Respect the Unstable

The practical, physical (and sometimes dangerous) consequences of control must be respected, and the underlying principles must be clearly and well taught.

By Gunter Stein

Feedback control systems are all around us in modern technological life. They are at work in our homes, our cars, our factories, our transportation systems, our defense systems—everywhere we look. Certainly, one of the great achievements of the international controls research community is that the design principles for these systems are well developed and broadly understood by control engineers, so that the systems operate productively and safely in so many applications.

In this article, I want to talk about two trends that threaten to undermine this achievement. My objective is to heighten our awareness of these trends and hopefully bring about an appropriate response to them.

The first trend has to do with the applications themselves. Among the abundance of control systems operating today are increasing numbers of dangerous ones. Society trusts our technology. We are permitted to do things with automatic controls that cannot be done manually and that, if done improperly, can have dire consequences for property, the environment, and human life. Most, but not all, of these dangerous applications involve open-loop unstable plants with divergence rates violent enough to elude manual control. This characterization motivates the title of the article, and I will describe specific examples of such applications.

The second trend has been evident at our conferences, and certainly in our journals, over the years. This trend is the increasing worship of abstract mathematical results in control at the expense of more specific examinations of their practical, physical consequences. I will provide examples of this trend as well.

Gunter Stein's Bode Lecture

An understanding of fundamental limitations is an essential element in all engineering. Shannon's early results on channel capacity have always had center court in signal processing. Strangely, the early results of Bode were not accorded the same attention in control. It was therefore highly appropriate that the IEEE Control Systems Society created the Bode Lecture Award, an honor which also came with the duty of delivering a lecture. Gunter Stein gave the first Hendrik W. Bode Lecture at the IEEE Conference on Decision and Control in Tampa, Florida, in December 1989. In his lecture he focused on Bode's important observation that there are fundamental limitations on the achievable sensitivity function expressed by Bode's integral. Gunter has a unique position in the controls community because he combines the insight derived from a large number of industrial applications at Honeywell with long experience as an influential adjunct professor at the Massachusetts Institute of Technology from 1977 to 1996. In his lecture, Gunter also emphasized the importance of the interaction between instability and saturating actuators and the consequences of the fact that control is becoming increasingly mission critical.

After more than 13 years I still remember Gunter's superb lecture. I also remember comments from young control scientists who had been brought up on state-space theory who said: "I believed that controllability and observability were the only things that mattered." At Lund University we made Gunter's lecture a key part of all courses in control system design. Gunter was brought into the classroom via videotapes published by the IEEE Control Systems Society and the written lecture notes. It was a real drawback that the lecture was not available in more archival form. I am therefore delighted that *IEEE Control Systems Magazine* is publishing this article. I sincerely hope that this will be followed by a DVD version of the videotape. The lecture is like really good wine; it ages superbly.

—Karl J Åström, Professor Emeritus
Lund University, Lund, Sweden

0272-1708/03/$17.00©2003IEEE

Together, these trends threaten to undermine our good standing in society as masters of a technology that can be trusted.

Basic Facts About Unstable Plants

- Unstable systems are fundamentally, and quantifiably, more difficult to control than stable ones.
- Controllers for unstable systems are operationally critical.
- Closed-loop systems with unstable components are only locally stable.

The Punch Line

In slightly dramatized form, the message of this article can be summarized by drawing some contrasts. Consider the following mathematical statement:

Theorem: Given plant $g(s)$, compensator $k(s)$, and

$$\begin{aligned} gk(s) &= n(s)/d(s); \quad s(s) = 1/[1 + gk(s)] \\ Z &= \{z \mid n(z) = 0, \operatorname{Re}(z) \geq 0\} \\ P &= \{p \mid d(p) = 0, \operatorname{Re}(p) \geq 0\}. \end{aligned}$$

Then $s(s)$ is stable if and only if

$$\begin{aligned} s(s) &< \infty \quad \forall \operatorname{Re}(s) \geq 0 \\ s(z) &= 1 \quad \forall z \in Z \\ s(p) &= 0 \quad \forall p \in P. \end{aligned}$$

In words, this theorem states that the sensitivity function of a feedback system must not only be finite in the right-half plane, but it must pass through certain interpolation points corresponding to right-half-plane singularities of the loop. Most of us recognize this immediately as an elegant and compact description of control system constraints imposed by unstable, nonminimum-phase systems. It was formally developed in the context of parametrizing all stabilizing controllers, and it was popularized in the 1980s as part of the interpolation-theoretic approach to H_∞ optimization. (Of course, it was understood as stated above for single-input, single-output (SISO) systems as far back as the 1950s. Some historical notes on this theorem can be found in [1].)

Unfortunately, we are not as quick to recognize that this mathematical description includes some very dangerous systems. For example, the theorem applies to the JAS-39 airplane (the SAAB Gripen), which crashed on landing in 1989 in one of its first test flights. Figure 1 shows a video frame from the crash. Fortunately, the pilot survived, but the airplane was lost and its development program substantially delayed.

The theorem also applies to the Chernobyl nuclear plant, shown in Figure 2 as it appeared shortly after its accident in 1986. We are all familiar with the consequences of that accident—hundreds of people dead, hundreds of thousands evacuated, and hundreds of millions of dollars in cleanup costs.

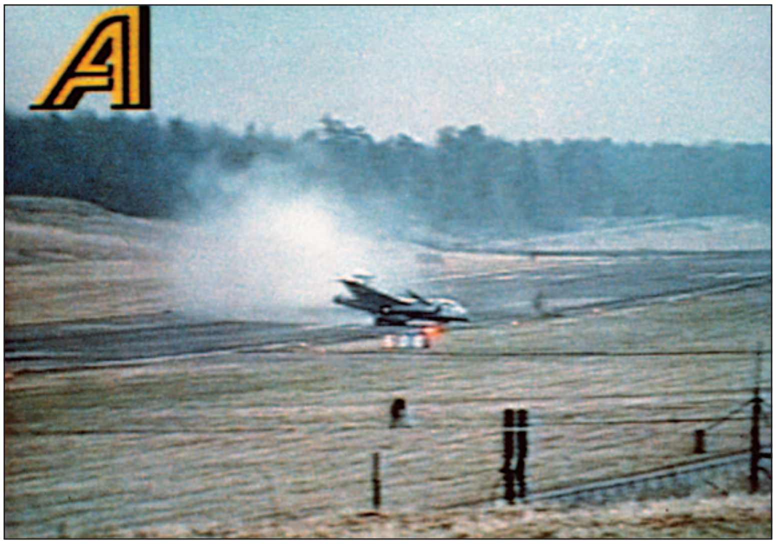
These and other examples dramatize the contrast between elegant mathematical statements and the real physical systems that they purport to describe. I have selected these two examples because both catastrophes involve explicit, traceable responsibilities of control systems. We will take a closer look at those responsibilities later.

My point is this: As society permits control engineers to operate more such dangerous systems, we who teach those engineers and fashion their tools cannot hide from responsibility under a cloak of mathematics. We dare not instill the notion that mathematical rigor is the only goal to strive for in control. We must also instill respect for the practical, physical consequences of control, and we must make certain that its underlying principles are taught clearly and well.

I want to explore this point by reviewing some basic facts about unstable plants. Again, we will consider *unstable* to be synonymous with *dangerous,* even though this is not all inclusive. I want to review these facts:

- unstable systems are fundamentally, and quantifiably, more difficult to control than stable ones
- controllers for unstable systems are operationally critical
- closed-loop systems with unstable components are only locally stable.

These facts should be well known to all of us, but as we will see, they are not always taken to heart.



©FAME PICTURES

Figure 1. *Gripen JAS39 prototype accident on 2 February 1989. The pilot received only minor injuries.*

Bode Integrals

The first integral applies to stable plants and the second to unstable plants. They are valid for every stabilizing controller, assuming only that both plant and controller have finite bandwidths. In words, the integrals state that the log of the magnitude of sensitivity function of a SISO feedback system, integrated over frequency, is constant. The constant is zero for stable plants, and it is positive for unstable ones. It becomes larger as the number of unstable poles increases and/or as the poles move farther into the right-half plane. (Technically, we must count all unstable poles here, including those in the compensator, if any.)

$$\int_0^\infty \ln|\,s(j\omega)|\,d\omega = 0$$

$$\int_0^\infty \ln|\,s(j\omega)|\,d\omega = \pi \sum_{p\in P} \mathrm{Re}(p)$$

Fact 1: Fundamental, Quantifiable Control Difficulty

One of the best-known examples of instability is the inverted pendulum, or the “broomstick balancing problem.” Those of us who learned our craft in the 1950s and 1960s know this problem well, either through the little cart experiments that appeared during those years in various university labs around the world or through textbook examples and simulations. Recall that this problem was motivated by the space race. Control engineers had to learn how to balance rockets on top of their plumes on their way to earth orbit. It is still an important problem today, in the post-Challenger era, as modern boosters, now built in several countries around the world, become more difficult to balance.

I bring up the inverted pendulum because it nicely illustrates our first fact. The illustration is this: I can obviously balance an ordinary stable pendulum without difficulty. I can also easily balance a long inverted pendulum, even in front of a room full of people. However, I find it more difficult to balance a shorter inverted pendulum, and I find it impossible to balance a very short one. You have probably tried this yourself. The exact lengths you can balance might be different, but the trend will be the same.

Certainly the theorem just cited has an explanation for this illustration hidden in it somewhere, and we could try to extract that explanation using all the modern machinery at our disposal. Perhaps we could compute some minimum H_∞ norms or even some minimum structured singular values achievable by human controllers. Of course, in our calculations, we would need to pay due attention to the inherent handicaps of such controllers, such as reaction time, neuromuscular lags, limb inertias, and many other uncertainties covering the fact that humans, and most everything else in the physical world, are not finite dimensional, linear, and time invariant.

The Bode Integrals

I will try to explain these observations in the frequency domain the way Hendrik Bode might have done it. It turns out that such an explanation is insightful, plain, and clear and is therefore preferable to many modern ones.

First, note that the difference between balancing long sticks and short sticks has to do with the location of the unstable mode. This is obvious from the linearized equations of motion of the stick. Under the simplifying assumption that all mass is concentrated at the end of the stick, these equations show an unstable pole at $\sqrt{g/L}$, where g denotes the acceleration of gravity and L denotes length. Divergence becomes more rapid as L decreases.

A frequency domain quantification of control difficulty in the face of such changing instabilities is captured by the Bode integrals (see sidebar). At the risk of sounding dogmatic, I believe that every control theoretician and every control engineer should know these integrals and understand their meaning. Unfortunately, we have not always taught them well.

My own background illustrates this last observation. You can judge for yourself, using your own background. During my entire control education as an undergraduate and graduate student in the 1960s, I ran across only the first integral, only

©AP/WIDE WORLD PHOTOS

Figure 2. *Chernobyl nuclear power plant shortly after the accident on 26 April 1986.*

once, and only in an optional reading of an unassigned chapter in one of the classical textbooks. This integral surfaced for me for the second time in the mid 1970s, referenced in a paper by Isaac Horowitz titled "On the Superiority of Transfer Functions over State-Variable Methods. . . ." It appeared as a perspectives paper in *IEEE Transactions on Automatic Control* amid a certain amount of controversy [2].

The second integral did not surface for me until 1983, in a talk by Jim Freudenberg at an IEEE Conference on Decision and Control in San Antonio [3]. If memory serves, someone pointed out at the time that this result was "just a version of Jensen's theorem," well known in mathematics for a long time. Perhaps this historical reference reduced the value of the result in the minds of some listeners, but it should not have, because the integral explains so much about the difficulties of controlling unstable systems.

A Bode Integral Interpretation

I like to think of Bode's integrals as conservation laws. They state precisely that a certain quantity—the integrated value of the log of the magnitude of the sensitivity function—is conserved under the action of feedback. The total amount of this quantity is always the same. It is equal to zero for stable plant/compensator pairs, and it is equal to some fixed positive amount for unstable ones.

Since we are talking about the log of sensitivity magnitude, it follows that negative values are good (i.e., sensitivities less than unity, better than open loop) and positive values are bad (i.e., sensitivities greater than unity, worse than open loop). So for open-loop stable systems, the average sensitivity improvement a feedback loop achieves over frequency is exactly offset by its average sensitivity deterioration. For open-loop unstable systems, things are worse because the average deterioration is always larger than the improvement. This applies to every controller, no matter how it was designed. Sensitivity improvements in one frequency range must be paid for with sensitivity deteriorations in another frequency range, and the price is higher if the plant is open-loop unstable.

It is curious, somehow, that our field has not adopted a name for this quantity being conserved (i.e., the integrated log of sensitivity magnitude), to put it on a par with some of the great quantities of physics such as mass, momentum, or energy. But since it has not, we are free to choose a name right now. Let me propose that we simply call it *dirt*. It is stuff we would rather not have around; the less we have, the better. I want to choose this name because it lets me liken the job of a serious control designer to that of a ditch digger, as illustrated in Figure 3. He moves dirt from one place to another, using appropriate tools, but he never gets rid of any of it. For every ditch dug somewhere, a mound is deposited somewhere else. This fact is most evident to the ditch digger, because he is right there to see it happen.

In the same spirit, I can also illustrate the job of a more academic control designer with more abstract tools such as linear quadratic Gaussian (LQG), H_∞, convex optimization, and the like, at his disposal. This designer guides a powerful ditch-digging machine by remote control from the safety of his workstation (Figure 4). He sets parameters (weights) at his station to adjust the contours of the machine's digging blades to get just the right shape for the sensitivity function. He then lets the machine dig down as far as it can, and he saves the resulting compensator. Next, he fires up his automatic code generator to write the implementation code for the compensator, ready to run on his target microprocessor.

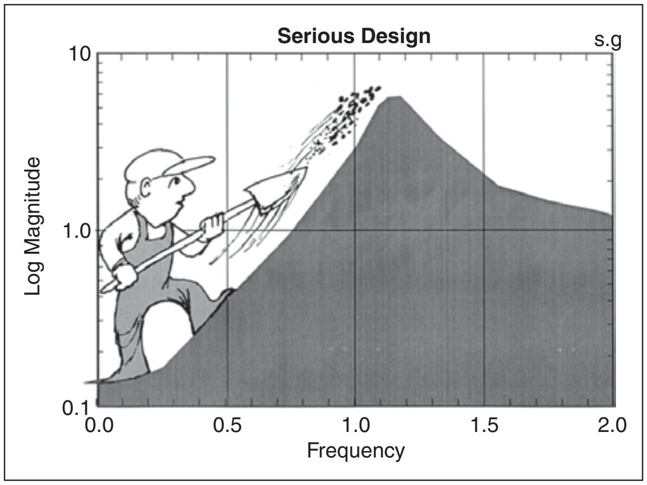


Figure 3. *Sensitivity reduction at low frequency unavoidably leads to sensitivity increase at higher frequencies.*

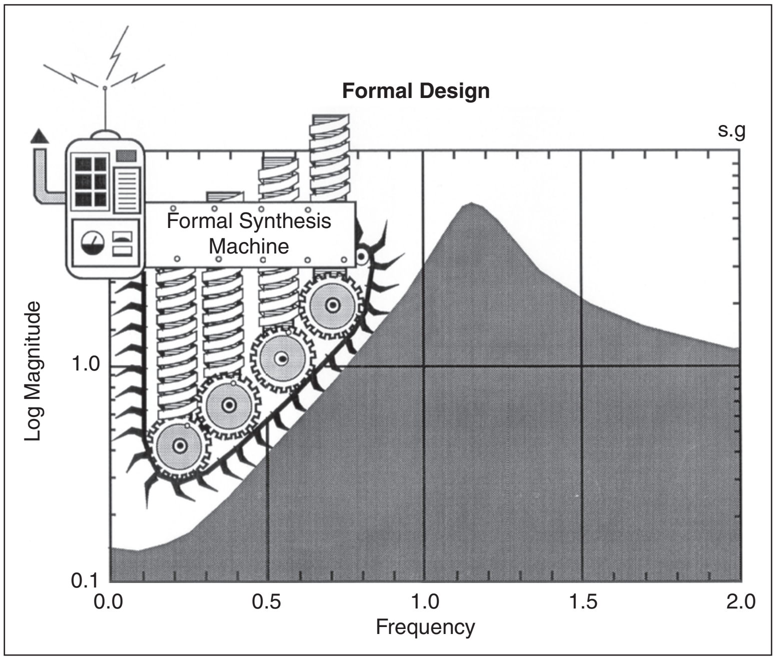


Figure 4. *Sensitivity shaping automated by modern control tools.*

He downloads that code automatically to the microprocessor and hits the power-on button on the control system. Indeed, this entire process can become so automatic and insulated that the designer may never look at what has actually happened to his control loop, and all too often the power-on creates a rude surprise.

Available Bandwidth

Many rude surprises that occur in automated design scenarios have to do with excessive bandwidth. The designer unwittingly allows the machine to dig too deeply, piling up dirt at high frequencies where it cannot be supported.

The notion that dirt piled at high frequencies needs support is not taken seriously enough in the theoretical community, even today. For example, an argument is sometimes made that the Bode integrals are not really restrictive because we only seek to dig holes over finite frequency bands. We then have an infinite frequency range left over into which to dump the dirt, so we can make the layer arbitrarily thin. The weakness of this argument is evident from standard classical theory. A thin layer, say with $\ln|s| = \varepsilon$, requires a loop transfer function whose Nyquist diagram falls on a near-unit circle, centered at $(-1 + j0)$ with radius $\approx (1 - \varepsilon)$, over a wide frequency range. This means that the loop cannot simply attenuate at high frequencies but must attenuate in a very precise way. The loop must maintain very good frequency response fidelity over wide frequency ranges.

But a key fact about physical systems is that they do not exhibit good frequency response fidelity beyond a certain bandwidth. This is due to uncertain or unmodeled dynamics in the plant, to digital control implementations, to power limits, to nonlinearities, and to many other factors. Let us call that bandwidth the "available bandwidth," Ω_a, to distinguish it from other bandwidths such as crossover or 3-dB magnitude loss. The available bandwidth is the frequency up to which we can keep $gk(j\omega)$ close to a nominal design and beyond which we can only guarantee that the actual loop magnitude will attenuate rapidly enough (e.g., $|gk| < \delta/\omega^2$). In today's popular robust control jargon, the available bandwidth is the frequency range over which the unstructured multiplicative perturbations are substantially less than unity.

Note that the available bandwidth is not a function of the compensator or of the control design process. Rather, it is an a priori constraint imposed by the physical hardware we use in the control loop. Most importantly, the available bandwidth is always finite.

Given all this, Bode's integrals really reduce to finite integrals over the range $0 \le \omega \le \Omega_a$, i.e,

$$\int_0^{\Omega_a} \ln|s(j\omega)|\,d\omega = \delta \quad \text{stable loops}$$

$$\int_0^{\Omega_a} \ln|s(j\omega)|\,d\omega = \pi \sum_{p \in P} \mathrm{Re}(p) + \delta \quad \text{unstable loops.}$$

All the action of the feedback design, the sensitivity improvements as well as the sensitivity deteriorations, must occur within $0 \le \omega \le \Omega_a$. Only a small error (δ) occurs outside that range, associated with the tail of the complete integrals. Since the details of this tail are not guaranteed by the design, the error can be either positive or negative. The design only guarantees that it will be small.

As an aside, you probably know that other constraints imposed by right-half-plane zeros also give rise to finite integrals, although in different variables [4]. I did not choose to use these additional integrals here because they still require infinite available bandwidth, and I believe that finiteness of available bandwidth is a more significant concept in control than nonminimum phaseness, even though it is not nearly so elegant.

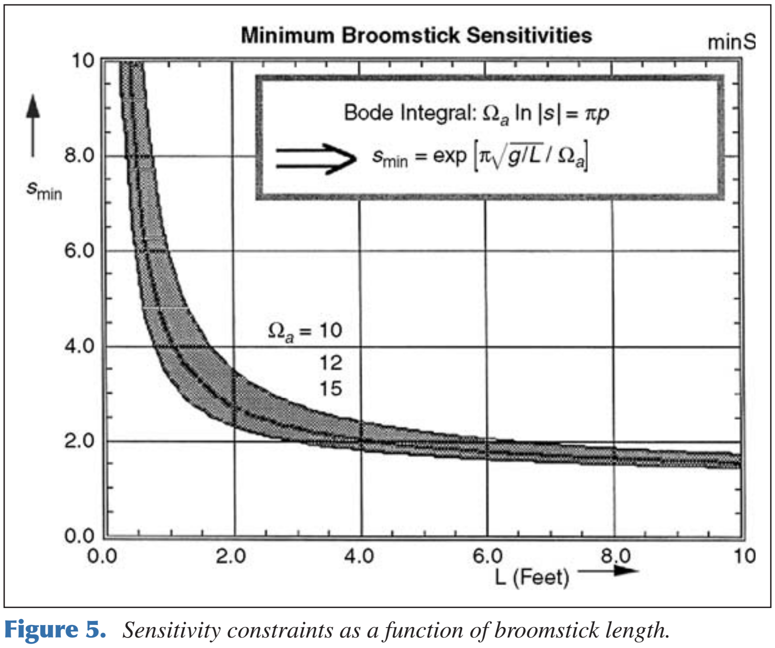


Figure 5. *Sensitivity constraints as a function of broomstick length.*

An Explanation of the Broomstick

These last integrals bring us to the point where we can explain the broomstick balancing illustration in a Bode-like way. First, what is the available bandwidth of the feedback loop in that experiment? Looking at the plant alone, it is fairly high. The stick is stiff, air drag is negligible, and little else prevents the stick from moving as required. The compensator, however, is another matter. Its physical implementation by a human operator has many complex limitations associated with perception, computation, and actuation of limbs. Many years of study and experimentation have gone into the characterization of these limitations, especially for piloting tasks in military airplanes. Since we obviously cannot go through all that here, let us simply agree that the compensator is good for a frequency range up to ≈2 Hz (say, 10-15 rad/s) and that its control strategy is to keep sensitivity as small as possible over that range (i.e., keep the loop's Nyquist diagram as far away from the critical point as possible everywhere).

The curves in Figure 5 show how well the closed-loop system will perform. These curves are simply restatements of the finite Bode integral, with sensitivity assumed equal to a constant minimum achievable value over $0 \leq \omega \leq \Omega_a$ and equal to unity elsewhere. The minimum achievable value is an exponential function of the ratio of unstable pole to available bandwidth (p/Ω_a) and with p given by $\sqrt{g/L}$, it is an explicit function of the broomstick length. Notice that a dramatic increase in sensitivity occurs below a foot and a half. These large sensitivities put the loop close to the critical point, and even minor imperfections in the implementation will cause instability. That, quite simply, is the reason we humans have trouble balancing short sticks, and other controllers with similar available bandwidth limits have trouble as well. This reason follows directly from the most fundamental conservation laws of feedback, Bode's integrals, and needs none of our modern mathematical machinery. More importantly, none of the modern results overcome these limitations. They can (and do!) hide them from us, but they do not remove them.

The X-29 Airplane Story

Although the broomstick-balancing illustration itself is only a toy, it represents some very serious physical systems with similar dynamics. In addition to the launch booster analogy already mentioned, there is also a direct analogy to unstable airplanes. As an example of the latter, I would like to tell a story about the X-29, the forward-swept-wing research airplane shown in Figure 6.

This airplane was built to demonstrate basic aerodynamic performance improvements that might be gained from new composite materials tailored specifically for aerodynamic efficiency. The wings are swept forward instead of aft, so that bending enhances lift instead of decreasing it at high angles of attack; a large canard surface is placed close in to take advantage of favorable interactions with the wing; and there are several other features as well. Once demonstrated, these technologies will make their way into next-generation airplanes.

The airplane was built by Grumman Aircraft Company, under U.S. Air Force, DARPA, and NASA sponsorship, and underwent successful flight tests for several years. Honeywell supplied the control hardware and software. Control laws were designed at Grumman, with supporting design activity at Honeywell, as well as several other places.

Although the airplane's controls were not a major focus of the flight demonstration, they are of interest here because they illustrate how extreme our worship of formal methods has become.You see, all of the various control design teams used modern digging machines early in the design process. As a result, we were well insulated from the fundamental difficulties imposed by the airplane's violent open-loop instability. We discovered only at the last moment that the vehicle was almost too unstable to control with the given hardware.

We will consider unstable to be synonymous with dangerous.

I want to relate this story by talking first about what makes the airplane unstable, then about the hardware features that restrict available bandwidth, and finally, I will put these together with Bode's integral to show the airplane's fundamental control limitations.

Static Instability

An airplane is open-loop unstable when its center of pressure (cp, the effective point of action of lift forces) is located ahead of its center of gravity (cg). Since lift forces grow in direct proportion to pitch (nose up or down) attitude, any initial pitching motion changes lift, which acts through the cp – cg offset to produce moments in the same direction, and the attitude diverges. In aeronautical circles, this condition is called static instability. The associated linearized dynamics look very much like broomstick-balancing equations. There are two roots—one stable, one unstable—approximately equal in magnitude.

Static instability does not happen arbitrarily. It is deliberately designed into an airplane by locating lifting surfaces

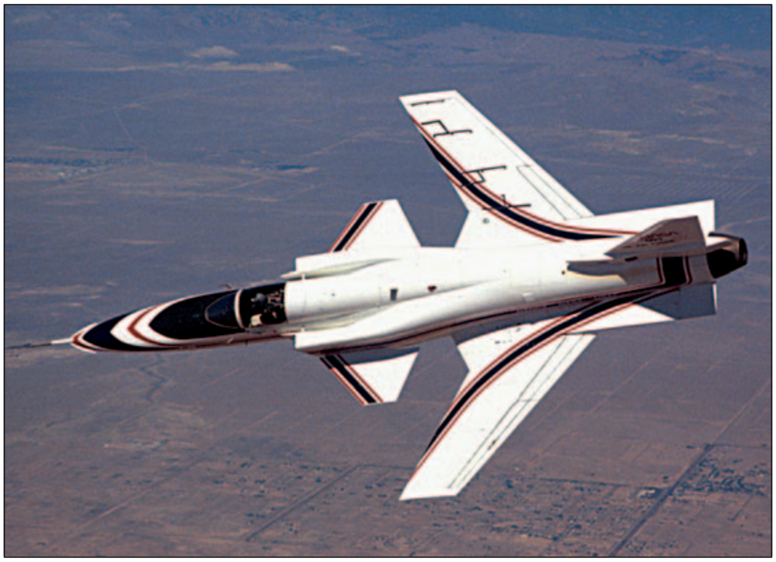
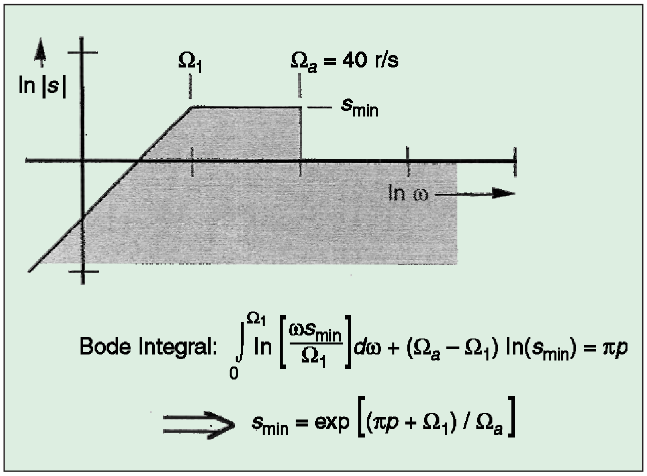


Figure 6. *NASA X-29 forward-swept-wing aircraft (photo courtesy of NASA).*

and distributing mass appropriately. In the years before full-authority automatic flight controls, most airplane designs constrained the placement of these elements to ensure stability over all flight regimes and all loading conditions. (The Wright brothers' airplane is a notable exception.) Today, however, when automatic controls are accepted and trusted in so many applications, there are important benefits to be gained by making the basic design unstable. For instance, note that a stable airplane requires a tail to balance the moment produced by lift acting through the cp – cg offset. The force produced by the tail is down, opposing the lift and making the overall configuration aerodynamically less efficient. The airplane also needs to carry the weight of the tail, which reduces payload. The less stable an airplane is, the smaller these performance and weight penalties. There are also other reasons for wanting instability, having to do with maneuverability and speeds of command response.

In the case of the X-29, the benefits of instability were desired in the transonic and supersonic flight regimes, so the airplane was designed to be modestly unstable in those regimes. Unfortunately, there is a basic aerodynamic phenomenon that moves the center of pressure of a lifting surface dramatically aft as speeds go from sub- to supersonic. As a result, the X-29's slight instability at supersonic speeds turns into a much more dramatic instability at subsonic speeds. Indeed, the airplane's real pole is as large as +6 rad/s in some flight regimes.



$$\text{Bode Integral: } \int_0^{\Omega_1} \ln\left[\frac{\omega s_{min}}{\Omega_1}\right] d\omega + (\Omega_a - \Omega_1)\ln(s_{min}) = \pi p$$

$$\Longrightarrow s_{min} = \exp\left[(\pi p + \Omega_1)/\Omega_a\right]$$

Figure 7. *Prototype X-29 sensitivity function.*

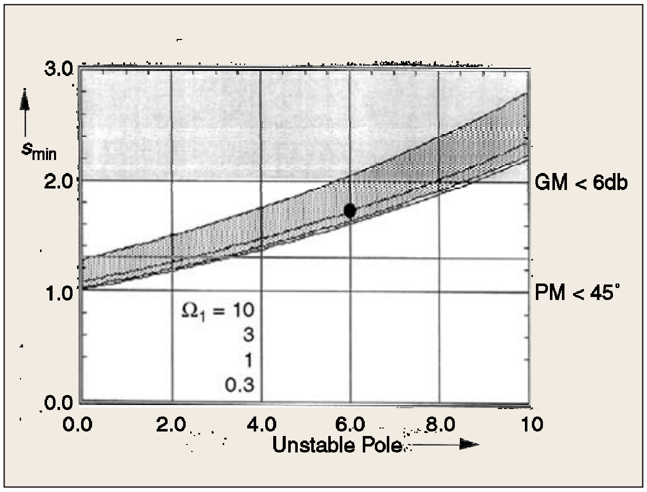


Figure 8. *Minimum X-29 sensitivities.*

Recalling the broomstick's formula ($p = \sqrt{g/L}$), flying this airplane manually corresponds to balancing a 1-ft-long stick. It is not something we can expect a pilot to do without automatic assistance, at least for very long.

X-29 Available Bandwidth

As with the broomstick, control difficulties associated with the X-29's unstable pole should not arise merely because the pole is large. Rather, they should arise if the pole is large compared with available bandwidth. Some of the major hardware elements in the control loop that limit this bandwidth include the following:

- *Sensors*: rate gyros and accelerometers, used for inner-loop stabilization. Their bandwidths, measured in the usual 3-dB-gain sense, are typically 120 rad/s or more.
- *Control processors*: digital computer systems sampling sensor data and computing surface commands at 80 Hz. This update rate can pass signals with good fidelity up to 30-40 rad/s (two to three samples per radian).
- *Actuators*: high-pressure hydraulic systems with servos to position each aerodynamic control surface. For the pitch axis, control surfaces include the canard, inboard and outboard flaperons along the wing, and strakes at the tail. The servo bandwidths, again measured by 3-dB gain, are approximately 20 rad/s. Basically, each of the servos is a position feedback loop around a hydraulic ram (an integrator). This gives a dominant first-order-lag response that retains its characteristics up to perhaps 70-80 rad/s, where valve dynamics, fluid compressibility, and local structural dynamics begin to take effect. The bottom line is that the available bandwidth of actuators is approximately 70 rad/s, at least for small signals.
- *Aerodynamics*: flow conditions around the airplane that map its geometric configuration (e.g., net orientation and control surface positions) into forces and moments. At low frequencies, this map is algebraic and well known, but because air itself has mass and momentum, the map becomes dynamic and poorly known on time scales comparable to characteristic length divided by velocity. For a 2-ft surface traveling at 200 ft/s, that time scale is 0.01 s, so the available bandwidth for aerodynamics is approximately 100 rad/s.
- *Airframe:* the mechanical structure linking the attachment points of actuators to the attachment points of sensors. (These are the important structural points affecting control. Other disciplines are, of course, interested in additional attachment points, such as where the wings are attached, where the pilot sits, etc.) The structure is rigid at low frequencies, but it begins to bend and flex dramatically at higher fre-

quencies. Typical fighter-sized airplanes have a first fuselage bending mode at ≈ 7.0 Hz (40 rad/s). The frequency of this mode, as well as its mode shape and damping characteristics, move around substantially with mass distribution (fuel, payloads) and with dynamic loading (maneuvers). This makes it very difficult to maintain good frequency response fidelity beyond the first mode frequency.

This list shows that the most severe limitation on available bandwidth for the X-29 is approximately 40 rad/s and comes from the airplane's mechanical structure and from the sampling rate of its computers. I have elected to describe the various other items, at least briefly, to emphasize the point that real physical systems have a multitude of limitations on available bandwidth, not just one or two that can be easily pushed out or removed. We cannot simply ignore them, as we often do in formal theories. Instead, we should fashion theories to clearly expose the consequences of these limitations, and we should honestly tell ourselves and our employers what the consequences are.

As mentioned earlier, in the initial X-29 designs, the various design teams used some of the formal digging machines, which hide these consequences. Not until late in the design process, and not without heated arguments among designers, did we all agree that they are indeed real and unavoidable.

X-29 Limitations

Using Bode's integral, the limitations are easy to demonstrate. We have a plant with an unstable pole, $p = 6$ rad/s, and with an available bandwidth, $\Omega_a = 40$ rad/s. We want to achieve a sensitivity function shaped approximately like the prototype shown in Figure 7. This prototype has small sensitivity at low frequencies, rising with $a+1$ slope up to Ω_1. It then stays flat up to Ω_a, so as to pay as small a penalty as possible, consistent with Bode's integral. Finally, it returns to unity (0 dB) beyond Ω_a. A formula for the smallest sensitivity penalty, $s_{\min}$, is easy to derive directly from the integral and is also shown in the figure.

Some $s_{\min}$ curves based on this formula are shown in Figure 8. We see that the penalty level rises as the airplane becomes more unstable and also as the frequency Ω_1, up to which we want good performance, increases. The design point for the X-29 is shown at $p = 6$ rad/s and $\Omega_1 = 3$ rad/s. The resulting smallest sensitivity penalty is ≈ 1.75. Classical designers will immediately recognize this to be marginal because traditional phase margins will not be satisfied. These margins are given in terms of $s_{\min}$ by well-known formulas [e.g., $PM = 2\sin^{-1}(1/2s_{\min})$]. Boundaries for standard military flight control specifications, a 45°-phase margin and a 6-dB gain margin, are shown in the figure.

Notice that we have not actually done any design work. We have made no reference to design methods or to big ditch-digging machines. Instead, we have just used a basic Bode integral calculation to determine that the situation is marginal.

To confirm that the situation is marginal, Figure 9 shows a Bode diagram of a loop transfer function corresponding to a realizable version of the prototype shape in Figure 7. This realizable version was found by taking a fourth-order approximation of Figure 7 (basically rounding the corners), selecting one free parameter (the new $s_{\min}$) to satisfy Bode's integral, and then solving explicitly for the loop transfer function (i.e., $gk(s) = s^{-1}(s) - 1$). Note that the resulting loop is well rolled off at the 40-rad/s available bandwidth and that 35° is the largest phase margin achieved by the prototype.

Based on what we have just shown, no controller can make $s_{\min}$ smaller under the given constraints. Thus, we should not be surprised that no acceptable controller was ever found, even with several design teams working on the problem. Indeed, the airplane flies today only because special specification relief was granted. As described in [5], the airplane's stability margins were actually measured explicitly in flight. Two curves from these flight tests are shown as Bode diagrams in Figure 10, one with slightly different control gains than the other. Note the remarkably close correspondence between the Bode-integral-derived prototype and the actual final flight control loop.

Based on this experience with the X-29, it is unlikely that airplanes now being developed, such as the Saab JAS-39, the Advanced Technology Fighter (ATF), and the National Aerospace Plane (NASP X-30), will be so violently unstable. A good rule of thumb, seen from Figure 8, is that available bandwidth should exceed the airplane's unstable pole by at least a factor of ten.

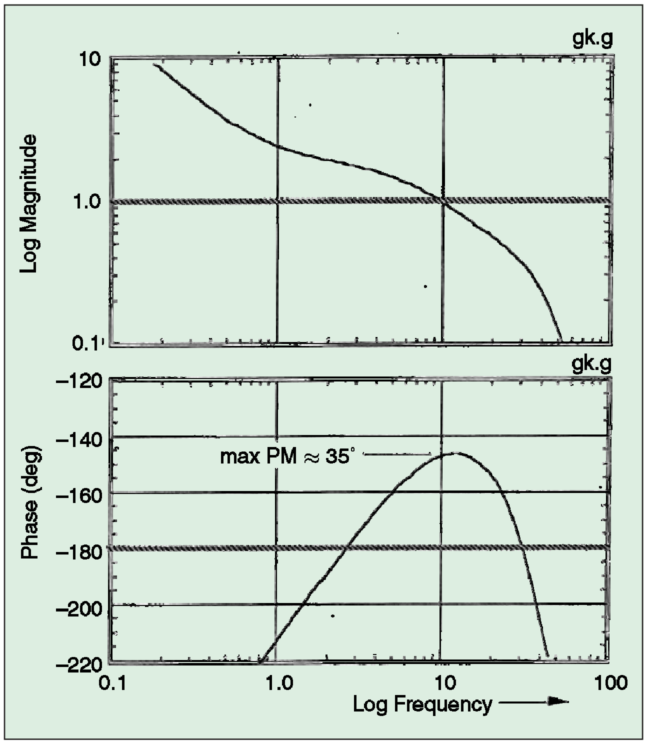


Figure 9. *Prototype Bode diagram for the X-29.*

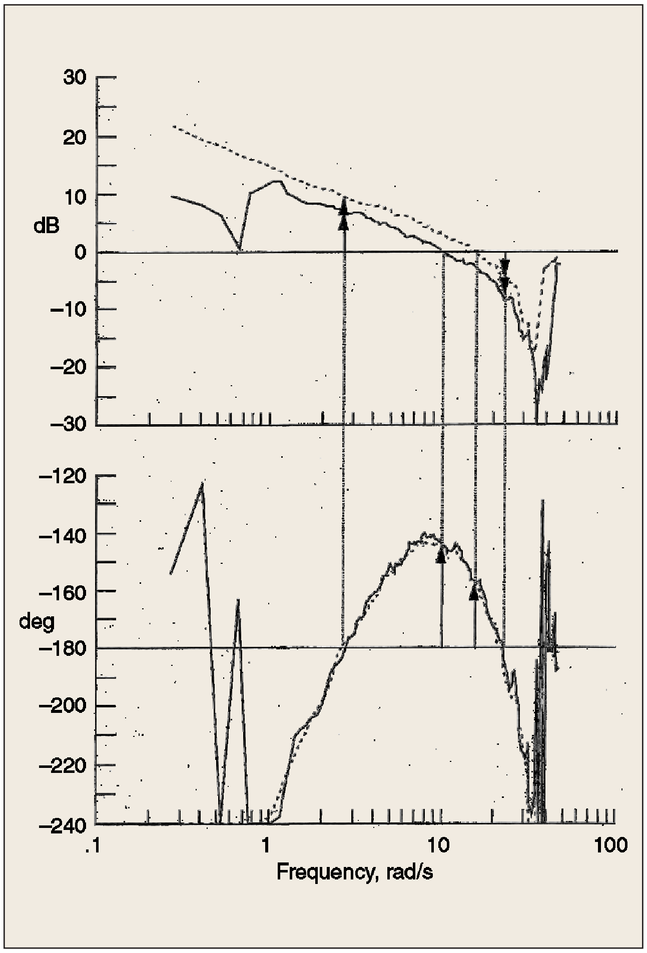


Figure 10. *X-29 flight data (courtesy Mr. J. Gera, NASA).*

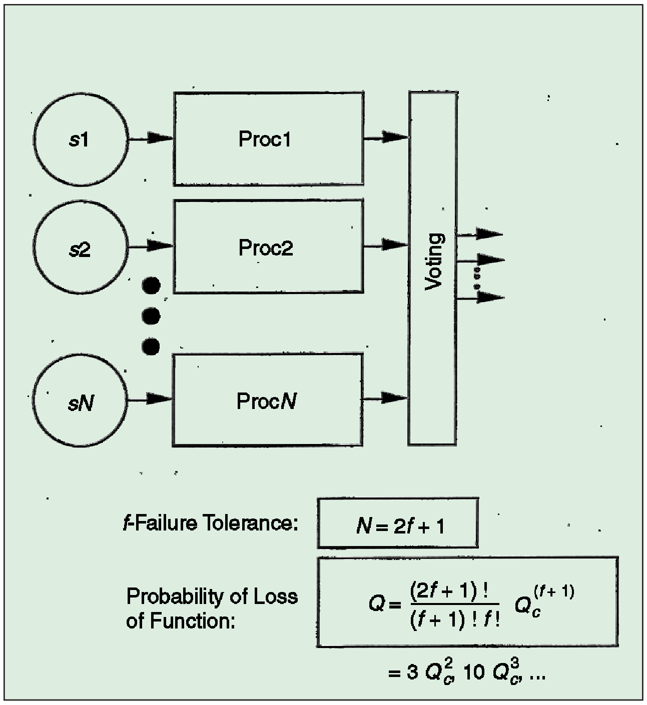

s1
s2
sN
Proc1
Proc2
ProcN
Voting
f-Failure Tolerance:
$N = 2f + 1$
Probability of Loss of Function:
$Q = \frac{(2f+1)!}{(f+1)!\,f!} Q_c^{(f+1)}$
$= 3\,Q_c^2,\ 10\,Q_c^3, \ldots$

Figure 11. *Basic redundant control architecture.*

Fact 2: Operationally Critical Controllers

We now turn to the second fact about unstable/dangerous systems that I want to emphasize, namely, that controllers for these systems are operationally critical. Controllers must work properly for the systems to be safe. In the case of airplanes, such controllers are called "flight critical." Bluntly stated, this means that if the controller fails, you eject or you die.

Although this may be obvious, it is well worth some discussion. In effect, today's control engineers working with dangerous applications must design and build control systems that are so nearly perfect that we are willing to stake our fortunes and sometimes our lives on them.

More often than not, the backbone of a control system involves electronic equipment, and it is safe to say that most of us have had experiences with such equipment—with our TVs, stereo systems, personal computers, etc.—that do not encourage excessive risk taking. Even the very best equipment fails at a rate of 10^{-3} to 10^{-4} failures per operating hour. The job of control hardware designers, therefore, is to build very reliable systems out of unreliable components. This can be done with careful use of redundant elements arranged in architectures that let the overall hardware/software system maintain its function even though components within it have failed. Indeed, an entire engineering subspecialty has evolved to handle such design problems. It is not my intent to cover much of this subspecialty here, but we should be aware of some of its basic concepts and issues. Certainly we need to recognize that this work is every bit as important to control as are the control theories and design tools that are our own specialty.

Perhaps the most basic architectural concept for building reliable systems out of unreliable components is shown in Figure 11. We simply replicate the entire hardware set of a control loop (sensors, processors, and input/output devices) in several channels and arrange for some sort of voting logic to keep only good channels in control. A simple majority-voting scheme suggests that it takes $2f+1$ channels to still be operational after f failures. This is because there are still $f+1$ channels that agree and can outvote the f failed ones. So if a system must be operational after one failure (so-called fail-op), we get a three-channel or triplex architecture. If operation is required after two arbitrary failures (fail-op squared), we get a five-channel architecture, and so on. (In practice, it is customary to claim fail-op squared capability even with only four channels. This relies on an unstated assumption that the two failures are not simultaneous and that the first has been isolated before the second occurs.)

The statistical reliability of the architecture also increases with the number of channels, as given by the formulae in the figure. Q_c is the failure rate of a channel and Q is the overall failure rate of the implementation.

Voting

Before this basic concept sinks in too firmly, let me caution that the picture presented in Figure 11 hides most of the difficulties and issues that make redundancy management a challenging engineering discipline. In particular, the figure makes reference to "some sort of voting scheme" to distinguish good channels from bad. The scheme is not shown, but consider how it might be built. The basic idea is to compare outputs of various channels and to proceed with the majority. However, the outputs are not discrete yes-or-no votes, but are digital words, 8 to 12 bits long. In separate channels, the words will be different because the sensors read different signals and processors run at different rates, and perhaps for other reasons as well. So we cannot simply vote, but must make comparisons against preset thresholds instead.

Figure 12. *Airbus A320.*

©AIRBUS, 2003

This raises the first big issue: How large must the thresholds be? Consider a case where each processor implements a control law that includes a sophisticated control algorithm—say, an adaptive law with explicit identification (e.g., parameter estimates evolving according to $d\phi/dt = e^T\phi + \cdots$). Each channel executes this law with different noisy sensor data. The channel differences will then consist of free integrations driven by noise, and from our first course in stochastic processes, these differences behave like Brownian motion and will exceed any particular threshold infinitely often. This is not good for voting!

Note that the fancy control law is not the culprit here. The same problem arises for any algorithm with free integrations or unstable dynamics, as well as for control laws with mode logic, saturation protections, and other discrete switches. It arises with any control algorithm for which small input differences can produce large output differences.

A common solution to this voting issue is to utilize cross-channel communication to "equalize" all channels (e.g., force them to synchronize clocks and to process identical data). Voting then reduces to simple bit-by-bit comparisons. Unfortunately, this solution only peels back the first layer of difficulty, because the communications hardware needed to equalize channels must also be very reliable. This requires still more levels of redundancy, fortunately restricted to fewer components [6]. Even with appropriately reliable communications, however, there remain certain failures that cannot be detected, namely those that produce identical symptoms in each channel.

Failures with identical symptoms are called "generic faults." At first glance they would appear to be unlikely (two or more identical failures at the same time), until we realize that undiscovered design errors in hardware or software have precisely this characteristic. In fact, the undeniable possibility of such errors has caused most bit-by-bit synchronous architectures to also carry one or more backup channels, using different control laws, different hardware, and different software. The X-29, for example, has a triplex digital system as its primary controller, but carries three backup channels with very basic control laws implemented in analog electronics. The Space Shuttle has a quadruplex digital system as its primary and a fifth dissimilar digital channel as a backup.

Heterogeneous Redundancy

As an alternative to identical primaries with dissimilar backups, other architectures use dissimilar primary channels—different hardware, software, and control laws. Dissimilar channels make the voting problem more difficult (how to set thresholds?), but they alleviate problems with generic faults.

This alternative, incidentally, is one in which some of us have placed a lot of trust. It is used on the Airbus A320 transport aircraft, shown in Figure 12. This airplane was certified for commercial service in 1989 and is now flying regular routes in Europe and the United States. Although the hardware architecture of the A320 does not fit exactly into the prototypes we have looked at, it basically consists of four dissimilar channels. Two channels are built out of one brand of microprocessor, and two others are built out of a second brand. The software in each of these similar hardware channels is different, developed by different design teams using different programming languages. Unfortunately, I do not know much about the control laws themselves, but presumably they are similar enough to permit reasonably tight voting thresholds.

In the context of this article, it is important to point out that the A320 is not statically unstable. However, it is a fly-by-wire airplane, which means that there are no mechanical connections between the pilot's input device (a sidestick hand controller, not the traditional control wheel) and the main control surfaces. Only electronic connections exist through the control computers. If these connections fail, the airplane can continue to cruise and can be brought down with the aid of a low-authority mechanical trim system. On the other hand, complete loss of control functions during critical phases of autoland (e.g., just before touchdown) could well be catastrophic.

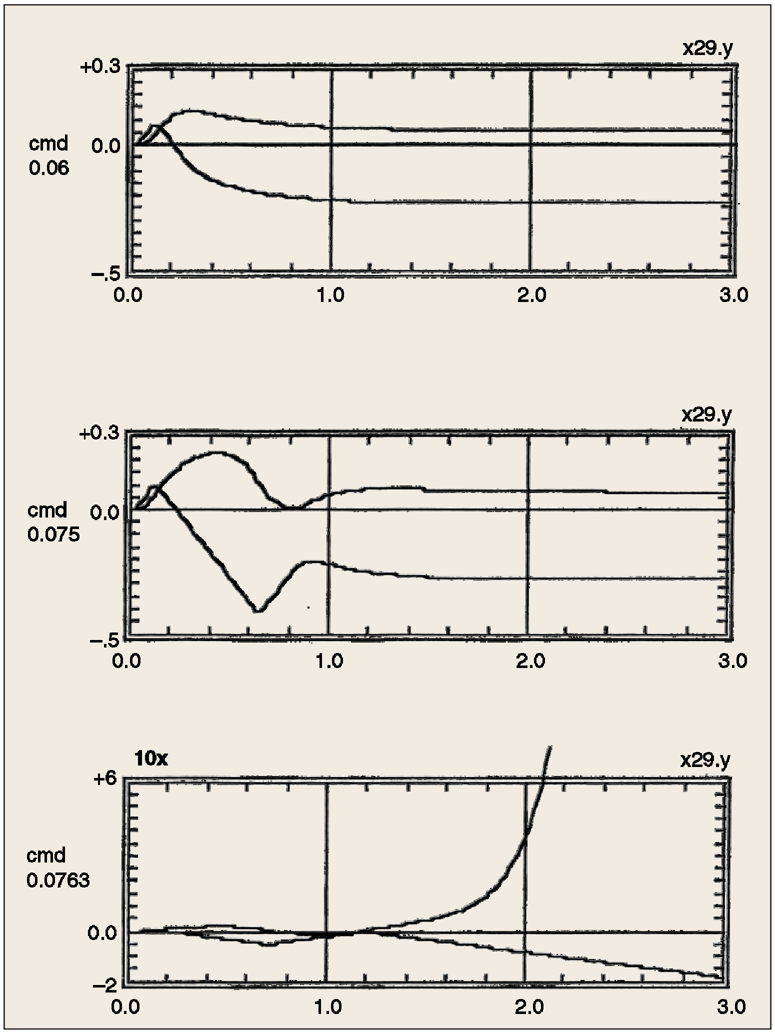


Figure 13. *Simulation traces for the X-29 demonstrating rate limit constraints.*

So in the spirit of my discussions so far, the A320 qualifies as a dangerous system that we should treat with appropriate respect. And the A320 is only the beginning. Both Boeing and McDonnell Douglas expect to build their next transports as fly-by-wire also. It is perhaps comforting to know that neither of these manufacturers expects to build statically unstable commercial transports in the immediate future. Designers at Boeing have told me that unstable fly-by-wire transports are not likely unless they are supersonic. The argument is the same as for the X-29. Relaxed stability for aerodynamic efficiency at supersonic speeds means pronounced instability at subsonic speeds.

Paper studies of new supersonic transports are ongoing but are not likely to produce commercial vehicles until the next century. But the next century is not far away! Engineers who will design the controls for those vehicles are being trained by us right here today. I, for one, still expect to be traveling when these new vehicles come online, so I have a vested interest that we teach the fundamentals well, not just the formal tools and the mathematics!

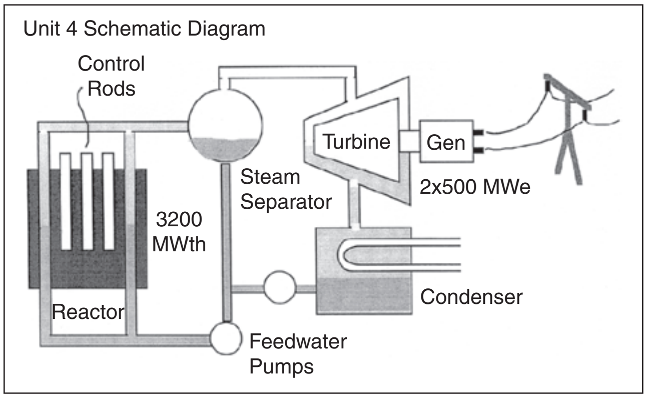


Figure 14. *Schematic diagram of Chernobyl Unit 4 reactor.*

Fact 3: Local Stability

Finally, let us turn to the third fact I want to review: Closed-loop systems with unstable components are only locally stable. This should be well known to all of us. An unstable system cannot be stabilized globally with bounded control authority. For linear systems, this follows directly from the equations. Consider motion along an unstable eigenvector. If the control signal is constrained to have a bounded component along that vector, then there exist initial conditions large enough to keep the state-derivative positive, and the system diverges. This argument holds both for magnitude limits on the controls and for rate limits.

An example of local stability due to rate limits is given in Figure 13. This figure shows responses of the prototype X-29 controller from Figure 9 implemented with rate limits and excited by several different step commands. We see that small commands give linear responses. There is about 100% overshoot, consistent with the (unavoidably) large sensitivity of the design. As commands increase, the control rate saturates, and with very little additional command, the system diverges. Please keep the general characteristics of these traces in mind—the triangular waveform of the control at the edge of instability and finally the dramatic divergence with control rate fully saturated. Later I will present some very similar plots, only much more frightening.

For control engineers, local stability means that special care must be taken to avoid excessive commands (i.e., limits on operator inputs). We must also verify that worst-case upsets due to external disturbances do not drive the closed-loop system out of its region of attraction.

These observations bring us back to our first example, the Saab JAS-39 airplane in Figure 1. This airplane was statically unstable, with divergence rates about half as severe as the X-29. Like the X-29, it carried a triplex digital control system with three analog backup channels. As one might expect, extensive investigations have been going on ever since the airplane's accident, with final official conclusions still not out. However, various reports in the media suggest that the control hardware did not malfunction. Instead, unstable behavior of the control laws in the face of surface rate limits is being cited as the cause. The control laws have already been redesigned to cure the problem (see [7], for example).

Heuristically, we can think of rate limits as acting to reduce available bandwidth for large signal levels (i.e., the controls can only follow position commands whose frequency-amplitude product is less than the rate limit). This suggests that the effects of rate limits can indeed be alleviated by redesigning control laws for lower assumed available bandwidth. According to Bode's integrals, however, such redesigned controllers will achieve poorer sensitivity performance, especially for unstable problems. So the improvements that can be gained for the JAS-39 will necessarily be limited.

Once again, it is evident that unstable/dangerous systems must be treated with more than casual respect.

The Chernobyl Story

Finally, to make this point most dramatically, let me recount the story of the nuclear accident at Chernobyl (Figure 2). On 28 April 1986, news came out of Ukraine that a nuclear power plant had destroyed itself two days earlier and had released significant amounts of radioactive contaminants over a wide area. Short of nuclear war or impending long-term climate changes, this kind of accident certainly looms large in the public mind as one of the more serious threats to our well being.

Whether we choose to recognize it or not, control played a major role in that accident. The plant's hardware did not fail. No valve hung up, no electronic box went dead, and no metallurgical flaw caused a critical part to break. Instead, the reactor control system systematically drove the plant into an operating condition from which there was no safe way to recover. This is true, at least, if we count the control system's hardware, its human operators, and its operating policies as part of the system.

The Plant

To tell this story, we need a few facts about the plant itself. The information comes from a seminar given by Herbert Kouts of Brookhaven National Laboratory summarizing the accident [8]. I want to emphasize, however, that the very simplified control interpretations I am about to make are my own, so the blame for any errors and absurdities is mine as well.

The plant at Chernobyl consisted of four units, each laid out as shown in Figure 14. Unit 4 is the one that experienced the accident. It had a boiling water reactor that took in water at the bottom and heated it to produce a mixture of water and steam. The steam was separated in steam drums and drove conventional turbine-generators to supply power to the distribution grid. There were two turbines rated at 500 MWe (electrical) each and two complete water/steam flow circuits through the reactor. The water/steam circuits operated at about 1,000 psi pressure, with a boiling temperature inside the reactor of about 540 °F.

Chernobyl
Events of the Night of 25 April 1986

- 1) Power had been brought down during the previous day to around 700 MWt, the edge of the legal low-power operating limit, ready to run the test.
- 2) A switchover was made to different flux detectors, better suited for power sensing at this low level, which was apparently a standard procedure. During the switchover, however, the operator neglected to engage the power-hold mode. This oversight set up the conditions for the accident.
- 3) Without automatic power hold, reactor power dropped rapidly to 30 MWt. The operator halted this drop and recovered to 200 MWt by withdrawing control rods.
- 4) With automatic control of power and manual control of feedwater, the plant successfully maintained 200 MWt. Because feedwater flow settings were high, however, the steam void in the reactor dropped to zero. Lower reactivity was compensated for by pulling even more control rods. Only six to eight rods remained in the reactor, far fewer than the minimum number (30) required by operating regulations.
- 5) To avoid automatic shutdown triggered by out-of-range steam drum and feedwater signals, the operator disabled the associated automatic scram control circuits.
- 6) Recognizing excessive feedwater, the operator finally reduced pumping rates. The steam void recovered, producing increased reactivity. Automatic power controls responded to keep power regulated. This response was rate limited and barely stable.
- 7) Next, in preparation for the actual intended test, the operator disabled the automatic scram circuits associated with turbine trip signals.
- 8) Finally, the test was actually started. Steam was cut to the test turbine. The steam void began to rise, and the power controller responded by inserting all three available banks of control rods at maximum rates. This was too little control authority, applied too slowly. A huge power rise followed, up to an estimated 300,000 MWt (100 times rated capacity). The reactor was destroyed. Steam at primary working pressure was released into the reactor containment chamber. The 1,000-ton cover plate of the chamber blew off, and the entire radioactive debris was exposed to the environment.

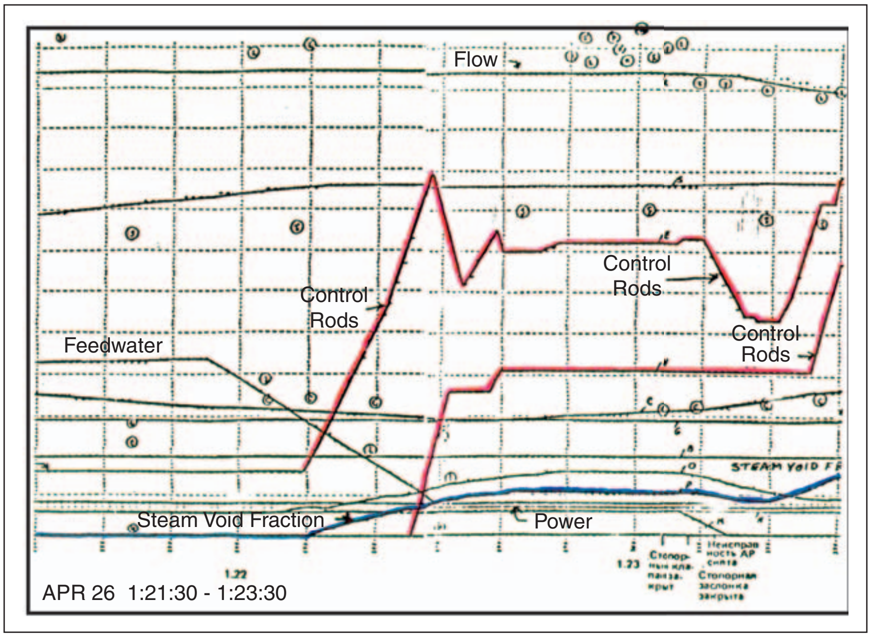


Figure 15. *Trace of reactor events preceding Chernobyl disaster.*

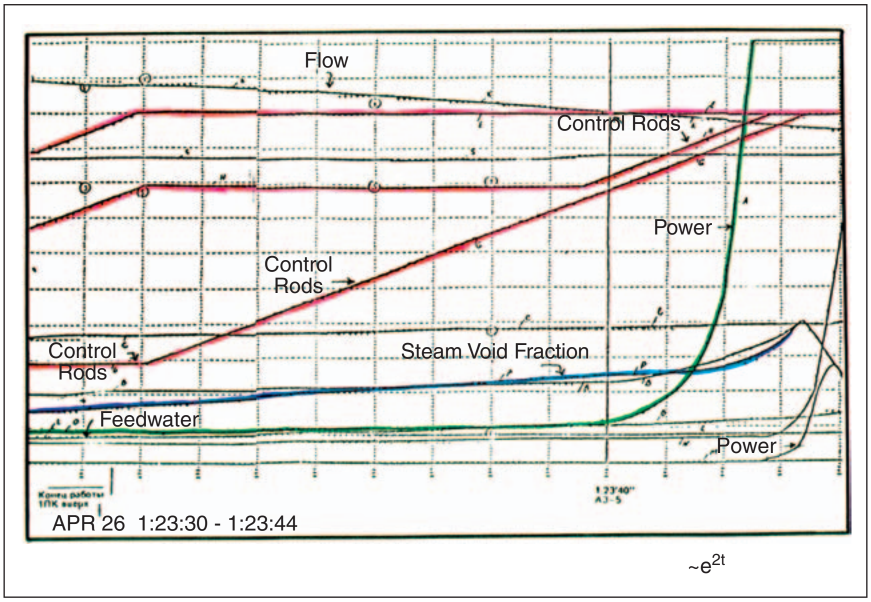


Figure 16. *Trace of power increase associated with the Chernobyl disaster.*

The reactor itself consisted of a graphite core with many pressure tubes (≈1,600) through which the water flowed to be boiled. Each tube contained several fuel rods of uranium oxide that produced the nuclear fission reaction. The generated thermal power was measured by ion flux devices and controlled by several banks of control rods, inserted or withdrawn from the core to moderate the reaction. A second major control input was provided by the feedwater pumps, which regulated flow rates into the bottom of the reactor.

The reason this plant is of interest is that its reactor design has a so-called positive void coefficient. The void refers to the ratio of steam to water in the pressure tubes. The void coefficient refers to the gradient of thermal power with respect to the void. In Chernobyl's design, this coefficient is positive at low power, leading in the linearized sense to a first-order, open-loop instability. Heuristically, as more steam fills the tubes, the reaction becomes stronger (water is apparently a better moderator than steam). This produces more steam, which produces more reactivity, and so it goes.

At nominal operating power, 3,200 MWt (thermal), the positive void coefficient is swamped by other effects and the reactor design is stable. For this reason, operating policies only permitted sustained operation above a minimum power threshold of 700 MWt. Of course, to get the reactor started and stopped, it had to transition through the unstable region.

An Experiment

The accident occurred in connection with an experiment conducted on the electrical side of the plant. The operators wanted to evaluate a scheme for drawing electrical power from a turbine-generator coasting down after a trip from the main power grid. During such a trip, electrical power at the plant is momentarily lost, the reactor scrams (shuts down) automatically, and backup diesels are started to resupply local electrical equipment at the plant. The new scheme would go through this sequence without local power interruption.

The test was scheduled in conjunction with a routine reactor shutdown for maintenance. The intent was to slowly bring power down into the 700- to 1,000-MWt range, load the test turbine-generator with some of the circulation pumps, cut steam to the turbine, and scram the reactor.

A summary of the events of that evening is given in the sidebar. Traces of the last few seconds of this event are shown at the tail end of Figure 15 and continue on an expanded time scale into Figure 16. We can see a slight steam void drop as pressure increased momentarily after steam was cut to the turbine. The steam void then began to build steadily, the control rods dropped in at maximum rate, and finally the power rose uncontrollably. Simple time-to-double calculations approximated from the final power rise place the reactor's unstable pole between 1.5

and 2.0 rad/s, equivalent to a 10-ft broomstick. Again, except for signs, these traces are very similar to Figure 13.

Back to the Punch Line

Why have I taken the time to describe these details of the Chernobyl accident? Certainly I do not want to focus on its tragic consequences, nor do I want to make a case for antinuclear advocates. I simply find this accident to be the most compelling example of blatant disregard for the basic facts I have restated here. We all claim to know these facts, to respect them, and to teach them. Yet the operators at Chernobyl did not appear to know them. Indeed, it can be argued that the plant's designers did not know them either. How else would they be content to write paper regulations prohibiting operation at low power and not insist on reliable hardware to enforce such regulations? On at least two occasions, the operators were able simply to shut off critical safety circuits. Had these shutoffs not been made or not been possible, there would have been adequate time to scram the reactor automatically and to save the plant, even after item 6 of the fateful sequence of events. In redundancy management terms, this possibility qualifies as a generic fault, a single design flaw able to defeat the safety of the entire control architecture.

The Chernobyl accident also provides the most compelling motivation I can think of for us to improve the way we do our job. This reactor control application, as well as the airplane applications I talked about earlier, illustrate that society does indeed permit control engineers to operate dangerous systems. The number of such applications increases steadily. Not all of them have such dramatic consequences as Chernobyl, but they are dangerous nevertheless. Control designers and operators appreciate and respect the practical, physical consequences of these applications only to the extent that we, their teachers, value and instill that appreciation. Unfortunately, our behavior over the past few years at conferences, in our journals, and I suspect also in our lecture halls, places little value upon it. Instead, our behavior tends to uphold mathematical rigor as the only virtue to strive for in control. This trend is incompatible with the trend in applications.

We must place renewed emphasis on stating and teaching the principles of our subject clearly and well. The applications out there are simply too serious for us to hide from responsibility under a cloak of mathematics.

Historical Notes

Much has changed since this article was first presented as the Bode Prize Lecture in December 1989. Here are some brief notes of how certain topics in the lecture have played out:

1) The SAAB JAS-39 accident was indeed attributed to unstable oscillations involving actuator saturations. Control laws were redesigned and retested. The new design experienced a second accident in August 1993, for similar causes. Another redesign followed, this time successful, and the aircraft reached a production milestone with the delivery of 30 aircraft completed in 1996. A prototype of the USAF F-22 fighter also experienced unstable oscillations in 1992, barely avoiding loss of the aircraft.

2) The X-29 research aircraft completed its flight test program without incident. It was retired in 1992 after 374 test flights.

3) After a shaky start, the Airbus A320 fly-by-wire commercial transport has accumulated a strong record of safety and performance and continues to be in service worldwide. It has been joined by Boeing's 777, also fly-by-wire, with a similar strong safety and performance record. No supersonic transports are foreseen in the near future. Design studies during the late 1990s showed their economics to be unattractive using currently projected material and propulsion technologies.

4) The sarcophagus at Chernobyl continues to stand as a stark reminder of the need to "respect the unstable."

References

[1] S. Boyd, C. Barratt, and S. Norman, "Linear controller design: Limits of performance via convex optimization," *Proc. IEEE,* vol. 78, pp. 529-574, Mar. 1990.

[2] I.M. Horowitz and U. Shaked, "Superiority of transfer function over state-variable methods in linear time-invariant feedback system design," *IEEE Trans. Automat. Control,* vol. AC-20, pp. 84-97, Feb. 1975.

[3] J.S. Freudenberg and D.P. Looze, "Sensitivity reduction, nonminimum phase zeros, and design tradeoffs in single loop feedback systems," in *Proc. Conf. Decision and Control,* San Antonio, TX, 1983.

[4] J.S. Freudenberg and D.P. Looze, *Frequency Domain Properties of Scalar and Multivariable Feedback Systems.* New York: Springer-Verlag, 1988.

[5] J. Gera and J.T. Bosworth, "Dynamic stability and handling qualities tests on a highly augmented statically unstable airplane," in, *Proc. AIAA Guidance, Navigation and Control Conference,* Monterey, CA, 1987, AIAA Paper 87-2258.

[6] M. Pease, R. Shostak, and L. Lamport, "Reaching agreement in the presence of faults," *J. Assoc. Comp. Machinery,* vol. 27, no. 2, pp. 228-234, Apr. 1980

[7] *Aviation Week Space Techno.,* Feb. 13, June 26, Oct. 16, 1989.

[8] H. Kouts, "The Chernobyl accident," Seminar Notes, Brookhaven National Lab., Upton, NY, 1986.

Gunter Stein is a chief scientist (retired) of Honeywell Technology Center (now Honeywell Labs). He received a Ph.D. in electrical engineering from Purdue University in 1969. His technical specialization is in systems and control, particularly aircraft flight controls (fighters, transports, and experimental vehicles), spacecraft attitude and orbit controls, and navigation systems for strategic, tactical, and commercial applications. From 1977 to 1997, he also served as adjunct professor in electrical engineering and computer science at MIT, teaching control systems theory and design. He is also active in the development of computer aids for control system design. He was elected Fellow of the IEEE in 1985, awarded the IEEE Control System Society's first Hendrik W. Bode Prize in 1989, elected to the National Academy of Engineering in 1994, and was awarded the IFAC's Nathaniel Nichols Prize in 1999. He can be contacted at 168 Windsor Ct., St. Paul, MN 55112, U.S.A., gunterstein@aol.com.