On the Effect of Transient Data-Errors in Controller Implementations

Gäfvert, Magnus; Wittenmark, Björn; Askerdal, Örjan

Published in:
Proceedings of the 2003 American Control Conference, vols 1-6

DOI:
10.1109/ACC.2003.1244059

2003
Abstract

Computer-level faults leading to data errors in computations are predicted to occur increasingly frequent in future microprocessors. This work discusses the impact of such errors on closed-loop performance in implementations of digital control systems. A method to render a control system more robust to data errors by introducing artificial signal limits and then combine them with an anti-windup scheme is presented and exemplified.

1. Introduction

As computers, rather than electro-mechanical systems, are increasingly used for implementing control algorithms, control systems become more vulnerable to computer level failures due to faults in semiconductor devices. Such faults are classified as: (i) transient faults which are short-duration faults that are induced by neutron and alpha particles, power supply and interconnect noise, and electrostatic discharge, (ii) intermittent faults which are re-occurring short-term faults that occur due to marginal hardware or aging effects, and (iii) permanent faults which have the same causes as the intermittent faults, but reflect irreversible physical changes. The trend with decreasing transistor and interconnect dimensions, lower power supply, and higher operating frequencies contributes to increasing the occurrence rates of transient and intermittent faults, while improvements in design and manufacturing have led to decreasing permanent failure rates, less than 15 FIT (Failures In Time, or failures in 10^9 hours of operation) for microprocessors in 2001 [4]. This paper will therefore focus on transient faults.

A microprocessor is built from SRAM memory cells, latches, and combinational logic. Faults in the microprocessor may either result in control-flow errors, in which case the instruction execution order is erroneous, or in data errors, in which case the executing program delivers erroneous results. Many control-flow errors may be detected with watchdog timers. Modern microprocessors have built-in protection against transients faults in memory cells (cache) using e.g. error-correcting codes and parity checks, while the remaining parts are essentially unprotected. Methods to implement protection against transient faults in other locations than memory cells often involve redundant micro-controller subsystems (ALU etc.) as in [15], but they result in more complex and more expensive devices. Hence, remaining failures are data errors due to transient faults in the combinational logic or latches. In [14] it is predicted that the transient fault rate per chip of combinational logic circuits will increase nine orders of magnitude from 1992 (~10^{-7} FIT) to 2011 (~10^2 FIT), when the failure rate will be comparable to that of unprotected memory elements. The corresponding fault rate for the present generation of microprocessors is ~1 FIT. The trend of increasing fault rates is also reported in [8], where it is predicted that a 32 Mbit static memory implemented in a 0.1 \mu m process will fail on the average each 5.7 years at sea level. A result that is expected to hold also for other logic circuits, such as flip-flops, latches, registers, and combinational circuits.

In safety-critical applications it is imperative to have error detection and recovery functionality to meet the high demands on dependability. In some areas, such as the aircraft industry, fault-tolerance is achieved by redundant hardware and high-end devices. In more cost-sensitive areas, such as the automotive industry, expensive solutions may not be feasible. Assuming a microprocessor with a (constant) transient failure rate of 10 FIT that hosts e.g. a braking functionality in a car, the mean time to computer failure (possibly leading to catastrophic failure for the vehicle) is more than 11,415 years for a single vehicle. For a series of 100,000 cars, the mean time to failure in one of the cars is 42 days. If a microprocessor executes a control algorithm that is likely to tolerate a transient data error, this device may not necessarily have to be as fault-tolerant, and can thus be less expensive. Therefore it is of interest to investigate the impact of transient data errors on hardware hosting implementations of safety-critical control algorithms.

Understanding the effect of data errors on general computer functionality is an intensively researched area, e.g., [12]. Methods for analyzing the effects of timing errors on control systems were presented in, e.g., [9, 19]. Analysis of effects on system stability of data errors caused by EMI bursts was in-
vestigated in [10]. However, catastrophic failures in a safety-critical system may occur before the system reaches instability, e.g., if some constraint on the control error is exceeded. Recent results [5, 18] show that many data errors will have a limited effect on control performance, i.e., control systems often have an inherent resilience or inertia to data errors. This path is pursued further in this paper. It is discussed how the implementation of the control algorithms influences the tolerance to data errors resulting from computer node failures. Related work is published in [2], where classical frequency-domain methods are used to investigate the effects of computer failures on linear feedback control systems.

The paper is organized as follows: In Section 2 a general model for the effect of data errors on a control algorithm is stated. Thereafter, controller realization and scaling is discussed in Section 3. Sections 4 and 5 show how computed signal bounds together with an anti-windup scheme may improve the recovery from data errors. This is illustrated with an example in Section 6. The results are discussed in Section 7, and summarized in Section 8.

2. Modeling Data Errors Caused by Computer Node Faults

Let

\[ z(k+1) = \Phi_c z(k) + \Gamma_c^u u_c(k) + \Gamma_c^y y(k) \]
\[ u(k) = C_c z(k) + D_c^u u_c(k) + D_c^y y(k) \]  

(1)

be a state-space realization of a general linear two-degrees-of-freedom discrete-time controller with internal state \( z \), command signal \( u_c \), process output \( y \), and control signal \( u \). The controlled process is assumed to be a linear time-invariant system

\[ Y(s) = G(s)U(s) \]  

(2)

with zero-order-hold sampling. A pseudo-code implementation of (1) would be

```
repeat:
    uc := read_from_interface();
    y := read_from_sensors();
    u := compute_control_signal(z,uc,y);
    write_to_actuator(u);
    z := update_controller_state(z,uc,y);
    wait_for_next_sample;
end repeat;
```

This algorithm would be implemented and executed in a computer node like in Figure 1. Erroneous computation results due to transient data errors would eventually be stored in \( u \) or \( z \), and would propagate to the controlled process, and through the feedback loop. As illustrated in the figure, the internal components of a computer node are: communication controllers, memories, microprocessors, and internal communication buses. All these components may be affected by faults. As the communication controller handles the information exchange with other nodes, quantities measured by sensors may become erroneous due to faults in the communication controller. The data path of a microprocessor consists of caches, registers, buses, and functional units (i.e., ALU, multiplier, etc.). If a fault occurs in any of these parts, it will affect the ongoing calculation if the faulty part is activated, rendering the result of that calculation to be incorrect. The communication inside the computer node is, generally, performed using buses. If a fault occurs on such a bus, the data currently being transferred will be affected and the result may be a data error in any of the calculated data.

Data errors may be regarded as bit-flips in the digital representation of the affected variable or signal. Single-bit errors tend to be more common than multiple-bit errors [1]. In a \( N \)-bit fixed-point representation with \( M \) fractional bits the number range is \( R = [-2^{N-M-1},2^{N-M-1} - 2^{-M}] \), with a resolution of \( Q = 2^{-M} \). With this representation, the magnitudes of the errors will be in the same range as the control signals (assuming that the data have been properly scaled). In the IEEE floating-point standard with \( f \) fraction bits, and \( e \) exponent bits the range is \( R = [-2^{-2f},2^{2f-1}-1] \cdot (2-2^{-e}) \cdot 2^{2f-1} - 1 \), and bit flips in the most significant exponent bits may lead to very large errors. A reasonable model of data errors is additive impulse disturbances with magnitude of rectangular distribution within the number representation range. The data error disturbances may then be included in (1) as

\[ z(k+1) = \Phi_c z(k) + \Gamma_c^u u_c(k) + \Gamma_c^y y(k) + \eta_z(k) \]
\[ u(k) = C_c z(k) + D_c^u u_c(k) + D_c^y y(k) + \eta_u(k) \]  

(3)

where \( \eta_z(k) \) and \( \eta_u(k) \) are impulse disturbances due to data errors affecting the computation of the controller state and control signal, respectively. Since transient data errors occur sporadically, it may be assumed that an error manifests in either \( \eta_z(k) \) or \( \eta_u(k) \).

3. Controller Realizations

With a change of state variables \( \tilde{z}(k) = T z(k) \) in (3) the controller realization becomes

\[ \tilde{z}(k+1) = \Phi_{\tilde{z}} \tilde{z}(k) + \Gamma_{\tilde{z}}^u u_c(k) + \Gamma_{\tilde{z}}^y y(k) + \eta_{\tilde{z}}(k) \]
\[ u(k) = \tilde{C}_c \tilde{z}(k) + \tilde{D}_c^u u_c(k) + \tilde{D}_c^y y(k) + \eta_u(k) \]  

(4)

with \( \Phi_{\tilde{z}} = T \Phi_c T^{-1} \), \( \Gamma_{\tilde{z}}^u = T \Gamma_c^u T^{-1} \), \( \Gamma_{\tilde{z}}^y = T \Gamma_c^y T^{-1} \), and \( \tilde{C}_c = C_c T^{-1} \).

Note that \( \eta_{\tilde{z}}(k) \) is not transformed in (4), since it appears
internally in the controller implementation. Combining the controller realization (4) with a state space realization of the sampled controlled process (2)

\[ x(k+1) = \Phi x(k) + \Gamma u(k) \]
\[ y(k) = C x(k) \]  
results in the closed-loop dynamics

\[ \xi(k+1) = \Phi_c \xi(k) + \Gamma_c \eta_c(k) + \Gamma_n \eta_n(k) \]
\[ y(k) = C_c \xi(k) \]
\[ u(k) = C_c^{*r} \xi(k) + D_n^{*r} u(k) + \eta_n(k) \]  
with \( \xi = (x, z) \) and

\[ \Phi_c = \left( \begin{array}{cc} \Phi + \Gamma D_c \Gamma C & \Gamma C T \Phi_c T^{-1} \\ \Gamma T C & \Gamma T \Phi_c T^{-1} \end{array} \right) \]
\[ \Gamma_c = \left( \begin{array}{cc} \Gamma D_c \Gamma C & \Gamma C T \Phi_c T^{-1} \\ \Gamma T C & \Gamma T \Phi_c T^{-1} \end{array} \right) \]
\[ C_c = \left( \begin{array}{cc} C & \Gamma C T \Phi_c T^{-1} \end{array} \right) \]

The impulse-responses from \( \eta_c(k) \) and \( \eta_n(k) \) to \( y(k) \) are \( h_{n_c}(k) = C_c^{*r} \Phi_c^{k} \Gamma_c^{k}, k > 0, \) and \( h_{n_n}(k) = D_n^{*r} D_n^{k}, k > 0. \) Inspection of \( \Phi_c^{k} \) reveals that the structure

\[ \Phi_c^{k} = \left( \begin{array}{cc} * & * T^{-1} \\ T & T + T^{-1} \end{array} \right) \]

is preserved for all \( k > 1 \). Hence, the impulse-responses will have the structures \( h_{n_c}(k) = C \Gamma C T^{-1} \) and \( h_{n_n}(k) = C \Gamma \) for \( k > 1. \) Obviously \( h_{n_n}(k) \) is invariant under the change of state variables, while \( h_{n_c}(k) \) is not. With a large diagonal state scaling \( T = \Lambda, \) the output response to impulse errors on the states can be made arbitrary small. However, the state scaling is constrained by the available numeric precision.

The problem to find the controller realization, i.e. find the \( T, \) such that the output response \( y(k) \) to the disturbance \( \eta_c(k) \) is minimized with respect to some measure, under the constraint that the state scaling is kept appropriate for the available numeric representation is treated in e.g. [11, 13]. The results are optimal realizations in the context of roundoff errors, which have the same problem structure, but where the disturbances are close to stationary noise processes. With \( y(k) = H_n(q) \eta_n(k), z(k) = H(q) u_c(k), \) solutions to the minimization of \( ||H_n||_p \) with respect to \( T, \) subject to the state scaling constraint \( \|H_n\|_q \leq \gamma \) are presented for \( p, q \in \{2, \infty \} \) in [13]. For data errors it may be more natural to regard \( p = q = 1, \) i.e. peak-to-peak gain. Even more natural would it be to minimize \( ||H_n||_\infty, \) since the disturbances are assumed to be impulses. However, in practice it is likely more feasible to optimize the realization with respect to the all-time present round-off noise than rare and sporadic data errors. In the following it will thus be assumed that the transformation \( T \) is given, and we concentrate on the analysis of the influence of \( \eta_c \) and \( \eta_n \) on the closed-loop performance. Hence, we can consider the representations (3) and (5) of the controller and the process.

4. Signal Bounds

If a bound on the command signal \( u_c \) is known, then \( l_\infty \) bounds on controller states and the control signal may be computed using \( l_1 \) norms on impulse responses [7]. The controller state and control signal are given by

\[ z_c(k) = \sum_{j=0}^{k} h_c(k-j) u_c(j) \]  
\[ u(k) = \sum_{j=0}^{k} h_u(k-j) u_c(j) \]  
where \( \{h_c(k)\} \) and \( \{h_u(k)\} \) are the impulse-response sequences of the closed-loop system from the command signal \( u_c(k) \) to the state variable \( z_c(k) \) and control signal \( u(k). \) Bounds on \( z_c \) may be computed by applying the Hölder inequality on (8a):

\[ |z_c(k)| \leq \sum_{j=0}^{k} |h_c(k-j)| |u_c(j)| \]  
\[ \leq \left[ \sum_{j=0}^{k} |h_c(j)|^p \right]^{\frac{1}{p}} \left[ \sum_{j=0}^{k} |u_c(j)|^q \right]^{\frac{1}{q}} \]  
where \( \frac{1}{p} + \frac{1}{q} = 1. \) If the maximum absolute value of the command signal is known \( |u_c(k)| \leq M, \) then Equation (9), with \( p = 1 \) and \( q = \infty, \) gives

\[ |z_c(k)| \leq M \sum_{j=0}^{k} |h_c(j)| \]  

Correspondingly a bound on \( u(k) \) may be expressed as

\[ |u(k)| \leq M \sum_{j=0}^{k} |h_u(j)| \]

It is straightforward to include the effects of other bounded inputs to the closed-loop system, e.g. load disturbances, in the bounds of Equations (10). The bound (10a) may also be used for \( l_1 \) state scaling, as described in [7]. If the closed-loop system is well designed with proper damping, the bounds are not expected to be very conservative, which is also confirmed in [7]. If the signal bounds are exceeded it can be concluded that an error has occurred in the system. Hence, the signal bounds may be used on-line to actively detect deviations from normal operation, similar to the approach in e.g. [16]. Note, however, that in the cited work the bounds appear to be computed on the controller in open-loop, which result in over-estimation of the bounds by factors of magnitudes in comparison with the closed-loop bounds. The approach taken in the present work is to introduce explicit bounds in the controller, that correspond to those of (10), and then use well-known anti-windup methods to handle these signal limitations in a graceful manner. This will give the system an inherent robustness to data errors that exceed the signal bounds.
5. Anti-windup

In the presence of control signal limitations, the control signal actually delivered to the controlled process will be 

\[ u(k) = \text{sat}(v(k)), \]

where \( v(k) \) is the linear control signal. When the output signal is saturated the feedback path is broken and the controller states are driven in open-loop, leading to deteriorated performance or even instability. If the controller has integral action this phenomenon is denoted integrator windup. To inhibit this behavior various anti-windup schemes may be applied [6]. Anti-windup should always be implemented in a controller with actuator saturation. In this work an explicit artificial limitation according to (10b) is introduced to make the system robust to data errors. In practice an actuator limitation will also be present. If the actuator saturation limit is smaller than the artificial limit (10b), then the smallest limit should be used. The observer-based anti-windup of [3] is a general method where the control signal error is fed back to the controller. With observer-based anti-windup the controller (3) is modified as

\[
\begin{align*}
    z(k+1) &= \Phi_c z(k) + \Gamma_c^w y(k) + \Gamma_c^w u_c(k) \\
            &+ K (u(k) - v(k)) + \eta_z(k) \\
    &= \Phi_c z(k) + \Gamma_c^w y(k) + \Gamma_c^w u_c(k) \\
            &+ K u(k) + \eta_z(k) \\
    v(k) &= C_c z(k) + D_c^w y(k) + D_c^w u_c(k) + \eta_y(k) \\
    u(k) &= \text{sat}(v(k))
\end{align*}
\]

(11)

with \( \Phi_c = \Phi_c - KC_c, \Gamma_c^w = \Gamma_c^w - KD_c^w, \) and \( \Gamma_c^w = \Gamma_c^w - KD_c^w \). The gain \( K \) is chosen as to obtain the desired observer dynamics given by \( \Phi_c \). The anti-windup scheme will now reduce the effect of data errors that are causing the controller output to exceed the estimated limits. Note that the observer based anti-windup operates on all controller states, while certain other schemes only operate on the integrator state. Hence, errors in any of the controller states are eliminated due to the anti-windup. Also note that the observer-based method does not require any additional states to be introduced. In the context of data errors this is important, since data errors affecting explicit anti-windup states would not be handled gracefully by the system. The closed-loop system resulting from combining (11) with (2) is shown in Figure 2.

6. Example

As an illustration of the inherent tolerance to transient faults that may be achieved with the combination of signal bounds and anti-windup, we study the control of the simple servo process

\[ G(s) = \frac{100}{s(s+10)} \]

(12)

A discrete-time two-degrees-of-freedom tracking controller is synthesized using the polynomial pole-placement design method of [3]. The controller is designed for a closed-loop realization \( z_1 \) is an integrator state, and \( z_2 \) may be interpreted as an observer state. The command signal \( u_c \) is assumed to be bounded by \( |u_c(k)| \leq 1 \). The \( l_1 \) state-scaling of (10a) is applied such that \( |x_i(k)| \leq 1 \). The control signal bound according to (10b) is \( |u(k)| \leq 1.88 \). The corresponding open-loop bound according to [16] is \( |u(k)| \leq 341 \). The anti-windup scheme of (11) is implemented, with \( K \) chosen as to obtain dead-beat dynamics, to obtain the closed-loop system of Figure 2. In Figure 3 the closed-loop step-response to a command-signal unit step is shown. Note that the computed state-bounds are very tight, while the control signal bound seems to be a little more conservative. In Figure 4 the closed-loop response to a data error \( \eta_z(k) = 5 \delta(k) \), affecting the computation of the integrator state \( z_1 \), is shown in the case when the artificial signal-limit and anti-windup are not applied. It can be seen how the integrator state slowly recovers, while the control error grows large. In presence of actuator limitation large data-error amplitudes even result in instability. Figure 5 shows the corresponding response with applica-
tion of the artificial signal limit $|u(k)| \leq 1.88$, in combination with dead-beat anti-windup. In this case the controller state recovers within a few samples, and the control error is moderate. Figure 6 shows the largest absolute control-error that results from various magnitudes on the data errors $n_{q_1}$, $n_{q_2}$, and $n_{q_3}$. Four different system settings are shown for comparison: (i) No control-signal limits or anti-windup. Here the graph is a straight line with a slope equal to the $l_1$-norm of the impulse response. (ii) Dead-beat anti-windup. Note that the response to data errors affecting the control signal ($n_{u}$) leads to larger control errors in this case. This is because the data error is interpreted as a saturation by the anti-windup, which is fast enough to react immediately. Important to note is also that the control error magnitude does not increase with the data error magnitude for large data errors. (iii) Dead-beat anti-windup only on the integrator state (this is common in e.g. PI-controllers). This is normally sufficient for handling actuator limitations, but in the case of data-errors the performance is inferior compared with full-state anti-windup. (iv) Dynamic anti-windup with a bandwidth of $2\omega_{n_d}$. Here the response to data errors entering the control signal is better, since the anti-windup is too slow to react immediately on the error impulse. The response to data errors on the state is, however, worse than for the dead-beat design, as is clear from the figure. The recovery time is also significantly slower compared to the dead-beat design, which may be seen from time-domain response plots.

7. Discussion

In general it can be noted that the integrator state $z_1$ is most sensitive to data errors, which seems very intuitive since it depends on feedback of the control error to decay, while the controller state $z_2$ is stable with a short time constant, and decays by itself when the loop is broken by saturation. Another general observation is that the control-signal bound is computed from bounds on the command signal, and in consequence is related to the expected magnitude of control errors during normal operation. Hence, the artificial limits and the anti-windup scheme will capture data errors resulting in control errors larger than those expected during normal operation. The proposed method may also be combined with the dynamic bounds of [17], to increase the coverage for data errors.

Since the controller state $z$ is bounded by (10a), it may seem natural to introduce explicit limits also for this variable.
However, simulations with saturations on the controller state indicate that the performance improvement is minor. The increased complexity resulting from additional saturations in the loop also makes the system difficult to analyze, even if it seems to perform well in presence of data errors.

In presence of stochastic signals such as measurement noise there will be a probability of false error detections. This may be handled by computing the resulting variance of the control signal. The deterministic control-signal bound (10b) is then adjusted with some measure depending on the variance. The size of the adjustment will determine the probability of false detections. Sporadic false detections will affect control performance as the anti-windup intervenes. By using slower anti-windup dynamics the noise sensitivity is decreased.

If a rate bound on the command signal is known $|\Delta u_c(k)| = |u_c(k) - u_c(k-1)| \leq M_\Delta$, then state and control-signal rate bounds may be computed in analogy with (10). An artificial rate bound on the control signal may then also be introduced in the controller, and used together with the anti-windup scheme. Note that the noise sensitivity will be worse than for the case with magnitude bounds, since the rate of the control signal will depend on the noise variance.

8. Summary

Data errors resulting from transient faults in the computer hardware may be modeled as impulse errors entering the control algorithm internally. The effect on closed-loop performance therefore depends on the controller realization. Previous methods to optimize controller realizations with respect to round-off errors in finite-precision numerics are applicable also in the context of transient data errors. Given a controller realization, robustness to data errors may be achieved by introducing artificial signal limitations based on $l_\infty$-bounds, in combination with an anti-windup scheme. As one may expect, the integrator state of the controller is most sensitive to data errors.

9. References


