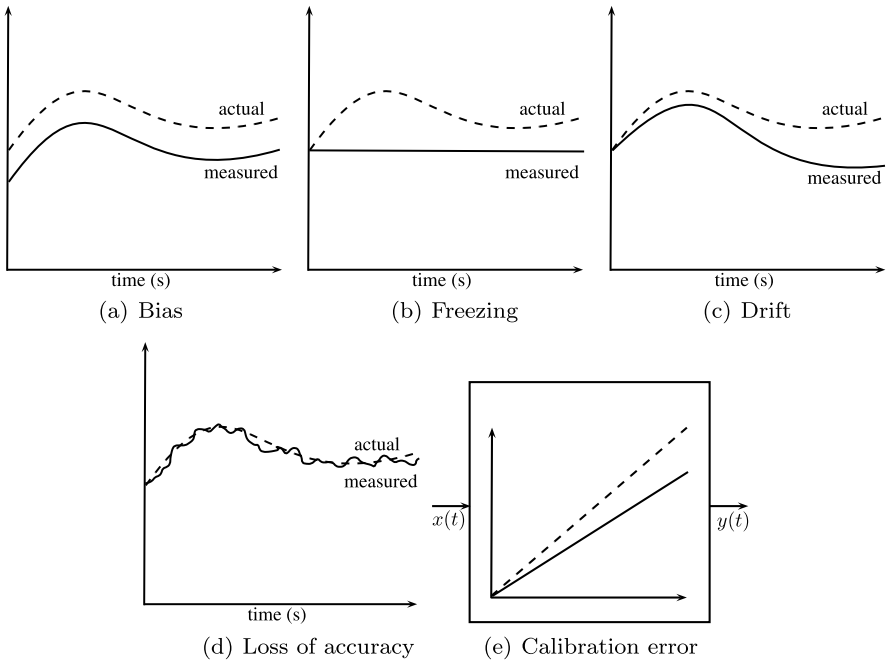

Fig. 2.2 Type of fault and failures on sensor (figure adapted from [32])

speed). This type of failure is considered in [234] and occurred in incidents such as Flight 85 (B-747, Anchorage, Alaska, 2002) [41] (which suffered a lower rudder runaway to full left deflection, causing the airplane to roll excessively) and flight 427 (B-737, Near Aliquippa, Pennsylvania, 1994) [9] (which suffered from a rudder runaway to its blowdown limits).

Note that the above faults and failures are related to the aircraft's control surfaces. Another type of fault that occurs in aircraft, is *structural damage*. Structural damage may change the operating conditions of the aircraft due to changes in the aerodynamic coefficients or a change in the centre of gravity. This constitutes a change to the dynamics of the system. Examples of failures that cause structural damage are wing battle damage [30], detachment of control surfaces, for example the rudder (flight 961, A310, Varadero, Cuba, 2005) [3] or engines (flight 1862, B-747, Amsterdam, 1992) [235], or detachments of some body parts of the aircraft e.g., the vertical fin/stabiliser (Flight 123, B-747, Japan, 1985) [41, 109] and (flight 587, A300, New York, 2001) [41], wing (DHL A300B4, A300, Baghdad, 2003) [41], fuselage skin or cargo doors (flight 981, DC-10, Paris, 1974) [41].

Figure 2.2 describes some typical sensor faults in aircraft. Bias is a constant offset/error between the actual and measured signals. Sensor drift is a condition whereby the measurement errors increase over time (and might be due to loss of sensitivity of the sensor). Loss of accuracy occurs when the measurements never reflect the true values of the quantities being measured. Freezing of sensor signals result in the sensor providing a constant value instead of the true value. Finally,

a calibration error is a wrong representation of the actual physical meaning of the quantity being measured from the electrical or electronic signals that emerge from the sensor unit itself.[2] Sensor faults/failures can occur due to malfunctions in the components in the sensor unit, loose mounting of the sensors and loss of accuracy due to wear and tear. An example of an incident resulting from sensor failures occurred in flight 124 (B-777, Perth, 2005) [2] which caused a flight control upset and contributed to the violent behaviour of the aircraft which necessitated the auto pilot and navigation unit being switched off.

It is interesting to mention that faults/failures can also be categorised in terms of time [247]. Abrupt faults or failures exhibit sudden and unexpected changes and are usually easily noticed by the pilot. An example of an abrupt failure is an actuator jam, or a hardover. Incipient faults, for example a slow drift in a sensor, are more subtle and the effect is not so obvious. However, incipient faults if left unattended for a long period of time might degrade the required performance of the system and might lead to abrupt and catastrophic failures. Incipient faults can be caused by operational wear and tear as the effect is negligible but becomes gradually worse before it fails abruptly.

2.2 Fault Tolerant Control: General Overview

In the literature, most of the motivation and research work in fault tolerant control involves solving problems encountered in safety critical systems such as aircraft. Applications can also be found in other systems, for example robots [169], space systems [252] and underwater remotely operated vehicles (ROV) [198]. Patton in [206] stated that

> ... Research into fault tolerant control is largely motivated by the control problems encountered in aircraft system design. The goal is to provide a 'self-repairing' capability to enable the pilot to land the aircraft safely in the event of serious fault ...

Zhang and Jiang [291] define

> ... fault tolerant control systems (FTCS) as control systems that possess the ability to accommodate system component failures automatically. They are capable of maintaining overall system stability and acceptable performance in the event of such failures. FTCS were also known as self-repairing, reconfigurable, restructurable, or self designing control systems ...

FTC is a complex combination of three major research fields [206], FDI, robust control, and reconfigurable control (see Fig. 2.3). Patton [206] also discusses the relationship between these fields of research. A typical active fault tolerant control systems (AFTCS) architecture is shown in Fig. 2.4. For most FTC schemes, when a fault/failure occurs either in an actuator or sensor, the FDI scheme will detect and locate the source of the fault. This information is then passed to a mechanism to

[2] Sensors, most of the time, provide measurements in terms of current or voltage and therefore require transformation to represent the actual physical meaning of the quantities being measured.

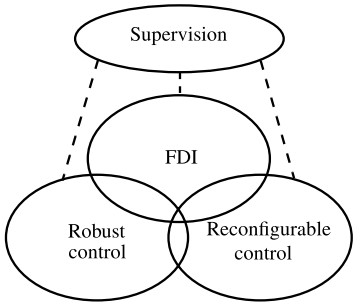

Fig. 2.3 Scattered areas of fault tolerant control research (figure adapted from [206])

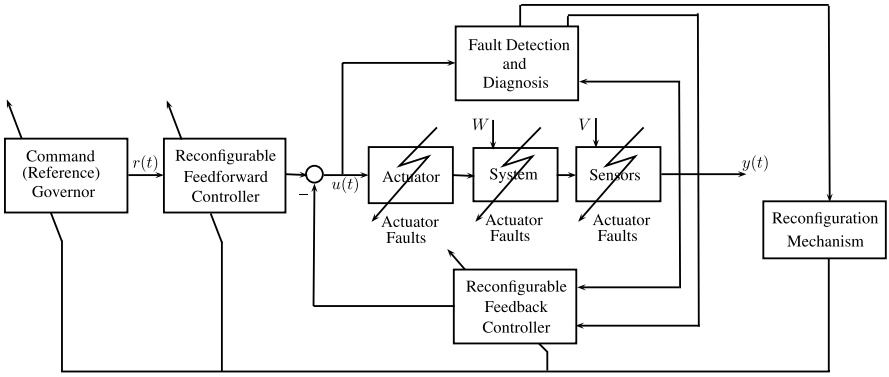

Fig. 2.4 General structure of active fault tolerant control systems (AFTCS) (figure adapted from [294])

initiate reconfiguration. The reconfigurable controller will try to adapt to the fault, therefore providing stability and some level of performance. Both the FDI and the reconfigurable controller need to be robust against uncertainties and disturbances.

Robust control is closely related to passive fault tolerant control systems (PFTCS) [206]. The controller is designed to be robust against disturbances and uncertainty during the design stage. This enables the controller to counteract the effect of a fault without requiring reconfiguration or FDI. In some robust methodologies, fault tolerant capability is limited, and importantly total, actuator failures cannot be handled. Some widely referred to surveys on FTC and FDI are [132, 143, 178, 206, 291, 293] and [51, 136]. Also there are recent publications (books and edited monographs) such as [24, 25, 47, 75, 80, 179, 199] in the field of FTC and [26, 69, 134, 135, 159] for FDI.

Zhang and Jiang [291] gives a good bibliographical review of reconfigurable fault tolerant control systems. The paper also proposes a classification of reconfiguration methods which is based on the mathematical tools used, the design approach and the way of achieving reconfiguration. It also provides a bibliographical classification based on the design approaches and the different applications. Open problems and current research topics in active fault tolerant control systems (AFTCS) are also discussed.

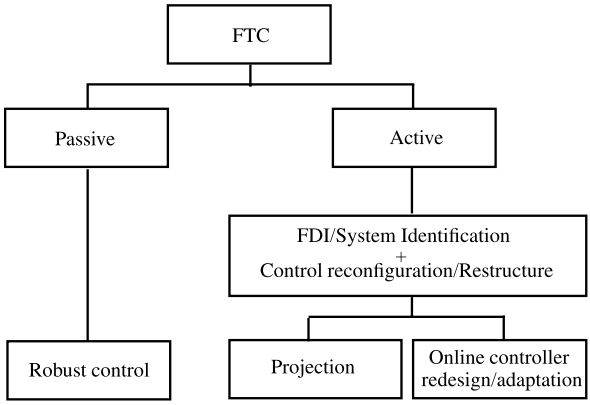

Fig. 2.5 Classification of FTC by [206]

Zhang and Jiang [291] and Patton [206], classify FTC into two major groups (see Fig. 2.5): passive fault tolerant control systems (PFTCS) and active fault tolerant control systems (AFTCS). In passive fault tolerant control systems, the controller is designed to be *robust against faults* and uncertainty. Therefore when a faults occurs, the controller should be able to maintain stability of the system with an acceptable degradation in performance. PFTCS does not require FDI and does not require controller reconfiguration or adaptation. AFTCS on the other hand responds to system component failures in an 'active' way by reconfiguration so that stability and acceptable performance of the entire system can be maintained [291]. Therefore, most AFTCS require FDI to provide the fault or failure information before reconfiguration can be undertaken.

Other surveys on reconfigurable control appear in [132] and [143]. The report [132], gives insight into many methods used for reconfigurable control for flight applications, while [143] gives a survey on reconfiguration methods used specifically for FTC in flight control applications. Table 2.1 presents a brief comparison of the FTC methods [143]. Note that in Table 2.1, the expression 'fault model' refers to the assumption that the faulty system is available and used in the design process. 'Actuator constraints' refers to the ability of the controller to handle actuator limits.

2.3 Redundancy

Redundancy can be categorised into two types; direct and analytical. In direct redundancy, actual physical hardware redundancy is available. In terms of sensors, two or three sensors that measure the same quantity is called double and triple redundancy. In normal operation, only one sensor is sufficient, however, two or three sensors are required to ensure reliable measurements in the case of faults. A voting system is a typical way to decide which channels are working correctly and which are faulty [112]. This hardware redundancy concept can also be extended to the actuators.

In terms of analytical redundancy, instead of having multiple sensors that measure the same signal, an observer that provides an estimate of the signals of inter-

Table 2.1 Current FTC methods comparison (table adapted from [262])

Method	Actuator failures	Structural failures	Robust	Adaptive	FDI	Fault model assumed	Actuator constraints	Linear model	Nonlinear model
Multiple model switching and tuning (MMST)		●		●	●			●	
Interactive multiple model (IMM)		●		●	●		○	●	
Propulsion controlled aircraft (PCA)	●		○			●		●	●
Control allocation (CA)	●					●	○	●	
Feedback linearisation	●	●		●	●				●
Sliding mode control (SMC)	○[a]	●	●[b]				●		●
Eigenstructure assignment (EA)		●				●		●	
Pseudo-inverse method (PIM)		●				●		●	
Model-reference adaptive control (MRAC)		●		●	●			●	
Model predictive control (MPC)	●	●	○	○	●	●	●	●	●

● means that the method has the property

○ implies that an author suggested that the approach could be modified to incorporate the property [143]

[a]SMC can handle partial loss of effectiveness of actuators but not complete loss

[b]SMC assumes robust control can handle all forms of structural failures

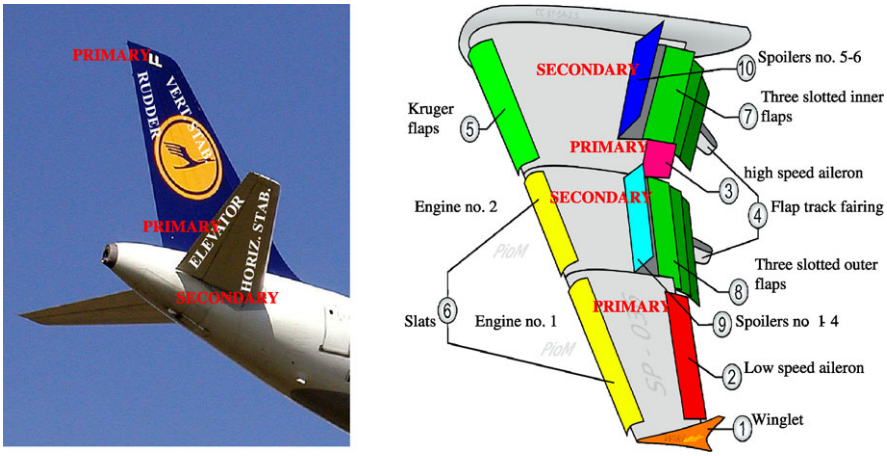

Fig. 2.6 Large transport aircraft: typical control surfaces (figures: Arpingstone and Piotr Jaworski via Wikipedia)

est provides analytical redundancy. There is no actual additional hardware implemented, instead some algorithm or mathematical model or observer runs in the control computer. This is desirable in many systems especially in aircraft and unmanned air vehicles (UAVs), since analytical redundancy eliminates the requirements for extra hardware therefore reducing weight and cost.

The development of new safety critical systems such as the re-entry vehicle [137, 138] allows the possibility of building in redundancy during the design process [206]. For many systems, however, the challenge is to use the existing available sensors and actuators to deal with faults/failures. In large transport aircraft, redundancy is already available in abundance. Even though it is not meant for the purpose of FTC, the use of these extra control surfaces provides the possibility of using them to obtain the same effect as the original control surface e.g., horizontal stabilisers can be used if elevators fail.

In large passenger transport aircraft, sensors are typically triple redundant [34, 35]. In view of the aerospace industry's attempts to reduce the 'carbon footprint' left by aircraft, many manufacturers have tried to reduce the consumption of fuel by designing high efficiency engines, and also by reducing weight by eliminating hardware redundancies, replacing them with analytical ones (observers to estimate the aircraft states). This is also beneficial in the development of cheap, robust and maintenance-free UAVs. Due to the low production cost, there is no requirement for repair, and instead, the whole unit is replaced.

In aircraft, a control surface, for example the rudder, can have three different hydraulic actuators running from three separate lines to three independent hydraulic pumps [35]. This means most control surfaces will have triple redundancy. In terms of the control surface itself, there exist secondary control surfaces that can be used in an emergency in an unconventional way to achieve the same effect as the primary control surface (see Fig. 2.6). In large passenger transport aircraft for example, the spoilers which are typically deployed to reduce speed can also be used differentially

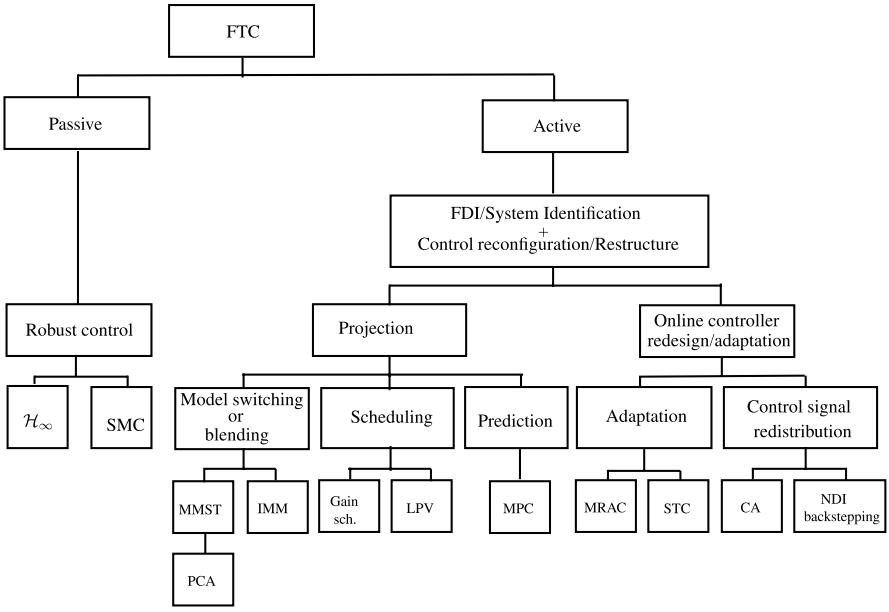

Fig. 2.7 Classification of FTC—how FTC is achieved

to create roll which normally is achieved by using ailerons; also engines can be used differentially to create yaw, which is typically achieved by using the rudder; and finally the horizontal stabiliser (see Fig. 2.6), which is normally used to set the angle of attack, can also replace elevators for pitch modulation.

2.4 Fault Tolerant Control

Figure 2.7 gives a general overview of how FTC is achieved. The top level of the tree diagram is based on the one proposed in [206]. The lower level is based on the different approaches for achieving FTC discussed above.

Passive FTC is usually based on robust control ideas and therefore handles faults/failures without requiring information from an FDI scheme. Active FTC (AFTC) in general requires some information on the faults/failures that occur and therefore typically FDI is required. AFTC can be divided into two sub-groups: projection type FTC; and online reconfiguration/adaptation. In projection based FTC, controllers are designed a priori for all possible faults/failures that might occur in the system. The projected controller will only be active when the corresponding fault/failure occurs. Projection based FTC is sub-divided into three categories which are model switching or blending, scheduling and prediction. AFTC is based on reconfiguration or online adaption. Here, two further sub-components have been proposed: FTC which is achieved through adaptive control; and FTC which can be

achieved through redistributing the control signals (control allocation). A discussion of these different strategies is provided in the following subsections.

2.4.1 Adaptation

Motivated by the design of autopilots for high performance aircraft in the 1950s, adaptive control was proposed as a way of dealing with a wide range of flight conditions [230]. Adaptive control is used in order to automatically adjust the controller parameters to achieve the desired performance. There are two approaches in adaptive control: so-called direct and indirect adaptation [14, 76, 143]. In indirect adaptation, there are two stages in designing the controller. First, the system parameters need to be estimated. In the case of linear systems, the matrix pair (A, B) needs to be estimated due to changes in the operating conditions e.g., faults/failures. The next step in the indirect adaptation approach is to use this information to design the controller. In the direct adaptation approach, the controller is designed directly without estimating the system parameters.

Model-reference adaptive control (MRAC) and self tuning control (STC) [230] are two popular methodologies. In self tuning control, online parameter estimation is required for the controller adaptation. Meanwhile, in the MRAC, the unknown parameters are not perfectly estimated, but rather are tuned and adjusted so that the output of the plant follows the desired trajectory (the output of the reference model) by making the tracking error converge to zero.

2.4.2 Switching or Blending

The idea of using multiple models for reconfigurable control was introduced in the early 1990s [196]. Multiple model schemes have been motivated by the problem of coping with changes in operating conditions and varying flight envelopes. Most early classical control methods were based on linear methods, and multiple model schemes seemed an ideal extension to solve the problem of changing operating conditions. When implementing on a real system, usually linear controllers need to adapt to changes in operating conditions since the controller is only guaranteed to be stable near the linearisation condition. Therefore using multiple model schemes is one way to ensure that the controller can be designed so that stability and performance can be guaranteed for a wide flight envelope.

From an FTC point of view, the ‘bank’ of controllers acts as a backup and ‘offline’ dormant controllers are only activated when faults for which a particular controller is designed, occurs. This method depends on the FDI scheme providing the correct information on the type and location of the faults/failures to enable the correct controller to be switched on. The ‘bank’ of models must contain all the possible faults and failure modes. An FDI scheme can be created by comparing the current

plant states and the outputs of all models in the bank [30]. Essentially, the model with the smallest error is the 'nearest' model to the actual plant and therefore its associated controller can be switched on. More elaborate descriptions of the switching rule are available in [195, 196]. These papers discuss the stability of the switching schemes as well as the performance after a switch has occurred. In [196], a review of the most recent development in MMST has been presented.

The switching between models and controllers sometimes introduces undesired transients. Therefore bumpless transfer methods [48, 81] are sometimes needed to reduce the effect. Another disadvantage is that some faults that occur are not predicted a priori. For example, in several flight incidents, unthinkable failures have occurred e.g., the Bijlmermeer incident in Amsterdam [233], where two engines detached from the right wing and caused unforeseeable effects on the aerodynamics of the aircraft due to the damaged airframe. A significant disadvantage of this method is its dependency on the robustness of the FDI scheme to identify the correct model and controller pair to be activated. Another disadvantage highlighted in [143], is the scheme's inability to handle multiple faults/failures. The survey in [143], gives a brief introduction to MMST. More detailed descriptions can be found in [195, 196]. The application of multiple model ideas in terms of FTC for aircraft systems can be found in [30, 111] and recently in [10].

Even though MMST can be used to tackle the problem of varying operating conditions, in some cases, to obtain a linear model that exactly matches the varying plant is hard to achieve; since hundreds (if not thousands) of linear models and controllers are needed to match every possible flight condition including faults/failures. In the 'Interacting Multiple Model' (IMM) [214, 290] approach, the idea is to obtain a set of linear models based on a few carefully chosen flight conditions and to design linear controllers at these selected operating conditions. When the operating conditions change (or faults/failures occur), an estimated plant output or control input is obtained by blending the predetermined models.

The main assumption used in IMM is that every possible flight condition including faults/failures can be modelled as a convex combination of the predetermined linear models. The second step is to obtain a control signal based on a blend of predefined controllers [290] or online control law calculations using the probability weight provided by the IMM estimator.

In the first step of the IMM scheme, [290] and [214] proposed the use of a bank of Kalman filters to calculate the probability of the individual faults/failures. This probability is also used to obtain a weighted average of each predefined linear model to estimate the state of the plant. In the second step, a bank of controllers is predesigned based on the anticipated faults/failures that might occur [214, 290]. The idea is that during faults/failures, the eigenvalues of the closed-loop system need to be as close as possible to the nominal no fault conditions. The reconfiguration of the controller comes from the online use of the probability weighted average to determine the blending ratio for the control input from the predefined/projected controllers when a fault or failure occurs [214, 290].

In [290], it has been shown that system faults[3] can be handled. In comparison to the MMST, the IMM has the ability to cope with non-anticipated faults/failures [214, 290]. One problem of IMM schemes is finding the right balance of blending/probability weights to get the best model match. The IMM method is also heavily dependent on the FDI scheme to correctly identify the faults/failures. Details of IMM schemes can be found in [149]. In [290], an integrated IMM approach is discussed where both FDI and FTC are integrated. The application of IMM to an Eagle-Eye UAV can be found in [214].

2.4.3 Prediction

Unlike many other control paradigms which came ostensibly from the academic community, the development of predictive control/model predictive control (MPC) was initiated in the process industry. This is due to the fact that the concept and the mathematical description is easy to understand by most control engineers in industry. Therefore it is no surprise that (other than classical PID control), MPC is the most widely used and implemented method in the process control industry [176].

The original idea for MPC is to allow the production process to run as close as possible to the process limits (both physical and safety) without violating any of the limits, in order to maximise production and therefore profit. The main benefit of MPC is its ability to handle limits and constraints. This is the main motivation for the study of MPC for flight control and especially FTC. Examples of MPC in the field of flight control and FTC can be found in [177, 178] respectively. During faults/failures, especially to the actuators, the remaining actuators will be driven to their limits [177]. MPC has the ability to handle this situation by including these limits in the optimisation process which is used to obtain the control signals. Structural damage can also be handled in MPC by modifying the internal/reference model [177].

MPC is an iterative control algorithm based on optimal control. The objective is to obtain predicted state trajectories in the future using the current states and the computed control signals. However only the first control signal from the optimisation is applied to the real actuators. Then the states are sampled again and the calculations are repeated. MPC is also known as receding horizon control [176, 178].

MPC in its most powerful form requires an online solution to the constrained optimisation problem [176]. However, with the current state of computer technology, online optimisation is still hard to achieve for systems requiring fast responses—such as aircraft. As in most FTC strategies, MPC is dependent on reliable FDI to provide information on the faulty system. In the case of actuator faults, the behaviour of the faulty actuator is needed from the FDI scheme so that a new constraint

[3] In terms of linear methods, system faults are the ones that affect the A matrix i.e., airframe or wing damage.

can be included in the optimisation process. In terms of tuning for flight control systems, there is still a lack of transparency in the design process [178], which typically requires trial and error and experience. The major benefit of MPC is that it can handle actuator constraints, and this has provided motivation for the study of MPC in flight control and FTC [178].

2.4.4 Control Signal Redistribution

The idea of the pseudo-inverse method (PIM) is to design a controller such that the poles of the system subject to a fault/failure condition will be as close as possible to the nominal closed-loop poles. The following equations give insight into the PIM method. Consider a linear system given by

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{2.1}$$

Assume that a state feedback gain F has been designed, and the control law is defined as

$$u(t) = Fx(t) \tag{2.2}$$

and therefore the closed-loop system is given by

$$\dot{x}(t) = (A + BF)x(t) \tag{2.3}$$

During faults/failures, the closed-loop faulty system can be represented by

$$\dot{x}_f(t) = (A_f + B_f F_f)x_f(t) \tag{2.4}$$

The idea is to obtain a gain matrix F_f so that the faulty system closed-loop performance will be as close as possible to the nominal (2.3) one. The plant matrices A and B and the gain F is assumed to be known a priori. The faulty system (A_f, B_f) can be obtained from online system identification or from FDI: then in principle, F_f can be obtained online. For a non-square B_f matrix, the pseudo-inverse of B_f provides some degrees of freedom. In [187], these degrees of freedom were used to redistribute the control commands in order to improve the closed-loop system stability [206].

Even though the concept is quite simple and easy to understand, the PIM has several drawbacks which hindered its further progress. As argued in [132, 143, 206, 284], the main drawback is the lack of a stability analysis. The other drawbacks highlighted by [206] and [284] are associated with the assumption that the state measurements are always available. Meanwhile [284] highlights the problem of lack of robustness when the system pair (A_f, B_f) from the system identification is not perfectly known.

Some suggestions are given in [106, 132, 206] for improving the PIM method. In [106], the concept of modifying the PIM (MPIM) is discussed. It is based on the combination of PIM with the theory of robust stability of systems with structured uncertainty [132, 206]. In [132], a bank of pre-computed F_f matrices, which ensure

a stable closed-loop system for all possible faults, was suggested. In [284], a robust 'control mixer' which relates to the IMM method was proposed. It is also interesting to point out the resemblance between the PIM approach and model following methods [206], where the closed-loop system is forced to follow pre-specified desired closed-loop dynamics.

In its early development, the idea of redistributing the control signals to the remaining healthy actuators was called 'restructuring' [206]. An early example is given in [131], where a 'restructuring controller' utilising a 'control mixer concept' is used to redistribute the control signals. Due to some drawbacks, this restructuring controller was not explored in the 90's. It has re-emerged in recent years as *control allocation* partly because of the development of high performance, highly redundant aircraft (such as [31, 37, 225, 269]) and improvements in computational power (which is necessary in order to solve online optimisation problems [20, 29, 77, 89, 144]).

Control allocation (CA) has the capability of redistributing the control command signals to the actuators especially during faults/failures. One major difference between CA and PIM is that in CA, the controller is designed based on a 'virtual control' signal and the CA element will map the virtual control to the actual control demand to the actuators. The benefit here is that the controller design is independent of the CA unit. Therefore, CA can be used in conjunction with any other controller design paradigm. Papers such as [122, 222] represent some of the recent work in this area.

CA has the capability of managing the actuator redundancy that exists in passenger aircraft [35] and modern fighter aircraft [97]. Not only is CA beneficial for FTC (see for example [38, 63]), it has also been used for different control strategies i.e., optimally using the actuators to reduce drag and increase efficiency. There is extensive literature on CA which discusses different algorithms, approaches and applications. Reference [89] discusses two (broadly) linked approaches (linear and quadratic programming) based on finding the 'best solution' to a system of linear equations. The work in [122] compares control allocation with optimal control design for distributing the control effort amongst redundant actuators. The authors in [39] demonstrate that feedback control systems with redundant actuators can be reduced to a feedback control system without redundancy using a special case of CA known as 'daisy chaining'. In this approach, a subset of the actuators, regarded as the primary actuators are used first, then secondary actuators are used if the primary actuators reach saturation. Other CA approaches which take into account actuator limits (using constrained optimisation) are discussed in [29, 31], while [121] discusses frequency weighted CA.

From an FTC point of view, the benefits of CA are that the controller structure does not have to be reconfigured in the case of faults and it can deal directly with total actuator failures without requiring reconfiguration/accommodation: the CA scheme automatically redistributes the control signal. As in MPC, another major benefit of CA is that actuator limitations can be handled by including the actuator constraint in the optimisation process. One of the drawbacks of CA is that, for linear systems, the pure factorisation of the input distribution matrix is a very strong

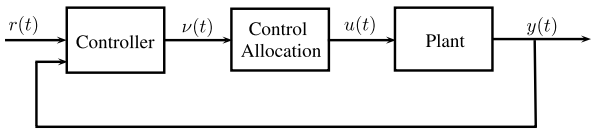

Fig. 2.8 Control allocation strategy

requirement and therefore some approximations have been made [38, 63, 122, 127]. In the case of optimal control surface deflection, linear or quadratic programming is required which is difficult to achieve online in real-time due to the requirements of high computational power during the optimisation process. There are only a few reported examples in the literature which have successfully implemented control allocation in real-time (see for example [63]).

CA occurs naturally in nonlinear methods like feedback linearisation and backstepping [20, 121, 225, 226]. It is based on separating the control law from the control allocation task (see Fig. 2.8). This is done by designing a controller to provide a 'virtual control' which will be mapped to the actual control signals sent to the actuators. Examples of the application of CA are given in [20, 37, 63, 236], while papers such as [75, 78, 122, 133, 144, 203, 222, 226, 263, 295] consider CA as an FTC strategy.

2.4.5 Robust Control ($\mathcal{H}_\infty$ Control)

Passive fault tolerant control relies on the robustness of the underlying control design paradigm. One of the most popular robust control methodologies developed during the 1980's was $\mathcal{H}_\infty$ control. It has become one of the most developed methods for multivariable control [178], with many applications ranging from industrial process control to aircraft control problems. Most robust control approaches do not require any information on faults and therefore work in nominal as well as in faulty conditions. The ability to deal with faults depends on the predesigned controller which is based on minimising the effect of uncertainty or disturbances on the system [178].

One disadvantage of $\mathcal{H}_\infty$ is the fact that in some cases, the controller is conservative in the nominal conditions in order to guarantee stability in the event of faults, the performance in the nominal condition is sometimes sacrificed for robustness. Another drawback is that the final controller is usually of a higher order than the system. In some cases model reduction is required to truncate the order of the controller (page 339 [178]). In the field of FTC, papers like [184] and the chapters in [178] describe some of the research in the area of flight control.

The theory of $\mathcal{H}_\infty$ control system design can be extended to Linear Parameter Varying (LPV) systems. Some general papers on Linear Parameter Varying (LPV) are [16, 204, 221, 273]. In the field of FTC, papers such as [105, 185] represent some of the research work in this area. Both of these papers have considered a LPV approach for dealing with faults/failures in a civil aircraft benchmark problem. The most recent LPV papers in the field of FTC are [200, 215, 223, 224].

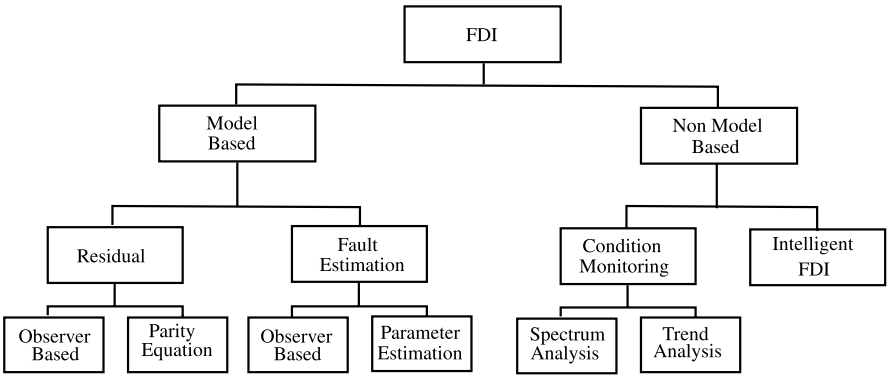

Fig. 2.9 FDI classification

2.5 Fault Detection and Isolation

In active FTC, FDI plays a vital role in providing information about faults/failures in the system to enable appropriate reconfiguration to take place. The main function of FDI is to detect a fault or failure and to find its location so that corrective action can be made to eliminate or minimise the effect on the overall system performance [247]. The IFAC technical committee, as stated in [136], makes the following definitions.

fault detection: determination of the faults present in a system and the time of detection.
fault isolation: determination of the kind, location and time of detection of a fault.
fault identification: determination of the size- and time-variant behaviour of a fault.

The interconnection of FDI with FTC is discussed in [206, 291, 293]. For most AFTC systems, the robustness of the FDI has a strong effect on the robustness of the FTC and this is discussed in [206, 289, 293]. Since the appearance of drive and fly-by-wire technology, there has been an increase in analytical redundancy. In analytical redundancy methods, the measured signals are compared to a mathematical model. The benefit of using analytical redundancy is clear: there is no need for redundant hardware to be installed, therefore reducing weight and cost. This is very useful for energy and weight reduction critical systems such as satellites and spacecraft.

There are many classifications of FDI in the literature [51, 136]. One obvious classification is model- and non-model-based FDI. In this book the emphasis will be on model-based FDI. Model-based FDI schemes can be grouped into two major categories; FDI using residual schemes and FDI which has the capability to estimate the faults.

Figure 2.9 represents a possible classification for FDI. The model-based residual classification is obtained from [51]. A brief description of a few key model-based FDI schemes are given below.

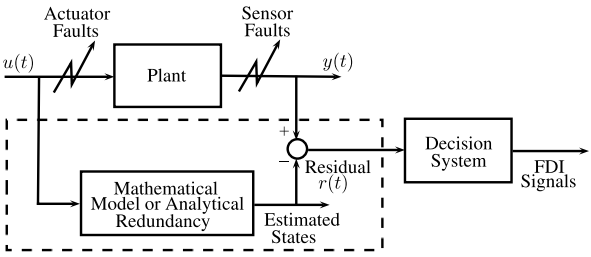

Fig. 2.10 Residual-based FDI

2.5.1 Residual-Based FDI

In residual-based FDI, signals from a mathematical model and hardware measurements are compared and the filtered difference forms a residual signal [51] (see Fig. 2.10). In nominal fault-free conditions, the residuals should be zero, and nonzero when faults/failures occur. This residual signal is usually applied with a threshold to avoid false alarms from disturbances or uncertainty. When the residual signal exceeds the threshold, a fault is said to occur. Usually in residual generation, a fault is detected and its location identified, but there is no further information on the fault.

A good deal of research has been focussed on residual-based FDI using different methods for various applications. In particular, [51] provides an excellent discussion on model-based residual FDI schemes covering all aspects including basic principles and robustness issues.

There are many benefits of using residual-based FDI. Most residual-based FDI systems are easy to understand and implement with many mature topic areas and examples of applications in the literature. For many systems, detection and isolation of the fault is sufficient to trigger the reconfiguration for FTC. For example multiple model controllers will switch on a particular controller when the designated failure occurs to the actuators or sensors based on the information about the location of the fault. However, for some FTC schemes, detecting and isolating the faults is not sufficient. Some FTC schemes require further information about the nature and behaviour of the fault.

2.5.2 Fault Identification and Reconstruction

Some FTC methods proposed in [275, 289, 296], require estimates of the actuator efficiency to allow the schemes to tolerate the faults/failures (see Fig. 2.11). In terms of sensor fault FTC, if the sensor fault can be estimated/reconstructed, this information can be used directly to correct the corrupted sensor measurements before they are used by the controller. This avoids reconfiguring or restructuring the controller. This is one aspect that will be considered in the later chapters of this book.

The Kalman filter is probably one of the most well known and used methodologies in industry. Conceived in the 1960's by Rudolf Kalman and made famous

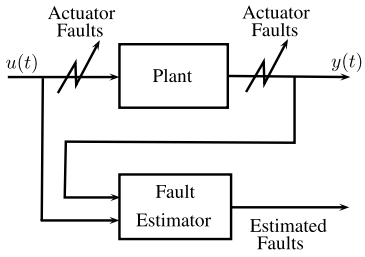

Fig. 2.11 Fault estimation based FDI

by its application in the NASA Apollo space program, the Kalman filter has found applications in many engineering systems (e.g., navigation, tracking targets such as aircraft and missiles using radar) as well as other fields such as economics.

A Kalman filter as summarised in [156], is an optimal estimator based on indirect, inaccurate and uncertain observations. It is recursive so that new measurements can be processed as they arrive. If all noise is Gaussian, the Kalman filter minimises the mean square error of the estimated parameters and therefore is optimal. Since its famous application in the Apollo space program, the Kalman filter has continued to be popular especially in industry for a number of reasons: (a) Kalman filters provide fairly accurate results in most applications due to its optimality and structure, and (b) Kalman filters have a recursive form and are suitable for online real-time digital processing and are easy to formulate and implement.

Standard Kalman filters act as an observer and therefore can be used to detect faults or failures by creating residual signals from comparing the actual and estimated outputs. The basic concept of the Kalman filter has been 'upgraded' to enable many applications, such as the extended Kalman filter for nonlinear systems [167] and for parameter estimation [107, 115] in which the parameters to be estimated are incorporated into the formulation as augmented states. Often this introduces bilinearities which can be overcome by the use of extended Kalman filters.

The Kalman filter can also be composed into a bank of Kalman filters [157, 158] or interacting multiple model Kalman filters (IMM-KF) [214, 290] in order to create a residual which can be used for fault detection. The IMM-KF uses the same IMM as used for controller reconfiguration which was discussed in Sect. 2.4.2. The Kalman filter also has been combined with the receding horizon (predictive control) method as shown in [162], which has the potential for fault diagnosis.

Another variant called the two stage Kalman filter [150, 151] has the ability not only to detect and isolate faults, but to estimate the effectiveness levels of actuators [140, 274, 275, 289]. This capability is a bonus for FTC schemes which depend on the effectiveness level of the actuator for reconfiguration [224, 296].

The early papers by Kalman can be found in [146, 147] whilst the most cited books and references are [13, 186, 237].

Using the same principles as those used for designing $\mathcal{H}_\infty$ controllers, an observer can be designed as a basis for a residual-based FDI scheme [185]. The idea is to allow the residual to be sensitive only to faults and robust against disturbances, modelling errors and noise [185]. This can be achieved by selecting the observer

gains (using LMI formulations for example) which minimise the $\mathcal{H}_\infty$ norm between the uncertainty and the residual signal. Fault detection filters using $\mathcal{H}_\infty$ techniques are amongst the most popular and mature FDI schemes in the literature [130] with many applications in industry including aerospace [184, 185, 244]. Apart from the $\mathcal{H}_\infty$ optimisation technique, other frequency domain design approaches for model-based FDI, including μ synthesis, are discussed in [51] and [98].

Applications of $\mathcal{H}_\infty$ for robust detection of faults can be found in [184, 185]. In [184] an integrated design of both controller and observer is considered. The integrated design proposed in [185] gives some insight on designing a controller that is not only robust against actuator faults but also considers the robustness properties of the FDI scheme in the design of the controller.

The early $\mathcal{H}_\infty$ methods have recently been extended to linear parameter varying models. It has been claimed that even though there are various FDI approaches for LTI, LTV and bilinear systems, there are only a few available methods for LPV systems [27]. Therefore, the focus of the work in [27] was to introduce FDI based schemes for LPV systems using an extension of the approach called the fundamental problem of residual generation. Other recent LPV papers in the field of FDI are [11, 28, 114, 116, 153, 168, 200, 245, 267]. FDI based on LPV systems has inherent performance and stability guarantees for a range of operating conditions compared to multiple model or gain schedule based FDI.

2.5.3 Parameter Estimation

Parameter estimation schemes provide a means of updating the system's parameters online in real-time and for controller reconfiguration. Parameter estimation is one of many methodologies which have been applied to aircraft. Aircraft contain many parameters (especially aerodynamic coefficients) which change, based on the operating conditions. These parameters are typically pre-estimated offline through wind tunnel and flight test before being used for modelling or control design. However, during faults/failures (especially structural damage, such as wing damage or missing fuselage/skin), no accurate pre-estimate is available and therefore these aerodynamic coefficients need to be obtained online.

Examples of parameter estimation methods appear in [107, 115] which use Kalman filters, and [57, 172] which use the two step method. In the two step method (TSM), the original state-parameter estimation problem is decomposed into a state estimation one and a subsequent linear parameter identification sub-problem [57, 192]. Other sources of information on parameter estimation of aircraft systems can be found in [180, 181, 193, 201, 260].

In [115], parameter estimation based on an extended Kalman filter is used for FDI in an automotive engine. One of the most recent papers for aircraft FDI is [144]. This paper proposes the use of online parameter estimation provided by the two step method [172] (which identifies and estimates the current aircraft parameters which change due to structural damage). Here, not only are the changes to the aerodynamic

coefficients used to detect faults/failures in the system, they are also used as part of the reconfiguration to achieve fault tolerance.

In most parameter estimation methods, in order to get good estimates, it may be necessary to introduce perturbation signals to make sure that all the plant modes are sufficiently excited [206]. For this reason, most parameter estimation methods work best in the presence of wind and gusts. However, in many practical applications, it is hard and not advisable to apply additional perturbation signals, especially when faults/failures or structural damage has occurred in the system.

2.5.4 Non-model-Based FDI (Intelligent FDI)

One of the main issues associated with model-based designs is the availability and quality of the model. Errors resulting from imperfect or inaccurate models will affect the performance of the fault diagnosis scheme [208, 210]. The use of robust model-based methods usually results in a design which is too conservative and insensitive to faults, too complicated or limited to certain classes of uncertainty [208]. Since the late 1990s there has been an increase in research on non-model-based FDI methods—especially those utilising artificial intelligence and 'soft computing' approaches such as neural networks, and fuzzy logic (see for example [26, 159, 160, 209, 212, 272]).

In [208], a combination of numerical (quantitative) and symbolic (qualitative) knowledge of the system in a single framework has been proposed. The idea was inspired by earlier work which uses observers for residual generation and fuzzy logic for decision making. The underlying concept is to structure the neural network in a fuzzy logic format which allows residual generation (through the rapid and correct training of the neural network to model the nonlinear dynamics of the system) and evaluation and diagnosis of the fault (through fuzzy logic). In [160], neuro-fuzzy modelling and diagnosis is considered with the addition of an adaptive threshold in the fault detection scheme, to achieve some level of robustness.

One of the benefits of using the intelligent approach, especially neural networks for FDI is its ability to model any nonlinear function [208]. In terms of FDI, neural networks have 'black box' characteristics and therefore the ability to learn from 'examples' and 'training', requiring little or no a priori information and knowledge of the system's structure [208]. Two major drawbacks of conventional neural networks are highlighted in [208]: namely, heuristic knowledge from an experienced expert cannot easily be incorporated, and the 'black box' characteristic means that its internal behaviour cannot be easily understood. Another drawback of neural networks is the lack of understanding of its internal behaviour, causing 'clearance problems'—especially for aircraft systems.

Recent research work can be found in [188, 213, 272] while application examples can be found in [5, 49, 189, 254, 287]. Examples of an intelligent approach for FDI in aircraft systems appear in papers such as [4, 238] and the references therein.

2.6 Summary

This chapter has presented a brief introduction to the fields of FTC and FDI. It includes definitions of terms regularly used in FTC and FDI such as faults and failures. This chapter also briefly discussed the possible types of faults and failures to actuators and sensors that can occur, together with a discussion on redundancy and its importance to FTC. Different methods to achieve FTC were discussed, ranging from robust control to control signal redistribution.

http://www.springer.com/978-0-85729-649-8

Fault Detection and Fault-Tolerant Control Using Sliding Modes

Alwi, H.; Edwards, C.; Pin Tan, C.

2011, XXVIII, 340 p., Hardcover
ISBN: 978-0-85729-649-8