LIGO PROJECT

Feedback Control System Design Notes
DRAFT

Alex Abramovici, Serge Dubovitsky, Jake Chapsky

June 12, 1997

Abstract

The present notes are an attempt to lay out a process which would help the practitioner in the lab take a control system design problem from requirements through design to a working system, without having to go into the theoretical intricacies associated with feedback control systems.

LIGO-T970233-00-D

Contents

1 Introduction: Feedback Control System Essentials

Carying out an experiment often involves forcing a parameter which is otherwise undetermined, or free runing, to track a reference, to a certain degree of accuracy. A common method employed in tracking applications is to use a feedback control system (FCS). The idea, illustrated in Fig. 1.1, consists of measuring the departure of x, the parameter of interest, from a reference value x_r using a *sensor*(s), processing the sensor output by a *compensator*, and using the processed signal to command an *actuator*(s) whereby appropriate corrections are applied. The total correction is supposed to cancell the free-runing value x_i of x and impart the reference value instead.

FCS design is covered by a highly theoretical discipline, which evolved through contributions by many illustrious authors during the last 150 years. The present notes, however, will trade generality for simplicity and rigorous optimization for the pursuit of "good enough" performance, so that the mathematics can be kept to a minimum. It turns out that this **low-math** approach can result in FCSs with good performance, adequate for many prctical situations in the lab. A few formulae of interest will be derived in the remainder of this Section; the rest of the necessary algebra will be delivered in small portions as the need arises.

At first sight, following the signals in Fig. 1.1 may appear confusing, due to the closed-loop character of the system. Understanding is greatly helped, though, if one assumes that the diagram represents a steady-state situation.[1] Following the arrows in Fig. 1.1, one notes that the difference between the "output" signal x_o on one hand, and the reference x_r plus its noise n_r on the other hand, are converted into an electrical signal $e_e = A(s) \cdot (x_o - x_r - n_r)$. $s = \alpha + i\omega$ is the Laplace variable. Laplace transforms are prefered to Fourier transforms in FCS work because the latter are not defined for such functions of practical interest as e. g. sine-waves. The symbols A, G, H_i and B_i in Fig. 1.1 thus represent transfer functions of the FCS components. The signal e_e is amplified and filtered by the compensator whereby noise is added, so that the Laplace transform of the voltage at the compensator output is $e_c = G(s) \cdot (e_e + n_G)$. This signal is further amplifiedand filtered by the actuator drivers,then converted by the actuators into correction signals

[1]or, alternatively, a frequency domain description. The latter statement will not be proven here.

$x_{ci} = B_i(s) \cdot H_i(s) \cdot e_c$. Two actuators are shown, because it is frequently the case that one actuator alone can not satisfy the tracking requirements. The correction signals are subtracted from the initial free runing value x_i, to yield the output $x_o = x_i - (x_{c1} + x_{c2})$. Eliminating the intermedite variables e_e and e_c from the above equations yields:

$$x_o = (x_r + n_r)\frac{L}{(1+L)} + x_i\frac{1}{(1+L)} - n_G\frac{L}{A(1+L)} \tag{1.1}$$

where $L(s) = A(s)G(s)[H_1(s)B_1(s) + H_2(s)B_2(s)]$ is called the open-loop transfer function or open-loop gain, because it is obtained by cutting the loop in Fig. 1.1 at some arbitrary point, then following the arrows and multiplying the Laplace transforms of the various elements as they are encountered. For $|L| \gg 1$, Eq. 1.1 reduces to:

$$x_o \simeq x_r + n_r - \frac{n_G}{A} + \frac{x_i}{L} \tag{1.2}$$

This equation can be interpreted as follows:

- With the loop closed, the contribution of the free-runing variable to x_o is x_i/L, that is the free-runing variable is suppressed by the loop gain L.

- With the loop closed, the main contribution to x_o is the value of the reference, x_r.

- Closing the loop as shown in Fig. 1.1 adds to the output the noise associated with the sensor and with the compensator.[2]

Thus, closing the FCS loop causes x to track the reference to accuracy ($x_i/|L|$ plus noise in the system), which illustrates the usefulness of FCSs in tracking applications, in particular when high gain and low noise can be obtained.

The correction signal is calculated from from Eq. 1.2:

$$x_c = x_{c1} + x_{c2} = x_i - x_o = (x_i - x_r) - n_r + \frac{n_G}{A} \tag{1.3}$$

which is valid in the high gain limit, as is Eq. 1.2. Eq. 1.3 shows that, for **high gain** and **low noise**, the FCS does what it is intuitively expected to

[2]it is assumed that compensator noise dominates the noise of the actuator drivers and of the actuators. These noise terms were thus disregarded.

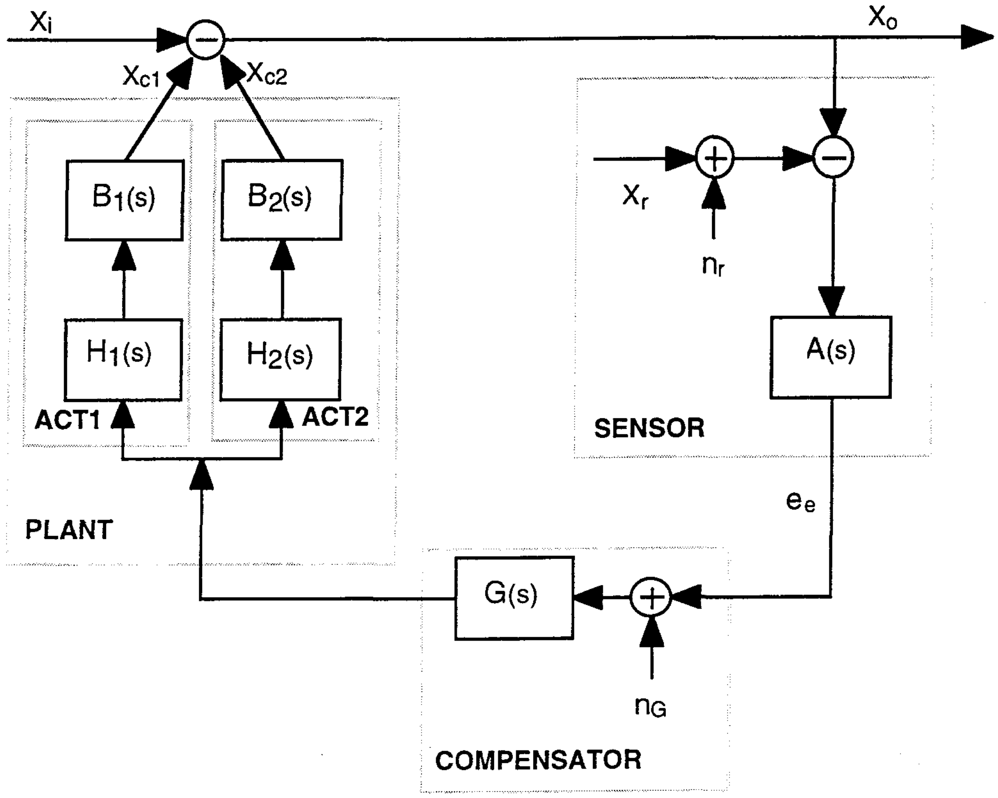


Figure 1.1: Schematic diagram of a tracking system. x_i: uncontrolled value of parameter x, x_o: controlled value, x_r: reference value, x_{c1}, x_{c2}: correction signals, n_r: reference error, e_e: sensor output, n_G: electronic noise refered to compensator input, s: Laplace variable, ACT1,2: actuators, which have arbitrarily been included in the plant. Each actuator consists of a driver with transfer function H and the actuator proper with transfer function B. G is the compensator transfer function, and A is the transfer function of the sensor. G and $H_{1,2}$ are measured in V/V, A is measured in V/units(x), and $B_{1,2}$ are measured in units(x)/V.

do, which is to apply a correction that converts the free-runing value into the reference value for the variable x, in other words it ensures proper tracking of the reference

Designing a FCS comes down to specifying the sensor, the compensator and the actuator(s) such that proper tracking, i. e. high gain and low noise, are achieved. An added requirement is that the FCS system be oscillation free, i. e. stable, as closed loop operation can engender instability.[3] Laying out a path for designing a working FCS with specified tracking accuracy and stable operation is the object of these notes. The procedure described in what follows is not unique or general, neither does it achieve optimum performance in a rigorous sense. Nonetheless, it provides a practical approach to many FCS design problems encountered in the laboratory.

[3]FCS stability is often the main topic in control systems textbooks

2 Feedback Control System Stability

2.1 Stability, Differential Equations and Laplace Transforms

The dynamics of many systems of interest can be described by differential equations; this allows stability to be defined and described in a precise mathematical language. In general, the evolution of a parameter x associated with the system is described by the n-th order differential equation with constant coefficients:

$$a_0 \frac{d^n x(t)}{dt^n} + \frac{d^{n-1} x(t)}{dt^{n-1}} + ... + a_n x(t) = f(t) \tag{2.1}$$

where $f(t)$ is called the forcing function. The general behavior of $x(t)$ is governed by the solution of the homogeneous counterpart of Eq. 2.1, i. e. by the solution corresponding to $f(t) = 0$, which can be written:

$$x_h(t) = \sum_{i=1}^{n} c_i e^{r_i t} \tag{2.2}$$

where c_i are constants depending on the initial conditions, and r_i are the roots of the characteristic equation associated with Eq 2.1:

$$\sum_{k=0}^{n} a_k r^k = 0 \tag{2.3}$$

The sum on the left-hand side of Eq. 2.3 is called $P_n(r)$, the characteristic polynomial associated with Eq. 2.1. When multiple roots occur, Eq. 2.2 has a slightly different form, but is still a sum of exponentials. A system is stable when $x(t)$ or any other function which describes the system are bounded as $t \to \infty$. It is worth noting that for complex roots $r_i = \sigma_i + \omega_i$, the real part represents a real exponential, while the imaginary part describes an oscillatory behavior. In other words, **a system is stable when the real parts of the roots of the characteristic polynomial are all negative or zero**, which ensures that the real exponentials in Eq. 2.2 are not becoming unbounded. An equivalent statement is that the characteristic polynomial should have no roots in the right-hand side of the s-plane.

In order to translate the above formulation of the stability criterion into common control system language, note that the Laplace transform[4] $X(s)$ of the solution of Eq. 2.1 is:

$$X(s) = \frac{F(s)}{P_n(s)} \tag{2.4}$$

where $s = \sigma + j\omega$. Note that the denominator in Eq. 2.4 is the characteristic polynomial associated with Eq. 2.1. $L(s) \equiv 1/P_n$ is called the **transfer function** of the system, as it allows to calculate the Laplace transform of the **output** $X(s)$ if the Laplace transform of the **input** $F(s)$ is known. In terms of $L(s)$, the condition that there be no roots of Eq. 2.3 in the right-hand side of the s-plane translates into the requirement that the **transfer function have no poles in the right-hand side of the s-plane**. When a feedback control system is built by closing the loop as in Figs. 1.1,2.3, the transfer function of the new system is $1/(1+L(s))$, according to Eq. 1.1,[5] and stability requires that this function have no poles in the right-hand side of the s-plane. Much of the classic feedback control system research revolved around using this condition as a starting point for deriving closed loop stability criteria in terms of the properties of $L(s)$. Two of the most frequently used criteria are given below without derivation:[6]

- **The Nyquist Criterion**
  The Nyquist stability criterion relies on the open loop transfer function calculated for real frequen-cies, that is on $L(j\omega)$. The imaginary part of L is plotted against the real part, for increasing values of ω. The closed loop system is stable if the curve does not encircle the -1 point on the real axis clockwise.[7] For illustration, Fig. 2.1 shows a Nyquist plot corresponding to a stable system.

- **The Bode Criterion**
  The Bode criterion requires that $\phi(L) < 180^o$ at frequencies where

[4]The definition of the Laplace transform and other transformations between the time domain and the frequency domain can be found in many textbooks, e. g. A. V. Oppenheim, A. S. Willsky,Signals and Systems, Prentice Hall, 1983

[5]in the absence of noise and for $x_r = 0$.

[6]see e. g. J. J. DiStefano *etall*, Feedback and Control Systems, Schaum's Outline Series, McGraw-Hill, 1990.

[7]While this is not themost general formulation of the Nyquist criterion, it is sufficient for most practical design ccases

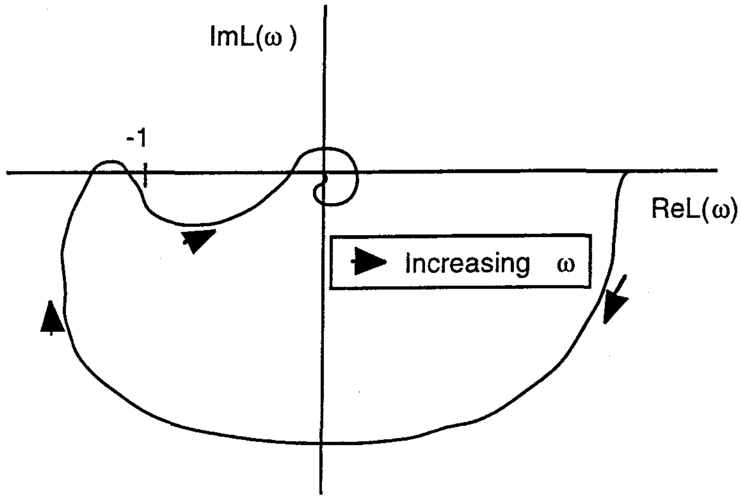


Figure 2.1: Example of Nyquist plot. The system is stable even though at some frequencies $\phi(G) > 180^o$.

$|L| = 1$. Often, use of the Bode criterion is based on plots of the magnitude and phase of the transfer function versus frequency, as in Fig. 2.5.

2.2 Poles and Zeros

Poles

Consider the RC network of Fig. 2.2. Assume the input is a sine-wave $e_i = \sin \omega t$.
The differential equation which describes this system is:

$$RC\frac{de_o(t)}{dt^n} + e_o(t) = \sin \omega t \tag{2.5}$$

with the solution:

$$e_o(t) = \frac{s_0}{\sqrt{s_0^2 + \omega^2}} \sin(\omega t - \phi) \tag{2.6}$$

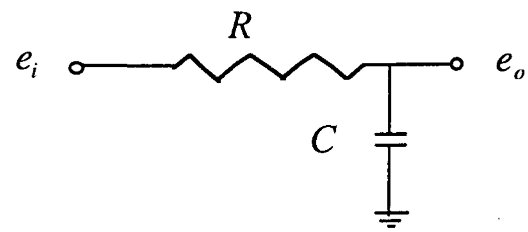


Figure 2.2: RC network used to illustrate the concept of pole.

where $\phi = \tan^{-1}(\omega/s_0)$, and $s_0 = 1/RC$. The output is a phase-shifted version of the input, with $\phi = -45^o$ for $\omega = -s_0$ and $\phi \to -90^o$ for $\omega \gg s_0$. For $\omega > s_0$, the input is also attenuated at the rate of a factor 10 per decade of frequency (20 dB/decade). In electrical engineering terms, the network of Fig. 2.2 is a low-pass filter. The Laplace transform of $e_o(t)$ is obtained from Eq. 2.5 by replacing $d/dt \to s$, $\sin \omega t \to S(s)$ and solving the algebraic equation. The result is:

$$E_o(s) = \frac{s_0}{s_0 + s} S(s) \tag{2.7}$$

The transfer function $s_0/(s_0+s)$, which describes the network, has a real **pole** at $-s_0$ in the complex s-plane. For physical frequencies, i. e. for $s = i\omega$, one finds that the amplitude of the input sine-wave is attenuated when $\omega > s_0$, as in Eq. 2.6. The imaginary unit multiplying ω describes a 90^o phase lag at frequencies well above s_0.

Another case of interest is the driven harmonic oscillator, described by the second order equation:

$$\frac{d^2x(t)}{dt^2} + 2\gamma\frac{dx(t)}{dt} + \omega_0^2 x(t) = \sin\omega t \tag{2.8}$$

with the solution:

$$x(t) = Ae^{(-\gamma+i\omega_1)t} + Be^{(-\gamma-i\omega_1)t} + \frac{\sin(\omega t+\phi)}{\sqrt{(\omega^2-\omega_0^2)^2 + 4\gamma^2\omega^2}} \tag{2.9}$$

where $\phi = \tan^{-1}\left[2\gamma\omega/\left(\omega^2 - \omega_0^2\right)\right]$. $\omega_1 = \sqrt{\omega_0^2 - \gamma^2}$ is real for small γ. The first two terms in Eq. 2.9 vanish after sufficient time has passed. The remaining term is attenuated for $\omega > \omega_0$, at a rate of ω^{-2}, i. e. a factor 100/decade (40 dB/decade). For $\omega = \omega_0$, $\phi = -90^o$ and, for $\omega \gg \omega_0$, $\phi \rightarrow -180^o$. The Laplace transform of $x(t)$ is:

$$X(s) = \frac{S(s)}{(s-s_1)(s-s_2)} \tag{2.10}$$

The transfer function $(s-s_1)^{-1}(s-s_2)^{-1}$, which describes the harmonic oscillator and has two complex conjugate poles at $s_{1,2} = -\gamma \pm i\sqrt{\omega_0^2-\gamma^2}$, displays the ω^{-2} behavior above ω_0 and the phase lag already seen in the solution of the differential equation.

Differential equations with order higher than 2 are more difficult to integrate than the examples considered above. However, the denominator of the corresponding Laplace transforms is a polynomial consisting exclusively of factors like the denominators in Eqs. 2.7,2.10, as long as the systems are described by linear differential equations with constant coefficients. Thus, the corresponding Laplace transforms will have single real poles or pairs of complex conjugate poles.

The preceding discussion of s-plane poles associated with systems described by linear differential equations with constant coefficients is consistent with the following statements, valid for either type of poles:

1. The presence of a pole is indicative of low-pass filter behavior.

2. For real poles, the cut-off frequency of the system, i. e. the frequency at which the low-pass filtering behavior sets in, is equal to the position of the pole.

3. For a complex pole, the cut-off frequency is equal to the imaginary part of the pole.

4. For either type of pole, the roll-off, i. e. the rate of attenuation above the cut-off frequency, is 20 dB/decade/pole.

5. Poles introduce a phase lag:

   - $\phi = -45^o$/pole at the position of the pole.
   - $\phi = -90^o$/pole at frequencies far above the position of the pole.

Zeros

The presence of poles in the transfer function associated with Eqs. 2.1,2.5,2.8 is related to the derivatives of the unknown function. If derivatives of the forcing function in the right-hand side of the differential equations are present, the transfer function will have a polynomial in the numerator. The roots of the latter are called zeros of the transfer function. Similar to the case of poles, the following statements can be made regarding zeros:

1. The presence of a zero is indicative of high-pass filter behavior.

2. For real zeros, the set-in frequency of the system, i. e. the frequency at which the high-pass filtering behavior sets in, is equal to the position of the zero.

3. For a complex zero, the set-in frequency is equal to the imaginary part of the zero.

4. For either type of zero, the roll-up, i. e. the rate of amplification above the set-in frequency, is 20 dB/decade/zero.

5. Zeros introduce a phase lead:

   - $\phi = 45^o$/zero at the position of the zero.
   - $\phi = 90^o$/zero at frequencies far above the position of the zero.

Both the magnitude and the phase shift of the frequency response are readily measured quantities. The properties of poles and zeros, listed above, can thus be used to gain insight about the number of poles and zeros of the system at hand, by inspection of the magnitude/phase plots.

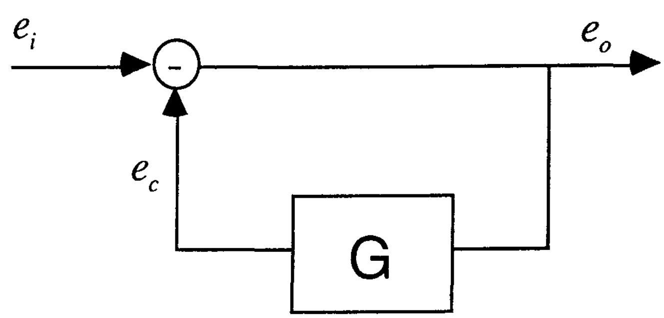


Figure 2.3: Diagram of error suppression feedback system, used as an example for stability discussion. Input, output, and correction sig-nals are shown.

Finally, it can be loosely stated that the steeper the roll-off, the larger the related phase lag and, on the other hand, the steeper the roll-up, the larger the related phase lead.[8] This often causes difficulties in the design of FCSs, when the desire for fast gain roll-off abround and above unity gain frequency is frustrated by instability caused by the associated high phase lag.[9]

2.3 System Stability: Low-Math Approach

The discussion of stability will be based on a simplified version of Fig. 1.1, shown in Fig. 2.3, where there is only one actuator, the sensor and the actuator have unit transfer functions and all signals are voltages. The corresponding simplified forms of Eqs. 1.1,1.3 are:

[8] At a deeper level, this relates to causality, via the Kramers-Kronig relations.

[9] steep filtering above unity gain is desirable in order to provide passive suppression of electronic noise in a frequency range where the loop is effectively open and thus no noise attenuation by the FCS is available.

$$e_0 = \frac{e_i}{1+G(s)} \tag{2.11}$$

$$e_c = \frac{e_i G(s)}{1+G(s)} \tag{2.12}$$

The ususal definition of stability will be used: a system is stable when it settles following a disturbance. The system of Fig. 2.3 will be tested for stability with a disturbance having flat spectrum with sharp cut-off:

$$X(f) = \left\{ \begin{array}{ll} 1 & |f| < BW \\ 0 & |f| > BW \end{array} \right\} \tag{2.13}$$

which has the sinc function $(1/\pi t)\sin 2\pi BWt$, plotted in Fig. 2.4, as its time domain counterpart. It is also worth noting that the single sided frequency bandwidth BW is equal to the inverse of the pulse duration T. The sharp cut-off in test disturbance spectrum makes it possible to explore the connection between stability and the properties of the system transfer function $G(s)$ in arbitrarily chosen frequency bands.

A particular transfer function $G(s)$, with three poles at 0.1 Hz, a zero at 0.31 Hz and unity gain at 1 Hz, will be used as an example. This function, plotted in Fig. 2.5, has a phase shift of 180.1^o at the unity gain frequency, thus the closed loop system is unstable, according to the Bode criterion.

In the absence of feedback, the system of Fig 2.3 is assumed to be stable. The onset of instability must then be connected with the presence of the correction signal, which is subtracted from the input signal in order to generate a null output. Inspection of Eq. 2.12 suggests that there are three distinct regimes for the correction signal:

Frequencies below the unity gain point of $G(s)$.

$|G(s)|$ is high by definition (see the example in Fig. 2.5), thus, according to Eq. 2.12, the correction signal is approximately equal to the input signal; the input disturbance is successfully suppressed and there is no ringing at the output.[10] Fig. 2.6 illustrates this point.[11] Indeed, when a 4s test pulse is fed into the system, pulse suppression and rapid settling is seen, even though, according to the Bode criterion, the system is expected to be unstable. Stable behavior in this regime is expected regardless of the phase of $G(s)$, as long

[10]the width of the test pulse would be larger than the inverse of the unity gain frequency, as its spectrum needs to drop to zero before unity gain is reached.

[11]System response was calculated using Matlab.

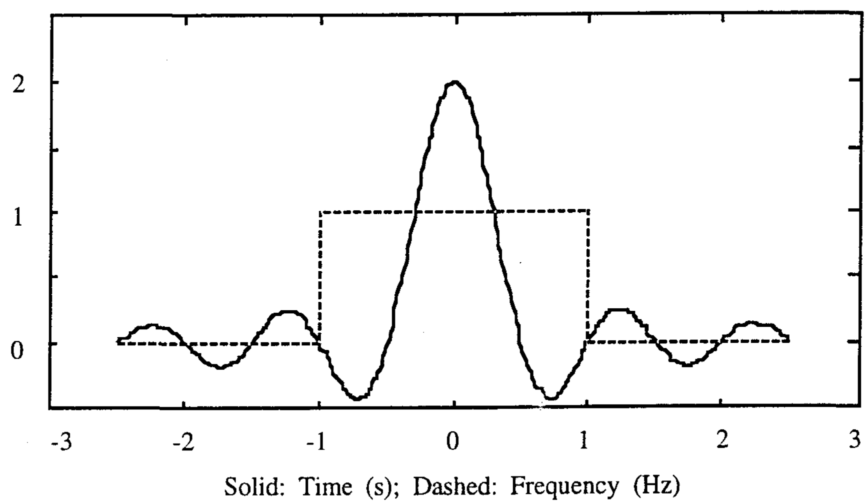


Figure 2.4: Pulse used to examine system stability. Solid line: time dependence, dashed line: frequency spectrum. The pulse shown has 1s width, measured between zero crossings next to the central peak. The single sided bandwidth is 1 Hz.

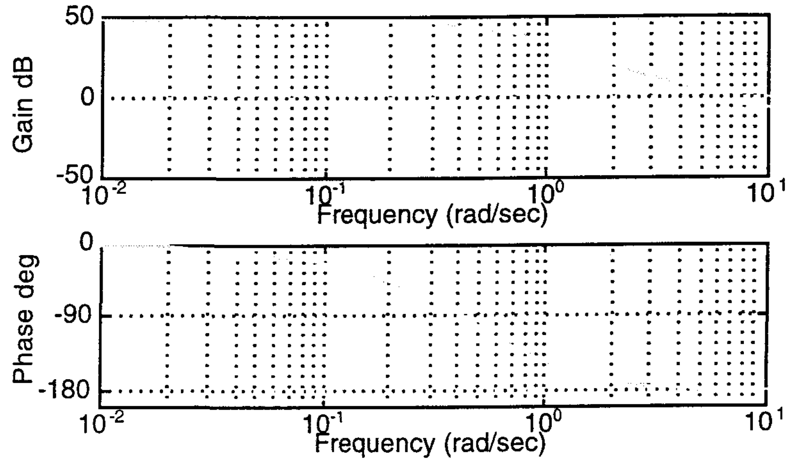


Figure 2.5: Bode plot of open loop sample transfer function used to illustrate this discussion.

as the gain is high. This is of advantage when high gain is necessary at low frequencies, in spite of a relatively low unity gain frequency, which requires that the gain be dropped rapidly. Since each simple low-pass filter contributes 90^o, the phase lag can stack up quite high.

Frequencies above the unity gain point.
In this regime, $|G(s)|$ is low, and the correction signal is small as well. Since little correction is applied to the input, the latter sails through the system with little attenuation, as if there were no feedback. In this regime there is no need to worry about instability.

Frequencies at and around the unity gain point.
From Eq. 2.12 one finds that system behavior is likely to depend strongly on the phase of $G(s)$, because of the expression in the denominator, where a complex number with modulus close to unity is added to 1.

- If the phase of $G(s)$, ϕ, is small, the correction signal will be smaller

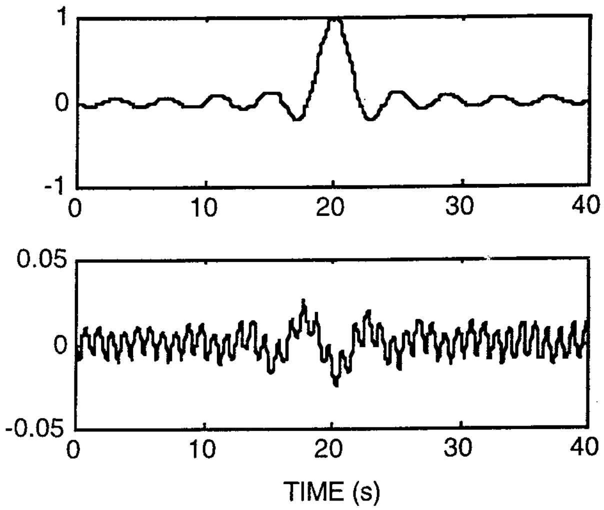


Figure 2.6: System response to a 4s pulse (0.25 Hz bandwidth). Upper trace: input pulse; lower trace: system output. Note that the pulse is suppressed 40 times, consistent with the open loop gain around 0.25 Hz (Fig. 2.5). Some 1 Hz ringing, excited by wideband noise due to the computation process, is present.

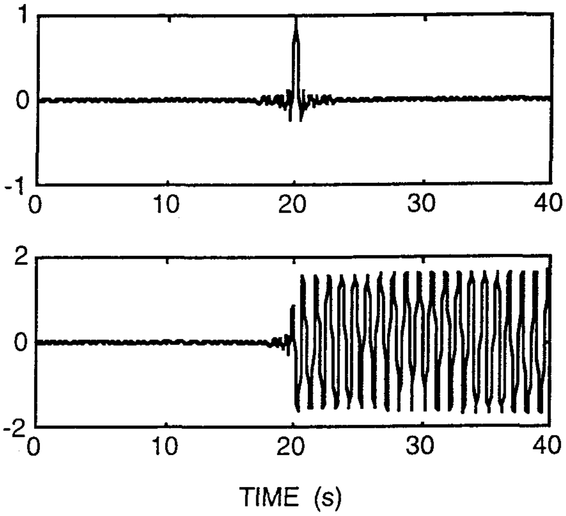


Figure 2.7: System response to a 0.5s pulse (2 Hz bandwidth). Upper trace: input pulse; lower trace: system output. Note that the ringing is slowly building up, as expected with a very low phase deficit (0.1^o) at the unity gain frequency.

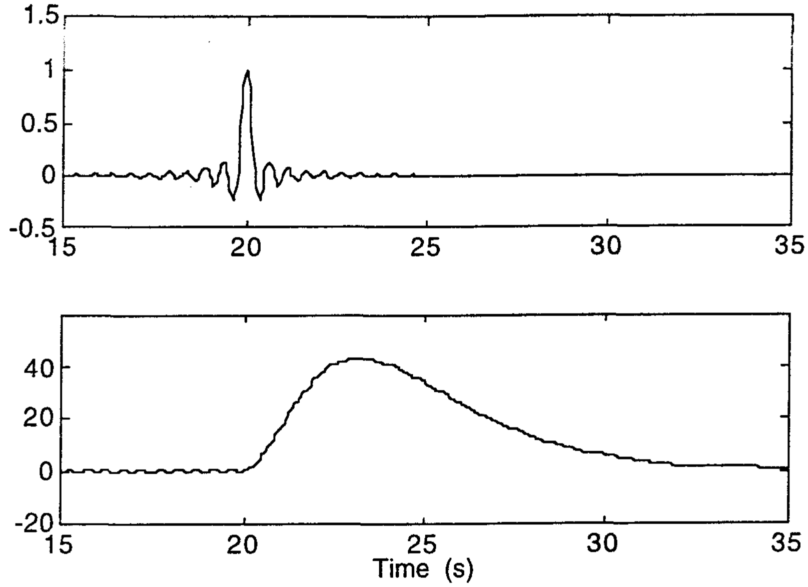


Figure 2.8: Upper trace: input pulse, lower trace: input pulse distorted and delayed by the feedback network. For this example, $G(s)$ has 3 poles at 0.1 Hz, unity gain frequency of 1 Hz, and approximately 256^o phase lag at unity gain. At the correction point, the processed pulse misses the input pulse which is there-fore not attenuated; the result is oscillation.

than and slightly out of phase with the input test pulse, but input attenuation will still occur, albeit less effectively. The test pulse will pass through with some attenuation, and the system will slowly settle.

- If ϕ is close to 180^o, bad behavior is expected, as the denominator in Eq 2.12 tends to vanish. The key effect here is that the correction pulse becomes very large. With real systems, this is made worse by saturation in the amplifiers, which usually adds to the phase shift. This, in turn, causes the correction pulse to miss the input pulse at the correction point, so that two pulses are now traveling trough the system. A third one is then generated, and there will be a never ending succession of pulses at the output, caused by just one input pulse. In other words, the system is drastically unstable. The adequate test pulse for this regime has a spectrum extending beyond the unity gain frequency and duration less than the inverse of the unity gain frequency. Fig. 2.7 illustrates that the system does, indeed, display ringing in response to such a pulse.

- If ϕ is well above 180^o, for example $\sim 270^o$, no anomalous correction signal is generated, as the two terms in the denominator of Eq. 2.12 do not cancel. After passing through the feedback network, the signal will be lagging 270^o behind the input pulse, which it thus misses, as illustrated in Fig. 2.8, and a series of pulses will be seen at the output. A real system, with noise at the input, would never really settle. It should be stressed here that for instability to occur, it is not necessary to have the denominator of Eq 2.12 vanish; any phase shift larger than 180^o would do.

Conclusion: to ensure closed loop stability, the phase of $G(s)$ should be less than 180^o when $|G(s)| = 1$. In practice, one usually requires that the phase lag at unity gain be kept below 120^o-135^o. The intuitive discussion above led to the Bode stability criterion. It is stressed that at frequencies where the gain is substantially higher than 1 there is really no restriction on the phase lag.

3 Example of Design Problem: Laser Frequency Stabilization

In order to make these notes easier to follow, an FCS design problem is presented in this section. The example which has been chosen, frequency stabilization of a Nd:YAG laser, contains many of the ingredients found in a wide variety of cases of interest. Various aspects of the example will be used in later sections to illustrate aspects of the design process.

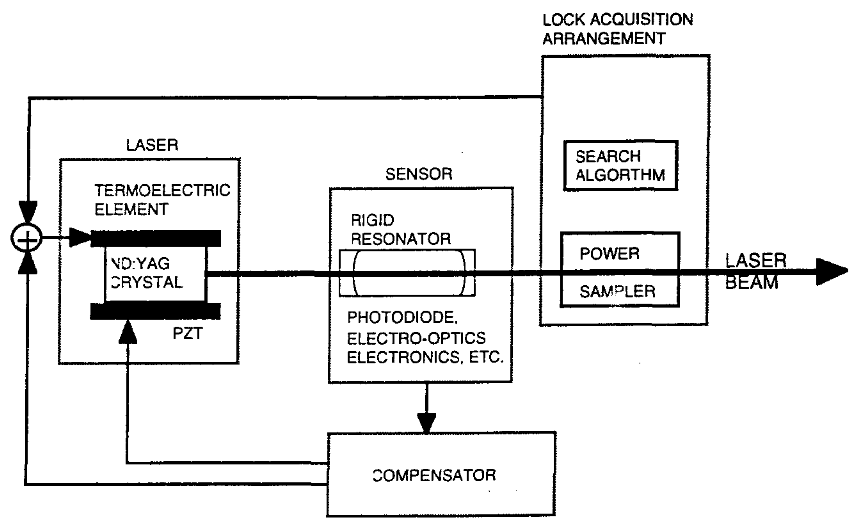


Figure 3.1: Concept of laser frequency stabilization system. The sensor and the actuators will be discussed in this section. Lock acquisition will be addressed later.

In terms of the language adopted so far, x, the variable to be controlled, is the frequency of the laser. The input variable x_i is the fre-runing value of the frequency, which is fluctuating, i. e. affected by frequency noise. The purpose of frequency stabilization is to attenuate the frequency fluctuations. In the example dicussed here, this is to be achieved by using a FCS which forces the frequency to track a stable reference. According to Eq. 1.2, x_i is attenuated by a factor equal to the magnitude of the open loop gain, $|L|$. If the frequency stabilization system only aims at obtaining a stable frequency,

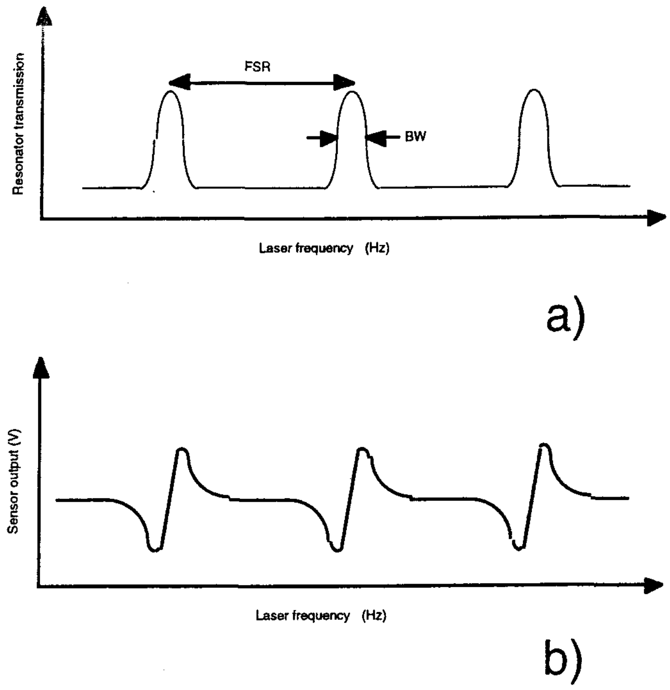


Figure 3.2: a) Power transmission of an optical resonator versus frequency. **BW**: resonator bandwidth, **FSR**: free spectral range, defined as $c/2l$, where c is the speed of light and l is the resonator length. b) Main features of output signal from a Pound-Drever-Hall frequency sensing arrangement.

but the value of the stable frequency is not important, the corresponding FCS belongs to a class called **error attenuation systems**.

3.1 System Concept

The concept for stabilizing the frequency of a Nd:YAG laser is shown in Fig. 3.1. The laser considered in this example is a monolithic Nd:YAG laser. The laser itself consists of a Nd:YAG crystal a few millimeters in size, which is the gain medium. The beam follows a closed path through the crystal, which is also the laser resonator. The laser frequency is determined by the legth of the optical path through the crystal. Slow frequency fluctuations result from changes in crystal temperature, while fast frequency fluctuations are caused mainly by fluctuations in the intensity of the pump source.[12]

[12] laser diodes are the prefered pump source for solid state lasers with output power up to a few Watts.

Sensor

Detecting laser frequency fluctuations relies on coupling the light to a rigid optical resonator, consisting of two mirrors attached to a rigid spacer, such that their optical axes coincide. Typically, mirror transmission is chosen at a fraction of a percent, so that multiple overlapping beams propagate and interfere with each other inside the resonator. When the resonator length is an integral multiple of half of the wavelength of the light, a standing wave patern builds up inside the resonator, a condition which is known as **resonance**. At resonace, the resonator transmission reaches its maximum value, and so does the power circulating inside the resonator. If the frequency changes by a **free spectral range, FSR**, defined as $c/2l$, the cavity length becomes equal to the next higher integral multiple of the half-wavelength and another resonance is reached, as shown in Fig. 3.2a. The dependence of transmission on frequency suggests the possibility of using the resonator as a frequency fluctuation sensor. The reference in this case is one of the resonant frequencies, which in turn is related to the length of the resonator; hence the choice of a rigid resonator, to ensure stable length. Measuring the transmitted power is not usefull for FCS implementation, since it is symmetric with respect to resonance and therefore indicates only that the frequency is off resonance, but not which way it is off. Among many ways devised to overcome this limitation, the Pound-Drever-Hall method has established itself for use in high performance laser frequency stabilization systems *** cite Drever-Hall paper ***. A sample output signal corresponding to the Pound-Drever-Hall sensing arrangement is shown in Fig. 3.2b. As long as the laser frequency is within the resonance bandwidth, the sensor output is approximately linear. The slope of the output, $\Delta V/\Delta\nu$, measured over a frequency band $\Delta\nu$ centered on resonance and expressed in V/Hz, is the **gain** of the sensor. A remarkable feature of the Pound-Drever-Hall sensor output is that when the frequency moves out of the resonance bandwidth, the slope levels off, i. e. the gain decreases. Thus, if this sensor is used in a FCS which performs well on-resonance, pulling the frequency off-resonance could seriously diminish system performance. Therefore the **resonance bandwidth** defines a **sensor range**. The laser frequency has to be within the sensor range for the system to function reliably.

The trace in Fig. 3.2b is obtained when the laser frequency is scanned slowly through resonance. For faster scanning rates,[13] the fact that the field is

[13]also called **Fourier frequencies** in order to distinguish them from the actual laser

stored in the resonator over time scales $\sim 1/BW$ influences the measurement; the faster the scanning rate, the lower the measured gain. Dependence of the gain and phase shift on the Fourier frequency, for a resonator with 200 kHz bandwidth, is shown in Fig. 3.3. The **Bode plot** of Fig. 3.3 can be seen as a model of the frequency fluctuation sensor, detailed enough for FCS design.

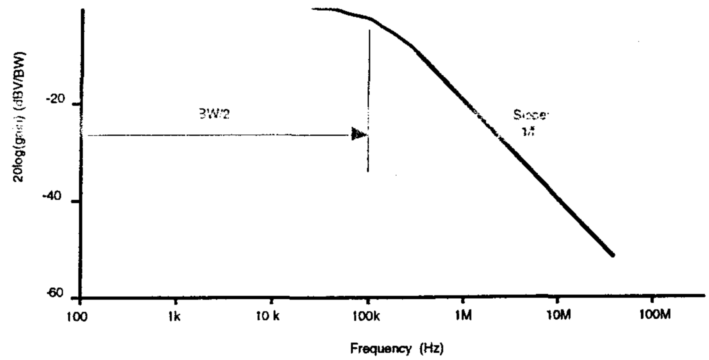
In summary, the sensor compares the laser light half-wavelength with the length of the resonator. When the latter is not an integral multiple of the former, a voltage proportional to the difference is generated at the sensor output.

Anything which changes the sensor output when the laser frequency does not change is defined as sensor noise or error. Examples of sensor noise/error are:

- Changes in resonator length due to ambient temperature variations.
- Changes in resonator length due to environmental vibration and acoustic excitation.
- Changes of the optical path inside the resonator due to fluctuations in air pressure, if the resonator is in air.
- Electronic noise associated with the sensor.

Actuators

The actuators for correcting the laser frequency are a provision for laser crystal temperature control and a piezo-electric device (PZT) bonded to the laser crystal. Both change frequency by changing the optical path inside the laser crystal. Temperature control provides for changing the frequency over a range of about 5 GHz, while the PZT can be used to tune the laser over approximately 100 MHz. Approximate typical Bode plots for the actuators are shown in Fig. 3.4. It can be seen from the plots that the temperature control has a substantially higher gain than the PZT. At the same time, temperature-induced frequency changes start showing a substantial phase shift around 0.1 Hz, while the PZT-induced ones do not show significant phase shifts up to about 1 MHz. Therefore, the temperature control can be used to effect slow, but large frequency corrections, while the PZT is appropriate for small amplitude, fast frequency corrections.



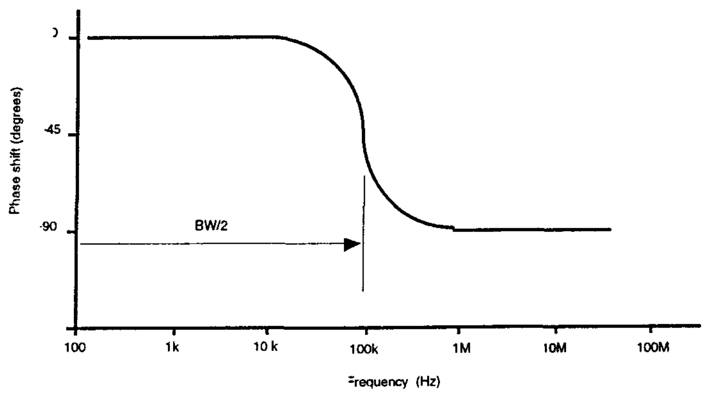


Figure 3.3: Sample Bode plot of output characteristic from a Pound-Drever-Hall frequency fluctuation sensing arrangement. Note that both the gain and the phase characteristics are flat from DC to roughly half the resonator bandwidth.

3.2 Tracking Requirement

A tracking requirement[14] could be formulated as follows:

1. The laser frequency shall be stable within a 10 kHz range, peak-to-peak, once lock to the reference resonator has been acquired.[15] Note that this is a really though requirement. Indeed, since for $\lambda = 1.064$ μm $\nu = 3 \cdot 10^{14}$ Hz, the laser frequency has to be stable to one part in $3 \cdot 10^{10}$.

2. There is no requirement as to what the center frequency of this range should be.

3. Lock acquisition should be automatic, and the acquisition sequence should be intiated automatically whenever the system is out of lock.

[14]more appropriately called frequency stability requirement, in this case

[15]i. e. the laser frequency is within sensor range and the FCS loop is closed and functioning. A short discussion of "lock" is given in Section 4.3.

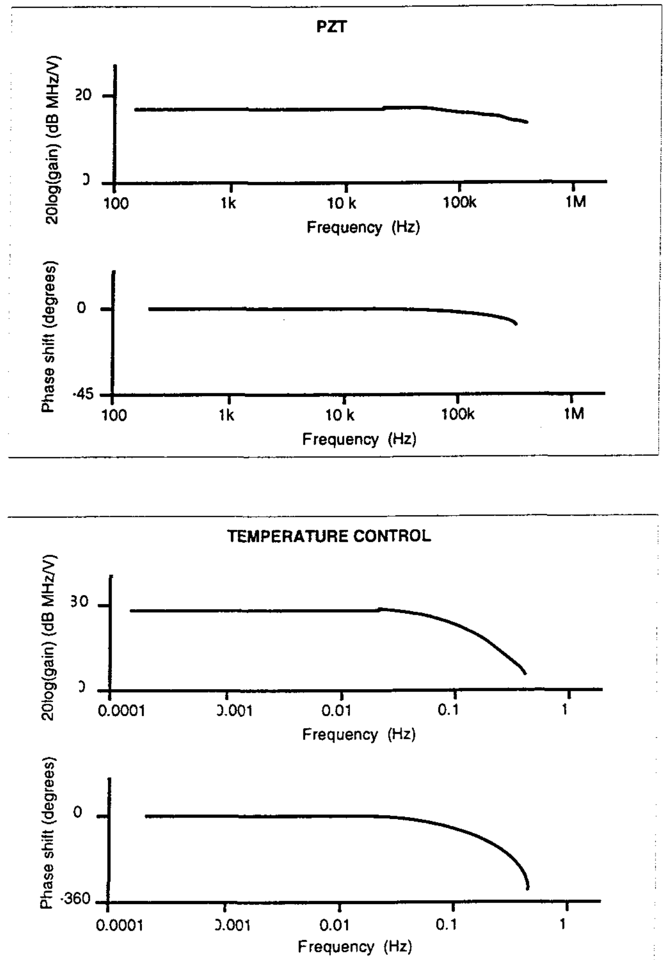


Figure 3.4: Sample Bode plots for frequency-tuning devices in a typical monolithic Nd:YAG laser made by Lightwave Electronics.

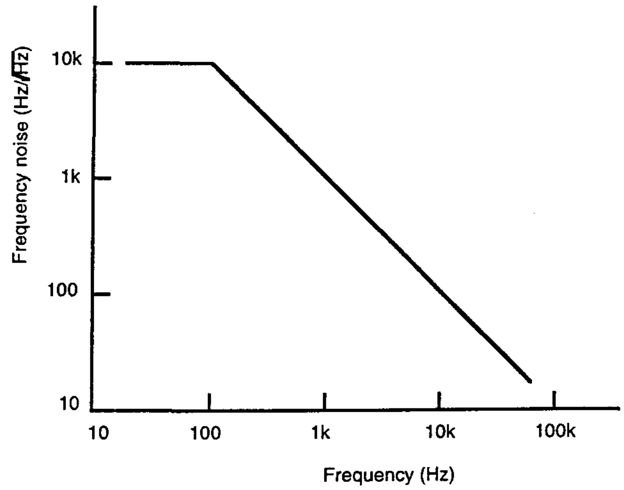


Figure 3.5: Example of upper limit for the frequency noise of a ~ 1 W monolithic Nd:YAG laser. At most Fourier frequencies, the frequency noise is lower than the curve shown. One benefit of adopting a generous upper limit is the simplicity of the spectrum, which makes the design process easier. Moreover, a safety margin is generated, which offers some protection against incomplete knowledge of the noise spectrum at the begining of the design work. If the plot is derived from a some frequency noise measurement, one should keep in mind that the units displayed on spectrum analyzer screens are usually root-mean-square (rms). A commonly used multiplier for converting rms units to peak-to-peak units is 5.

4. Once lock has been acquired, it should be maintained for continous time interavals of at least 10 hours.

3.3 Free Runing Laser Frequency Noise

For the purpose of this design example, the following free-runing frequency variations will be considered:

- **Drift:** over time scales of 100 s and longer, the laser frequency makes excursions within a 1 GHz range.
- **Noise at intermediate frequencies:** The frequency noise spectrum for a monolithic Nd:YAG laser is usually a fairly complicated func-

tion of frequency. Since the error attenuation ratio[16] is usually a fairly smooth function, it has be to chosen such that adequate suppression is provided at the peaks in the noise spectrum. Thus, it is sufficient to have an upper limit for the noise spectrum, which, for design convenience, should be a smooth function with simple frequency dependence. A sample upper limit for the free-runing frequnecy noise spectrum of a Nd:YAG laser is shown in shown in Fig. 3.5. The spectrum of Fig. 3.5 integrates to a total frequency error of 200 kHz rms which is equivalent to 1 MHz peak-to-peak.

- **Fast frequency fluctuations:** occasionaly the laser frequency undergoes step function-like jumps up to 50 kHz in amplitude, over time scales of ~10 ns.

3.4 Environmental Parameters

It will be assumed that the laser is to be used in air, in a normal laboratory environment. Since ambient temperature changes cause changes in laser frequency and reference resonator length, temperature is probably the most important environmental parameter, for this example. Thus, the requirement will be made that the laser maintain the specified frequency stability over a temperature range of 5° C, which exceeds the normal temperature variation in an air-conditioned room.

3.5 In-Band, Out-of-Band Frequency Ranges

Comparing the requirement of Section 3.2 that the laser frequency be stable within 10 kHz peak-to-peak with the free-runing error of Section 3.3, one notes the following:

1. The long term frequency drift has to be attenuated.
2. The frequency noise up to ~100 kHz has to be attenuated.
3. The frequency "jumps" over 10 ns time scales need to be attenuated.

In practice, it is exceedingly difficult to operate a closed loop system at ~ 100 MHz. Thus, the FCS for this problem will most likely be designed to

[16] i. e. $|L|$, according to Eq. 1.2

attenuate frequency noise at (Fourier) frequencies up to $\sim$ 100 kHz, which would take care of Points 1,2 above, while the frequency jumps of Point 3 will have to be addressed by other means, e' g' filtering by passing the entire laser beam through the reference resonator. As a consequence, it is convenient to divide frequencies into two cathegories:

- **In-Band (IB):** frequencies tracking errors exist **and** are attenuated by the FCS. For the present example, frequencies from DC to 100 kHz are IB.
- **Out-of Band (OB):** Other frequencies. For the present example, frequencies over 100 kHz are OB.

It is important to realize that even though OB frequencies are not addressed by the FCS, they are relevant to the design problem:

- OB disturbances can prevent achievement of the tracking requirement, as in the present example.
- OB disturbances can disrupt FCS operation, if they drive the parameter which has to be controlled outside the range of the sensor.
- In the presence of nonlinearities, OB effects can be frequency shifted IB. For example, steady high frequency OB signals could be rectified by slow electronic components. The FCS wold perceive this as a frequency deviation and attempt to correct it. If this effect is large enough to push the laser frequency outside the range of the sensor, the loop will be brocken and the FCS will cease to function.

4 Outline of Control System Design

4.1 Design Approach

4.1.1 Assumptions

The FCS design approach followed here is based on the following assumptions, which reflect reality in the laboratory and/or will considerably simplify the design process:

1. The end product of the design process is a system which complies with pre-determined performance requirements. Performance is not required to be optimal in a rigorous sense.

2. The starting point for the design consists of:

   - A concept for the tracking system
   - A tracking performance requirement
   - Some data on the free-runing behaviour of the variable x
   - Some knowledge of the environment in which the system is required to perform

3. Noise contributed by the compensator, n_G, can be made negligible with respect to reference noise/error n_r by proper comensator design. Noise contributed by the actuators and their drivers is negligible compared to n_G.

4. The design process has to contend with inaccurate performance requirements and incomplete knowledge of the spectrum of the free-runing values of x, the parameter to be controlled.

5. The dynamics of components to be used as sensors and actuators are incompletely known.

6. The behaviour of electronic components outside the linear regime is unknown or difficult to incorporate into a reasonably simple model.

7. Specifying subsystems and components, prototyping and experimental performance assessment are integral to the design process.

4.1.2 Error Budget

The contributions to the error budget are read off Eq. 1.2:

1. Sensor error, n_r consists of a slow variation of the nominal reference value, $n_{r,slow}$ due e. g. to thermal expansion, creep, component aging, drift in the electronics and other slow effects, and a fast fluctuation $n_{r,fast}$ due to fast environmental and internal fluctuations, e. g. acoustic excitation, vibrations, and electronics noise.

2. Compensator noise n_G can usually be made negligibly small by adequate design. Therefore, this term should not contribute to the noise budget.

3. Residual input fluctuation, x_i/L, due to finite loop gain.

In a well designe system output noise/error thus consists of sensor error and residual input error. In practice, it is usually easier to achieve high $|L|$ at low frequencies,[17] and to obtain low n_r at high frequencies.[18] The noise budget is therefore expected to be dominated by sensor error at low frequencies and by residual input error at high frquencies.

4.1.3 Design Guidelines

An immediate consequence of the assumptions made in Section 4.1.1 is that modelling which uses the initial design data will usually be quite inaccurate, and thus will not be able to reveal ahead of time many of the problems encountered in the course of building a working FCS. In view of this fact, and given the approximate nature of the requirements and of the design input data, the design approach discussed in these notes is based on the following guidelines:

1. Starting from the initial concept and from the requirements and the input data, assemble a specification for the sensor and for the actuator(s). Since these are usually long-lead items, it is desirable that their initial specification be adequate for being carried through the design process, with no further iteration.

[17] L has to roll off at high frequencies in order to ensure system stability.

[18] because suppressing drift and other low frequency sensor errors is generally difficult.

2. Assemble an approximate specification for the compensator, using a limited amount of FCS modelling and crude models for the sensor and for the actuator(s). At this stage, the emphasis should be mainly on lock acquisition[19] and on stability, and not on performance. The compensator is all electronics, thus its specification and design are most easily iterated, and does not need to ensure the required performance with its first version.

3. Design and build the compensator and all other necessary electronics. Its input noise should contribute negligibly to the tracking error, compared to reference noise/error.

4. Put together the system and close the loop. Many of the tough problems affecting the design task at hand will show up at this point, thus this stage should be reached as quickly as possible.

5. Diagnose the problems and achieve stable closed loop operation.

6. Run the system with the loop closed. Assess the ease of acquiring lock and measure the free-runing spectrum of x. Assess the long-term behaviour of the system. Determine if performance is adequate.

7. Refine the FCS model using experimental data obtained while runing the system.

8. Iterate the design in order to achieve:
   - smooth lock acquisition
   - reliable long-term operation and
   - specified tracking performance

9. Test-run the system and iterate the design again, if necessary.

The meaning of some of the above guidelines will become apparent when the issues are discussed in detail in the following subsections.

The fact that iterating the design is explicitly made part of the process relieves the need to go into excruciating detail in the initial phase of the work. This makes it possible to move forward fast and settle most details gradually,

[19] see Section 4.3

as the data become available. In this spirit, the various parameters will be devided into two categories:

1. The resolve-it-now (RIN) type includes parameters which need to be chosen at once in order to allow the design process to move along.
2. The worry-about-it-later (WAL) type includes parameters the definition of which depends on yet to be done analysis or on measurements which can be carried out only when a crude version of the system becomes operational.

The distinction between RIN and WAL parameters will be made frequently in what follows.

4.2 Input Data

The input to the design process consists of the following information:

- A concept for the system, including the type of sensor and acuator(s) and the lock acquisition method.
- A requirement on tracking performance which indicates the maximum allowed departure of the controlled variable x_0 from the reference value x_r, i. e.

  $$|x_o(f) - x_r(f)| \leq \mathcal{T}(f) \tag{4.1}$$

  It is useful to find out, by interaction with the customer, how much safety margin has already been included in the requirement, as a function of frequency. Knowing the frequencies where the requirement is "hard" and those where it is really a goal can make a big difference to the designer.

  The tracking requirement should also include a statement on the minimum timespan over which the system is expected to track continously.

- Some knowledge of the free runing value x_i of the parameter of interest. This input is likely t contain a large initial uncertainty, which will be removed when the system is run closed-loop.
- The range for relevant environmental parameters under which the system is required to display the specified tracking performance. It is important to identify the evironmental parameters which will have the heaviest impact on system performance.

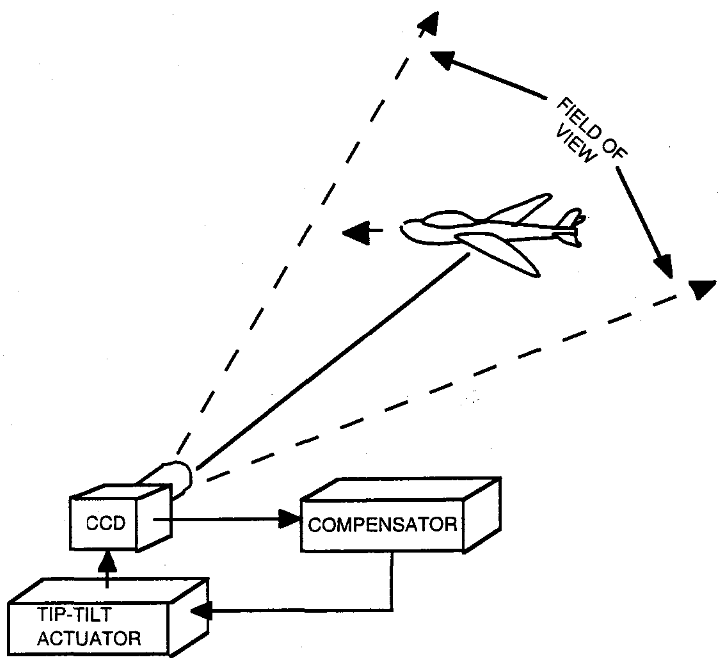


Figure 4.1: Example of a tracking system: a CCD camera is required to keep the image of the aircraft at the center of the CCD. In order to accomplish this, the camera itself in conjunction with appropriate software are providing an error signal when the image is off-center. The error signal is processed by the compensator and fed to a tip-tilt actuator, which corrects the pointing of the camera, to keep the image centered.

4.3 Lock Acquisition

In order for a tracking system to work, the loop shown in Fig. 1.1 needs to be closed, which in turn requires that all elements in the loop function properly. In the example shown in Fig. 4.1, the CCD camera, which is the sensor in the system, can provide an error signal only if the aircraft is within the field of view. If this is not the case, the input to the compensator cannot be processed into a signal which leads to successful tracking. In other words, the variable to be controlled, in this example the angle between the line of sight to the aircraft and the optical axis of the camera, needs to be within what is called the **range of the sensor**, in this case the field of view of the CCD camera. When the aicraft is outside the camera's field of view, the tracking cannot work and the system is **free-runing**. If, on the other hand, the image of the aircraft is on the CCD, the camera can track the plane. This situation is described by saying that the **system is locked**, or **in-lock**. Thus the need arises to ensure a transition, called **lock acquisition**, between the free-runing and the locked states of the system.

The arrangement of Fig. 1.1 does not explicitly contain any provision for lock acquisition. What is needed is some sort of **search algorithm**, which eventually brings the system within sensor range. In some fortunate cases one can rely on the random nature of the free-runing variable x_i to bring the system within sensor range; otherwise, an explicit search mechanism has to be designed into the system.

5 Step-by-Step Design

5.1 Sensor Specification

5.1.1 Sensor Range

Several considerations which help chosing the sensor range R will be discussed below. They all originate from the circumstance that, when the variable x is outside the range of the sensor, the loop is open and the FCS is not operational.

1. **IB requirement**
   When the system performs according to specification, the variable x still shows a residual error with respect to the reference. An upper bound to the IB residual error is calculated by integrating the spectrum of the tracking requirement defined in Eq. 4.1

$$e_{IB} = \left[\int_{IB} \mathcal{T}^2(f)df\right]^{\frac{1}{2}} \tag{5.1}$$

   If the system is to maintain lock, the residual error should not be capable of pulling x outside the sensor reference, i. e. $\mathcal{R} > 2e_{IB}$.

2. **OB requirement**
   Like in the IB case, the range of the sensor has to accomodate the OB error, that is $\mathcal{R} > 2e_{OB}$.

3. **Lock acquisition considerations**
   Before lock is acquired, x is typically outside the range of the sensor. The loop is therefore open, which usually means that the amplifiers making up the compensator are at least partially saturated.[20] When the lock acquisition subsystem brings x within the range of the sensor, the loop is closed, and tracking starts to take effect. However, due to amplifier saturation, attenuation of the tracking error is less than described by Eq. 1.2. The system needs some time to come out of saturation and accomplish the nominal error attenuation. This can happen only if x is within sensor range even when the residual error is larger than under nominal tracking conditions. In other words, the range of the sensor has to be wider than required by the IB condition

[20]similar to the case of a high gain operational amplifier with no feedback loop.

above. A good starting point is to require that sensor range be the larger of $20e_{IB}$, $2e_{OB}$. If during testing it turns out that lock acquisition is difficult, one may have to redesign the compensator electronics for faster recovery from saturation. This is usually simpler, easier and faster than re-specifying and rebuilding the sensor.

Going back to the example of the laser, $e_{IB} = 10$ kHz, and $e_{OB} = 50$ kHz. Thus, the range of the sensor, which is the bandwidth of the reference resonator, should be specified at ~200 kHz.

5.1.2 Sensor Error

According to Eq. 1.2, the tracking error $x_o - x_r$ can be no lower than the sensor error n_r. Therefore, sensor error is bounded by the tracking requirement:

$$|n_r(f)| < |\mathcal{T}(f)| \tag{5.2}$$

for all IB frequencies.

For design purposes, it is convenient to split sensor error into reference error and sensor electronics error, and to further split each of these into slow (e. g. drift, creep, etc.) and fast (arbitrarily defined as $f > 0.1$ Hz) components:

$$n_r = n_{ref;slow} + n_{ref;fast} + n_{el;slow} + n_{el;fast} \tag{5.3}$$

A good design will ensure that the reference error dominate at all IB frequencies:

$$\begin{aligned} |n_{el;slow}| &\gg |n_{ref;slow}| \\ |n_{el;fast}| &\gg |n_{ref;fast}| \end{aligned} \tag{5.4}$$

While designing the sensor for satisfactory noise performance, it is worth keeping in mind the following considerations:

1. The reference is the part of the sensor which is usually long lead. The success of the design may depend to a large extent on identification and careful analysis of the slow and fast contributions to reference error and ensuring they are low enough to comply with Eq. 5.2.

   $n_{ref;slow}$ and $n_{ref;fast}$ are RIN-type[21] parameters.

[21] resolve-it-now

2. In terms of the ultimate performance of the FCS, $n_{ref;slow}$ and $n_{el;slow}$ are equivalent. However, they affect tracking performance by different mechanisms:

   - In the absence of any other error, and under stable FCS operation, x_o tracks the reference; if the reference drifts, x_o drifts with it.
   - In the absence of any other error, x_o will differ from the reference by n_{el}. If n_{el} becomes larger than the sensor half-range, the FCS will loose lock. Thus, even though n_{el} can be considered a WAL-type[22] parameter, one should ensure successful closed-loop operation by selecting electronic components and construction techniques which are most likely to keep n_{el} well below the range of the sensor.

Refering to the example of laser frequency stabilization, Points 1,2 above translate as follows:

1. Since the resonant frequency changes with resonator length as $\Delta\nu/\nu = \Delta l/l$, and the stability requirement is that $\Delta\nu/\nu < 3 \cdot 10^{-11}$, the legth of the reference resonator has to satisfy $\Delta l/l < 3 \cdot 10^{-11}$. If thermal expansion is the only cause of length change, one needs to ensure that $\alpha\Delta t < 3 \cdot 10^{-11}$, where α is the coefficient of thermal expansion. With a resonator made of ULE (ultra-low expansion glass) for which $\alpha \sim 10^{-9}/^{o}$C, resonator temperaturevariations have to be limited to $\Delta t < 0.03^{o}$C, when the ambient temperature may change over a 5^{o}C range. Clearly, both thermal insulation and temperature control of the reference resonator will be necessary.

   A likely contribution to $n_{ref;fast}$ are acoustic and seismic excitations of the resonator spacer. While it is hard to make a general statement regarding the level of these disturbances, it is very likely that, in order to meet the frequency stability requirement that $\Delta\nu/\nu < 3 \cdot 10^{-11}$, one will have to place the reference resonator on a seismic isolator, in vacuum.

2. A typical range for the sensor output is $\sim \pm 1$V/BW. Thus sensor electronics noise should contribute significantly less than ± 0.1 V at the

[22] worry-about-it-later

sensor output in order to allow the system to stay within range once lock has been acquired. Of course, meeting the frequency stability requirement means keeping electronic contributions to the sensor output below ±0.05 V. Likely electronic contributions to sensor error are :

- Temperature-dependent offset voltages at the input of the op-amps
- Pick-up at multiples of the line frequency
- Electronics noise and electromagnetic interference not related to the power line

All these effects are hard to assess quantitatively in advance. In the early phase of the design it is enough to make sure that the various terms listed above do not add up to more than ±0.1 V. This will allow the loop to be closed and FCS tests to be carried out, while the sensor itself is subjected to independent tests aimed at reducing its output error to less than ±0.05 V, consistent with the stability requirement.

5.1.3 Sensor Gain

A more detailed block diagram for the sensor is shown in Fig. 5.1. The sensor head is defined as the part which converts the deviation of the parameter x from the reference x_r into an electrical signal. The purpose of a subsequent stage(s) as shown in Fig. 5.1 is:

- To amplify the signal from the sensor front end in order to make sensor error the dominant error in the system, in particular over compensator noise/error. According to Eq. 1.2, this is another way of saying that high sensor gain $|A(s)|$ reduces the total noise affecting x_o.

- To buffer the sensor front end from the compensator input, e. g. in case of a piezo-electric sensor with ~1 MΩ output impedance and a compensator with a low-noise input stage with ~1 kΩ input impedance.

- To allow the implementation of filtering needed for proper operation of the compensator and of the actuators. For example, it may happen that high frequency noise or pick-up are showing at the output of the sensor

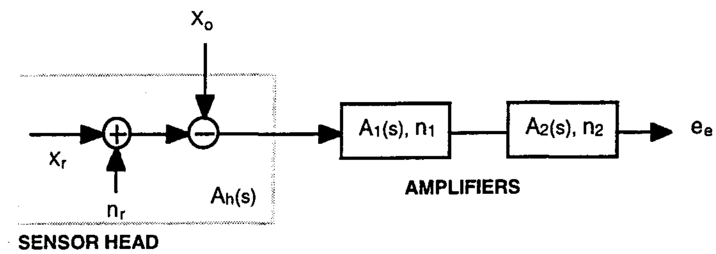


Figure 5.1: Multiple-stage sensor configuration. A distinction is being mde between the sensor front end, where the parameter to be measured is converted into electrical signals, and the amplifiers (two stages are shown), whch may include some filtering. The overall sensor transfer function, measured in units of V/[units of x], is $A(s) = A_h(s)A_1(s)A_2(s)$. Each stage is characterized by a gain and a noise/error contribution.

head, but cannot be tollerated by the slow compensator electronnics.[23] One solution is a filter/amplifier stage after the sensor front end.

The signal at the sensor output is:

$$e_o = A(s)\,(x_o - x_r - n_r) + A_1(s)A_2(s)n_1 + A_2(s)n_2 \tag{5.5}$$

where $A(s) = A_h(s)A_1(s)A_2(s)$. A good sensor design will ensure that the only significant contribution to sensor output error will come from the sensor head itself, that is:

$$\begin{aligned} |A_h(s)n_r| &\gg |n_1| \\ |A_1(s)n_1| &\gg |n_2| \end{aligned} \tag{5.6}$$

For the example of the laser stabilization arrangement, the tracking requirement corresponds to 50 mV at the sensor output. This is high enough to dominate compensator noise almost regardless of compensator design, thus no sensor output amplification is necessary. The only issue than is to design

[23]where it may be rectified and generat an unwanted DC offset in the system.

the sensor front end such that its own error/noise is much lower. e. g. by a factor 10, than the tracking requirement.

5.1.4 Sensor Noninearity

Sensor nonlinearity, defined as x-dependence of the derivative of sensor output with respect to x,[24] is not a crucial issue as long as it is small. Several examples of large nonlinearities, with their possible adverse effects on FCS operation, and tentative remedies are listed below.

- Gain reduction due to nonlineaarity, e. g. saturation, can be so severe that it may prevent the FCS to ensure proper tracking. The remedy is to design enough excess gain into the system to keep the overall gain high enough for proper tracking even when nonlinearity sets in.
- Signal distortion may cause phase shifts which erode the phase margin enough to render the closed-loop system unstable. The remedy is to design sufficient phase margin into the system to accommodate phas shift due to nonlinearity.
- A change of sign of the sensor gain would render the system intrinsically unstable. There is no easy remedy for this situation.

A general, albeit costly way to mitigate the existence of nonlinearities is to define the sensor range as the interval of x where the system is at worst mildly nonlinear. and then design the FCS to operate within this limited range. This entails adding a mechanism for lock acquisition, which is meant to bring the system within its linear range, and requires that there be enough control "strength"[25] available to keep the system tracking, once lock has been acquired. It may happen that keeping the system inside the linear range defines a tougher requirement than the original tracking specification.

The laser stabilization problem presents an example of extreme nonlinearity. As it can be seen from Trace b) in Fig. 3.2, outside the reference resonator bandwidth, the signal changes sign, rendering the closed-loop system unstable[26]. The range of the sensor thus has to be restricted to the resonance bandwidth, and a provision for lock acquisition, as shown in Fig. 3.1, has to be part of the design.

[24] which is the gain, or sensitivity of the sensor

[25] i. e. high enough open-loop gain L.

[26] as it causes positive feedback

5.2 Actuator Specification

Actuators are the part of the FCS which correct tracking errors detected by the sensor and filtered and amplified by the compensator. In order to ensure proper tracking, the actuators[27] have to satisfy two conditions:

1. Since, according to Eq. 1.2 the effect of the FCS is to cancel the free-runing value of the variable of interest, x_i and replace it with the reference value x_r, the correction range of the actuators has to be wide enough to cover the maximum variation of x_i:

$$\text{Actuator range} > 5x_{i;rms} = 5\left[\int_{IB} {x_i}^2(f)df\right]^{\frac{1}{2}} \tag{5.7}$$

2. The actuator has to be "fast" enough to be able to correct for errors at the high end of the IB range.

Returning to the example of the laser stabilization system, the requirement is to cancell frequency fluctuations of up to 1 GHz. The temperature control, which is capable of tuning the laser frequency over ~5 GHz, has plenty of range. However, inspection of the corresponding part of Fig. 3.4 shows that for frequency fluctuation rates[28] above 0.1 Hz temperature control just cannot correct the frequency of the laser anymore. Temperature control is to "slow" an actuator for correcting frequency fluctuations at Fourier frequencies from ~0.1 Hz to ~100 kHz, as required. A "faster" actuator is needed; as the upper box in Fig. 3.4 shows, the PZT has the necessary speed, as its magnitude plot hardly decreases at all up to 100 kHz. On the other hand, the PZT has a correection range of only 100 kHz, and is not able to cover the entire range of expected frequency error. This situation, calling for a slow, wide range actuator to be used in conjunction with a faster, narrower range one, is fairly common. Designing such arrangements for close-to-maximum gain and stable operation is covered in Section 6.2.

5.3 Open-Loop Gain Specification

The open-loop gain has to be specified while keeping in mind two aspects:

[27] which under the definitions used here include the actual correction device and thedriver
[28] i. e. Fourier frequencies

- The need to provide sufficient gain for achieving the prescribed tracking performance.
- The frequency dependence of the open-loop gain has to be consistent with closed-loop stability.

According to Eq. 1.2, if noise is disregarded, $|x_o - x_r| \simeq |x_i/L|$. On the other hand, the design has to ensure that the tracking requirement of Eq. 4.1, $\mathcal{T}(f) \geq |x_o(f) - x_r(f)|$, is satisfied. From these two equations it follows that:

$$|L(f)| \geq \frac{x_i(f)}{\mathcal{T}(f)} \tag{5.8}$$

which defines the minimum magnitute of the open loop gain L, as a function of frequency, necessary in order to satisfy the tracking requirement.

The lower bound for $|L|$, obtained from Eq. 5.3 may not satisfy the Bode criterion (Section 2). If this is the case, the bound on $|L|$ should be deformed to a curve which does satisfy the Bode criterion while still being consistent with Eq. 5.3.

The specification for $|L|$ is more easily used in subsequent FCS design steps if it is expressed in terms of a DC gain. and the number and positions of poles and zeros.

The derivation of a lower bound for the open loop gain L is illustrated using the sample tracking requirement for the stabilized laser, Section 3.2, which asks for the laser frequency to be within 10 kHz of the reference (assumed to be more stable than this requirement), while the low-frequency drift is up to 1 GHz, and the frequency noise at higher Fourier frequencies is bounded by the plot of Fig. 3.5. The somewhat arbitrary assumptions[29] will be made that the open loop gain shall be chosen so that the residual low-frequency and high-frequency errors are equal, each approximately equal to 3kHz, and that the residual noise above 10 Hz be white, e. g. constant as a function of Fourier frequency. The low-frequency gain should then be at least 1GHz/3kHz= 300000 = 110 dB. A total noise level of 3 kHz over a frequency span of 100 kHz corresponds to a constant spectral density of 10 $\mathrm{Hz/Hz^{1/2}}$. Thus, from Fig. 3.5 the gain has to be at least 1000 = 60 dB up to

[29] real life design work often demands that somewhat arbitrary choices be made, in order to keep the process moving. It is important that these choices be documented so that they can be tracked and modified, in case the design is faced with excessive difficulty. This is one of the circumstances which make design iteration necessary.

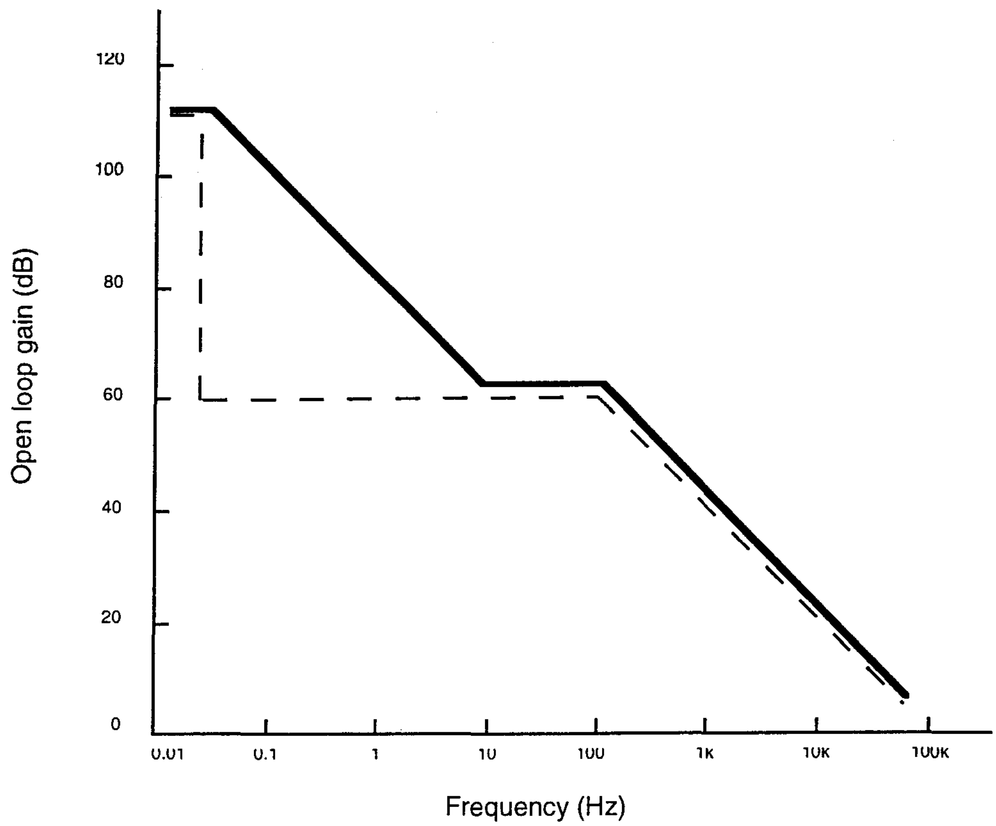


Figure 5.2: Example of open-loop gain specification, for the case of laser frequency stabilization. Dotted line: specification as derived from Eq. 5.3; solid line: modified gain specification, which can be implemented in practice.

100 Hz, with a $1/f$ roll-off at higher frequencies. The result of this process is illustrated in Fig. 5.2. The immediate result of using Eq. 5.3 has a vertical slope at low frequencies. In practice one cannot realize this kind of steep frequency dependence for the gain. Therefore, one has to modify the frequency dependence to a shape which is practicaly feasible, while encompasing the requirement set by Eq. 5.3.

At the frequency where $|L| = 1$ the slope of the gain plot in Fig. 5.2 is 20 dB/decade. This corresponds to a phase lag of 90° (Section 2.2), which, according to the Bode criterion predicts stable closed-loop behaviour. Had this not been the case, one would have had to further distort the gain curve upwards, untill it became compatible with closed-loop stability.

Finally, inspection of Fig. 5.2 and use of the rules listed in Section 2.2 yield the following open-loop gain specification in terms of DC gain, poles and zeros:

- DC gain: 110 dB
- Poles: 0.03 Hz (single pole) and 100 Hz (single pole)
- Zeros: 10 Hz (single zero)

5.4 Compensator Specification

From a pure control system point of view, the most important characteristic of the compensator is its transfer function $G(s)$. G is obtained from the definition of the open loop gain L given in Section 1:

$$G(s) = \frac{L(s)}{A(s)\sum_k H_k(s)B_k(s)} \tag{5.9}$$

where A, B_k and H_k are the transfer functions of the sensor, the actuators and the actuator drivers, respectively.

In order to facilitate electronics design and construction, as well as troubleshooting for proper close-loop operation, the compensator specification should include:

1. The transfer function as calculated with Eq. 5.9, expressed as a DC gain and a list of poles and zeros.
2. A normal signal input which should be characterized by:

- Input impedance, selected such that the sensor output is not overloaded.
- Input refered offset, which should be less than sensor error at the output of the sensor.
- Input refered noise level, which should be less than noise at the output oof the sensor.
- Type of connector

3. A test input which allows injection of a test signal under operating conditions. This input can be implemented as a summing point at the input of the compensator, or at the input of any of the compensator amplifier stages.

4. A normal output which should be characterized by:

   - Output impedance, which should normally be much lower than the input impedance of the actuator driver(s) which follow downstream.
   - Output range, selected to match the range of the actuator(s).
   - Maximum current drive capability, determined by the input impedance of the following actuator driver(s) and the output range.
   - Type of connector.

5. A monitor for the output of each amplifier stage, to assist the testing process.

6. A front-panel switch for changing the sign of the signal. This switch will help ensure that negative feedback is established once the loop is closed.

7. A front-panel gain adjustment knob, for changing the compensator gain[30] by two decades above and two decades below the nominal gain. This provision is necessary because the initial design is not accurate enough to predict the exact value of the gain "as built," and also because the necessary gain is somewhat uncertain, given the uncertainty in the initial design data.

[30]and thus the open-lop gain L

8. Input and output protection. This provision is necessary because close-loop operation is likely to engender system oscillation at one time or another, in particular during the trouble-shooting phase. Conditions may then occur which could damage the compensator or the actuator(s).[31]

5.5 Lock Acquisition

If the system is initially outside the range of the sensor, the loop cannot be closed, and the FCS is unable to perform its tracking function. While x_i is outside the range of the sensor, it is likely that the amplifying stages of the sensor and compensator will be saturated, thus driving the actuator toward the edge of its range. There are two distinct lock acquisition conditions:

1. **Spontaneous lock acquisition**
   This occurs if the natural variation of x_i is large enough to bring the system within sensor range occasionally. If the variation of x_i is slow enough to give the amplifiers time to come out of saturation while the system dwells within the range of the sensor, lock acquisition will "happen" without any special addition to the FCS. If spontaneous locking appears to be a possibility, the electronics should be designed with quick desaturation in mind.

2. **Induced lock acquisition**
   This is a situation where the system would "never" enter sensor range, if explicit provisions are not added to the FCS. The remainder of this subsection contains a short discussion of induced lock acquisition.

Locking can be induced in many different ways. Fig. 5.3 presents one example which illustrates what's involved. The main components of the arrangement in Fig. 5.3 are:

- An "out-of-lock" or "out-of-range" detector which continuously analyzes the output of the FCS sensor for anomalous behaviour, which indicates that the variable x is outside the sensor range, which means that the system is out of lock.

[31]For example, if the actuator is a piezo-electric device, oscillation may result in voltages much higher than the maximum rating of both the actuator and the compensator. A voltage clamp would provide adequate protection in this case.

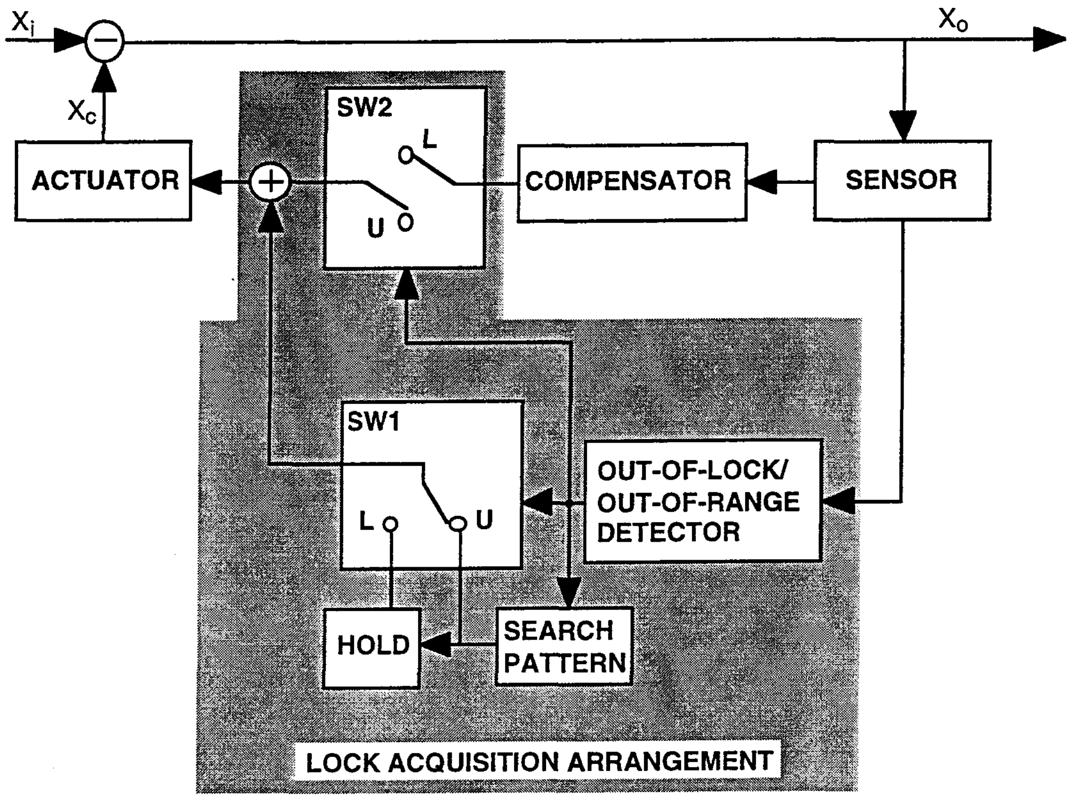


Figure 5.3: Concept of induced lock acquisition arrangement, shown in the shaded area of the picture. **SW1,2**: switches; **U**: system in unlocked (out-of-range) state; **L**: system locked.

- A set of switches.
- A search pattern generator.

Lock acquisition proceeds as folows:

1. Absence or loss of lock is detected by the out-of-lock detector.
2. The locking sequence is initiated by the out-of-lock detector setting both switches to the "U" position. SW2 thus disconnects the actuator from the compensator, which is saturated anyway, while SW1 connects the search pattern generator to the actuator input.
3. The search pattern generator is triggered, which causes the actuator to scan the otherwise free-runing variable x_i.
4. When x_i moves within sensor range, the FCS amplifiers in the sensor and the compensator come out of saturation.
5. The out-of-lock detector senses that the system is now within range, sets both switches to the "L" position, and commands the hold circuit to continously apply the last value of the search pattern generator to the actuator input. The FCS loop is thus closed and the system operates as designed, while the lock acquisition arrangement is basically out of the loop. Lock has been acquired.

Example 1: laser stabilization
For the stabilized laser shown in Fig. 3.1, the locked state corresponds to the laser field resonating with the reference resonator. In this state, the maximum amount of light is transmitted through the resonator. The out-of-lock/out-of-range detector can thus be implemented as a transmitted power measurement combined with a threshold detector; an in-lock state is declared when the transmitted light exceeds a preset threshold. The search pattern could be a succession of step increases in laser crystal temperature, effected by the thermoelectric element. As a result, the laser frequency would change in steps, untill it comes close to a reference cavity resonance. At that point, searching is stopped and the loop closed.

Example 2: aircraft tracking
For the tracking camera shown in Fig. 4.1, the locked state corresponds to

the image of the aircraft being held on the CCD array. The out-of-lock/out-of-range detector can thus be implemented as an algorithm which recognizes the shape or the motion pattern of the plane. The search pattern could be an outward spiraling motion of the camera, executed using the tip-tilt actuator. When the image of the plane is detected, the search is stopped and the loop is closed, so that the camera can track.

5.6 Making the System Work

5.6.1 Achieving Closed-Loop Operation

Of all the steps in the design and testing of a FCS, this is without doubt the most challenging. Indeed, in most cases proper closed-loop operation will not be established when a new system is being put together for the first time, even though the individual parts of the system, i. e. the sensor, the compensator and the actuator have been tested and are in good working order. The presence of tracking or the lack thereof can be determined by observing the signals e_e and e_c at the sensor and actuator output monitors shown in Fig. 5.4. When the system tracks, both sensor and the correction signal change with time. For sytems with very high open-loop gain, the sensor output may appear "frozen" during closed-loop operation, due to the large degree of suppression of the free-runing variable x_i, according to Eq. 1.2. On the other hand, if any of $e_{1,2,3,c}$ is either zero or the maximum value supported at the corresponding output, tracking is not occuring.

Since the various operating parameters are vastly different in the closed-loop and open-loop regimes, testing the system open-loop in order to find why the loop can't be successfully closed is going to be only marginally useful. The troubleshooting approach discussed in what follows is based on the observation that, when the system is brought within sensor range, all the components of the system are in a linear regime for a short time, untill linearity is lost because the loop malfunctions. The specifics of the transition from linearity to severe nonlinearity[32] hold the clues needed to make the loop work.

Troubleshooting can thus be conducted along the following three-step procedure:

1. Bring the system within sensor range by using the input normally used

[32] e. g. amplifier outputs getting stuck at one of the power supply voltages

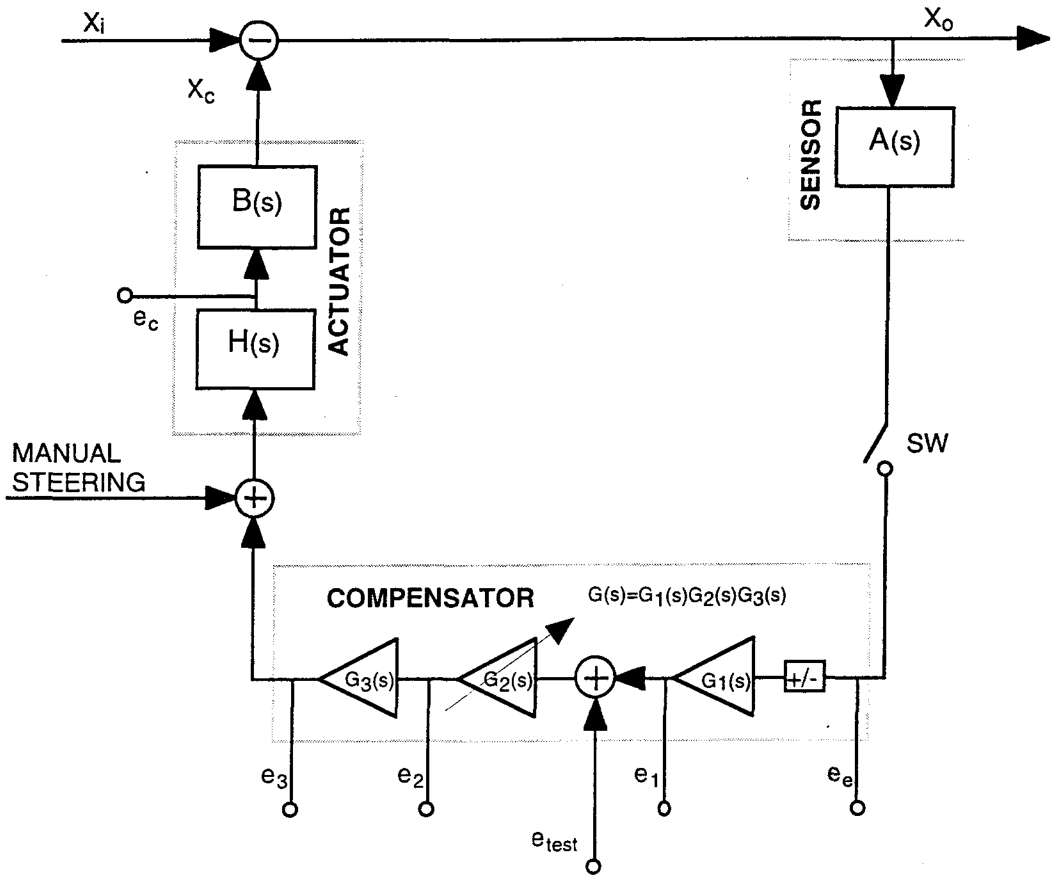


Figure 5.4: Modified version of Fig. 1.1, used in the discussion on troubleshooting the FCS and testing its performance. **SW**: switch for selecting open- or closed-loop operation; +/-: sign switch for ensuring negative feedback; e_e: signal at sensor output monitor; $e_{1,2,3}$: signals at output monitors for various compensator amplifier stages; e_c: correction signal monitor; e_{test}: signal at test input. The second stage of the compensator is shown as having variable gain, which is a provision for adjusting the open-loop gain. The input labelled "MANUAL STEERING," which under normal operation is the input for the lock acquisition arrangement (see Fig. 5.3), is used to manually bring the system wihtin the range of the sensor in the evaluation phase.

for lock acquisition, as shown in Fig. 5.4.

2. Record the linear-to-nonlinear transition at one or several of the output monitors shown in Fig. 5.4.

3. Analyze the above time record.

The tests will require an oscilloscope and a sufficiently fast transient recorder.

The most likely causes of closed-loop failure are:

1. Closed-loop instability

2. Insufficient open loop gain

3. Insufficient actuator range

A short discussion of each of these follows.

Closed Lopp Instability

If the loop is closed and the system is steered manually, observing the sensor output e_e will indicate when the system comes within range; the sensor output shows variation with time, within the known range for sensor output voltages. If lock is not acquired at this point because of closed loop instability, the sensor output will show an oscillation which builds up to a point where some amplifier in the system saturates, as shown in the example of Fig. 5.5. A first quick diagnostic consists of changing the sign of the open-loop gain by using the sign switch built into the compensator. If the cause of instability was positive feedback, it should now be possible to close the loop. If the loop appears to be working, changing sign should cause instability as in Fig. 5.5.

If changing the sign does not help, there is a genuine closed-loop instability. In other words, the phase lag fot $|L| = 1$ is more than 180°. To further investigate this possibility, one should attempt to set the sign correctly. Many times, this is not practically possible, and all the tests will have to be carried out for both signs. In some cases, the frequency of the oscillation is visibly different for the two signs settings. The correct sign then probably corresponds to the switch setting where the oscillation frequency is close to the nominal unity gain frequency of the FCS.

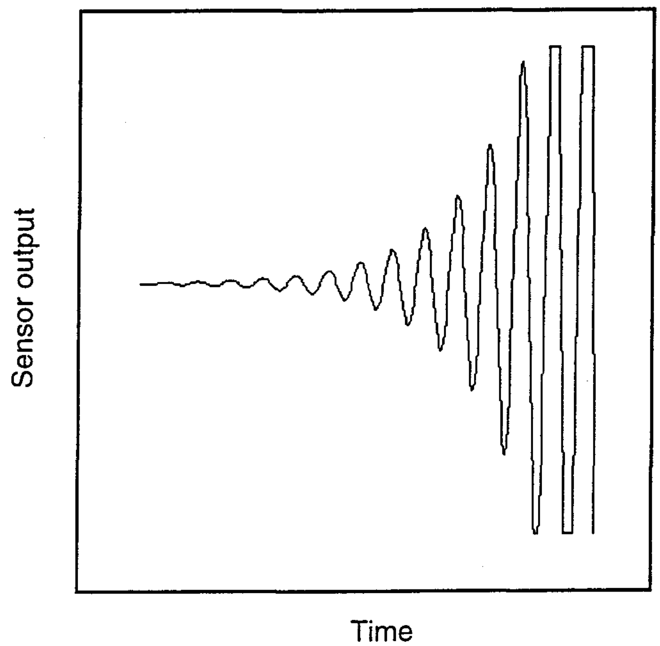


Figure 5.5: Buld-up of oscillation as a result of closed loop instability, as seen at the sensor output. Saturation is visible at the right hand side of the trace. Since the loop is closed, it is not obvious from the figure where in the system saturation occurs. To locate the first amplifier to saturate, one has to look at the output monitors downstream of the sensor.

Assuming that the feedback sign is correctly set, the next thing to look for is an incorrect overall gain value. The example shown in Fig. 5.6 illustrates one such situation. Thephase lag at 10 kHz is $\sim 35^o$ less than 180^o. Thus, if the unity gain is at 10 kHz, as in the middle trace, the closed loop should be stable. For both the upper and lower traces, however, unity gain occurs at frequencies where the phase lag is higher than 180^o, which will cause close-loop instability. sometimes the specific design for $L(f)$ is such that instability can occur only when the gain is too high. In any event, if improper gain is the cause for instability, both the dignostic and the remedy consist of changing the overall gain by using the manual gain control included in the design of the compensator.

If both changing the sign and changing the overall gain fail to eliminate the istability, the designer is likely to have one of the following problems:

1. Unaccounted for dynamics in the system causes additional phase lag.
2. Compensator as built induces more phase lag at unity gain than allowed for in the design.

In either case one will have to perform a piece-by-piece measurement of the open-loop gain L, followed by appropriate redesign.

Insufficient Open Loop Gain

If the system is free of instability, lock acquisition will occur oance the free-runing variable x_i has been steered within sensor range. In this mode, the sensor output will be $e_e = A(x_o - x_r)$, where A is the sensor gain. Disregarding noise, according to Eq. 1.2 this can be written:

$$e_e = \frac{x_i}{L} \tag{5.10}$$

If L is not high enough, e_e maight exceed the range of the sensor for large values of x_i. Then tracking is lost, and the sensor output may become a flat line, as in the example shown in Fig. 5.7. The immediate thing to try is to increase the overall loop gain by using the gain control of the compensator. This remedy is limited by:

- The limited range of any gain adjustment.

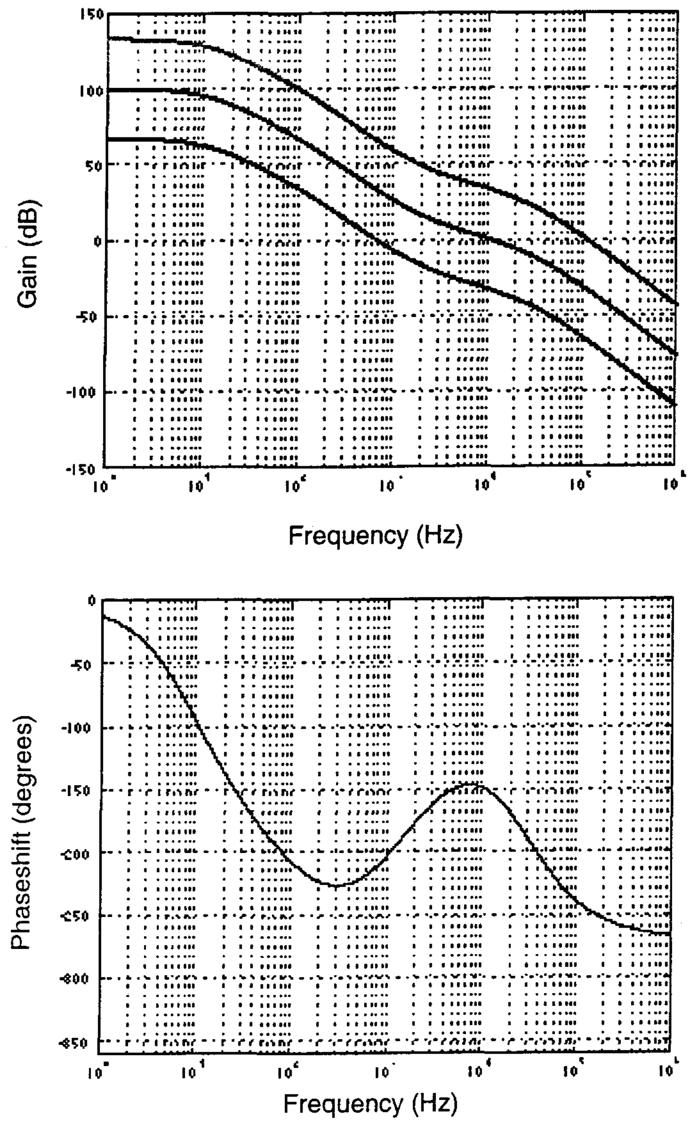


Figure 5.6: Bode diagram for a system which goes unstable when the gain is either increased or decreased from an intermediate value. The magnitude curve in the middle crosses 0 dB (i. e. unity gain) at a frequency where the phase shift is well under 180°, which makes the system stable, according to the Bode criterion. The upper and lower magnitude curves cross 0 dB at frequencies where the phase shift is well in excess of 180°, which makes the system unstable.

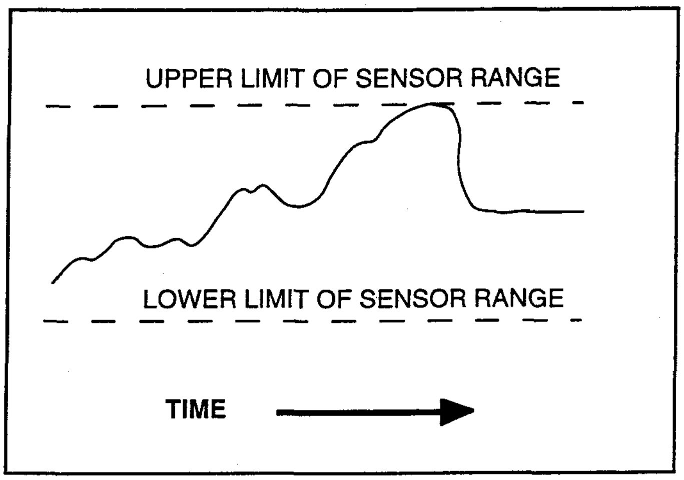


Figure 5.7: Time record of sensor output for a situation in which lock is lost due to insufficient open-loop gain.

- Increasing the overall gain moves the unity gain frequency to higher frequencies, so that the phase margin becomes lower and lower, untill the system is rendered unstable.

If this simple remedy does not help, one will have to measure $X_i(f)$, the spectrum of x_i, while the system is tracking, and increase the gain selectively around frequencies where $X_i(f)$ peaks. Methods for increasing the gain are discussed in Section 6.

Insufficient Actuator Range

If tracking for extended time intervals fails even though the system is stable and there is enough gain to keep the signal within the range of the sensor, the prime suspect is a limitation in the range of the sensor. Indeed, according to Eq. 1.2 the correction signal x_c is approximately equal to the free-runing variable x_i, when L is high and noise is disregarded. Therefore, with reference toig. 5.4, the actuator driver output is:

$$e_c = \frac{x_c}{B} \tag{5.11}$$

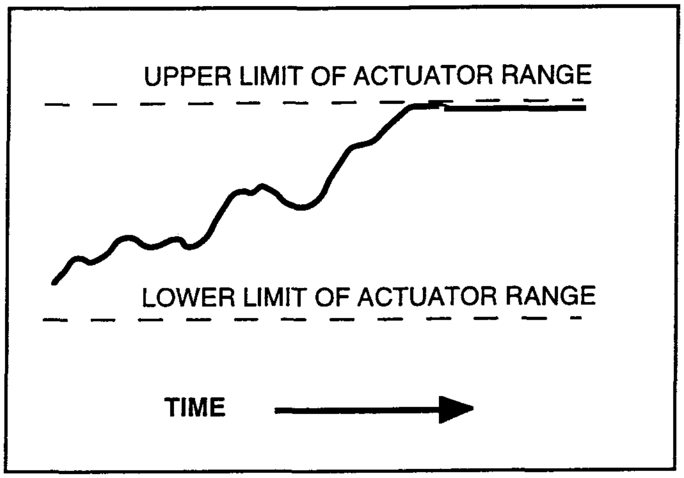


Figure 5.8: Time record of actuator driver output e_c for a situation where tracking fails as a result of insufficient actuator range.

If the free-runing variable becomes too large, the output of the actuator driver will fail to increase enough for proper tracking, i. e. its final stage will saturate, and tracking will be lost. An example of saturation at the driver output is shown in Fig. 5.8. In many cases, e_c will retain its saturated value even after x_i leaves the range of the sensor because of tracking failure,[33] because of the high gain in the system which amplifies small DC offsets beyond the output range of the actuator driver.

5.6.2 Measuring the Open-Loop Gain

Since the open-loop gain L is crucial in achieveing the specified system performance, it is important to determine what its actual magnitude and phase versus frequency are and compare them with the design values. With reference to Fig. 5.4, ideally one would open the switch **SW**, inject a sine-wave at the input of $\mathbf{G}_1$, measure the corresponding sensor output and take their ratio. Varying the frequency of the sine-wave over the band of interest would yield the open-loop frequency response. This direct approach is impractical in many cases, mainly because:

[33]leading to the sensor flat-linig at zero

- With the loop open (i. e. with the switch **SW** open), the free-runing variable may be outside the range of the sensor, and thus no useful signal would be present at the sensor output.

- The combined gain of the compensator and actuator driver may be large enough to force the actuator driver output into quasi-permanent saturation, whereas no change will be enforced on x_i and thus no signal would be present at the sensor output.

Neither of the two points above applies if the loop is closed, which is possible only when there is a useful signal at the sensor output and if the actuator driver is not saturated.

An open-loop gain measurement with the loop closed is described in what follows.

A sine-wave e_{test} with frequency f is injected at the test input of $\mathbf{G}_1$, the corresponding output $e_2(f)$ is measured and their ratio is calculated as a function of frequency. When the loop is open:

$$R_{\text{open}} = \left[\frac{e_2}{e_{\text{test}}}\right]_{\text{open-loop}} = G_2 \tag{5.12}$$

The calculation of R_{closed} for closed-lop operation is silmilar to the derivation of Eq. 1.1, with the result:

$$R_{\text{closed}} = \left[\frac{e_2}{e_{\text{test}}}\right]_{\text{closed-loop}} = \frac{G_2}{1+L} + \frac{AG_1G_2}{1+L} \cdot \frac{x_i}{e_{\text{test}}} \sim \frac{G_2}{1+L} \tag{5.13}$$

where $L = ABG_1G_2G_3H$ and the x_i term was neglected assuming that e_{test} is large. L is obtained by comparing Eqs. 5.12,5.13:

$$L = \frac{R_{\text{open}}}{R_{\text{closed}}} - 1 \tag{5.14}$$

5.6.3 Measuring the Free-Runing Variable $\mathbf{x_i}$

At the beginig of the design process, the spectrum or other quantitative descriptions of the free-runing variable are usually known with substantial uncertainty. Since the tracking performance of the system relies on x_i suppression by the loop gain, which is set by the designer to a certain magnitude, it is important to make a reliable assessment of the size of x_i.

According to Eq. 1.3, for high values of L the correction signal closely follows the free-runing variable, $x_c \sim (x_i - x_r)$, up to noise terms. Since $x_c = Be_c$ (See Fig. 5.4), the free-runing variable x_i can be calculated from:

$$x_i = x_r + Be_c \tag{5.15}$$

The input to the actuator, e_c, also called comand signal, is measured directly with the **loop closed**. Tis measurement should be repeated under a variety of conditions, in order to determine the upper limit $X_i(f)$ of the spectrum of x_i. $X_i(f)$ will henceforth serve as the worst case against which the tracking performance will be evaluated.

5.6.4 Evaluating Tracking Performance

As indicated in Section 4.2, the tracking requirement consists of two components:

- Maximum departure from reference.
- Tacking robustness, measured by the minimum amount of time the system is expected to track continously.

Maximum departure from reference refers to the quality of tracking. In some rare cases it is easy to see if tracking is appropriate. For the example where a camera is required to track an aircraft, the system performs adequately if the image of the flying plane is kept within the field of view of the camera, which is easily ascertained. In most cases however, testing for adequate tracking tends to be a more laborious process. Tracking error can be evaluated by inspection of Fig. 1.1:.

$$\delta(f) = |x_o(f) - x_r(f)| = \left|\frac{e_e(f)}{A(f)} + n_r\right| \leq \frac{e_e(f)}{A(f)} + n_r \tag{5.16}$$

$\delta(f)$ can be compared directly with the tracking requirement $\mathcal{T}(f)$, if the latter is given as a tracking error spectrum, as in Eq 4.1. If the tracking requirement is given as a root-mean-square value, it should be compared to the *rms* integral:

$$\Delta = \left[\int_{\text{In-band}} \delta^2(f) df \right]^{\frac{1}{2}} \tag{5.17}$$

If the sensor error n_r is **guaranteed** to be small enough, $\delta(f) \sim e(f)/A(f)$. In this case, since $A(f)$ is known, tracking performance can be assessed by simply measuring $e_e(f)$ several times, under different conditions with the loop closed, using these measurements to determine the upper limit $E_e(f)$ on $e_e(f)$, then using Eq. 5.16 to evaluate $\delta(f)$. It should be stressed however that unless a direct mesurement of $n_r(f)$ is carried out, this kind of "inside-the-loop" tracking performance evaluation is dangerously unreliable, as it would have to build on the **belief** that sensor error is small. While every effort has possibly been made to design a sensor with low error, the purpose of testing is to verify that the result is actually as good as desired. A more reliable test protocol is strongly advocated. While it is not possible to prescribe a test procedure that is satisfactory in all cases, because of the enormous variety of tracking applications, one method which often works is to use a sensor external to the feedback loop, as shown in Fig. 5.9. This method works if :

$$|n_{r;ext}(f) \ll n_r(f) \tag{5.18}$$

In this case. tracking quality is estimated by measuring e_{test}. For example, for a system designed to stabilize the temperature inside an enclosure, the external sensor would be a thermometer which is more accurate than the temperature sensor used in the loop.

Tracking robustness is easy to measure by letting the system acquire lock and recording the actuator driver output e_c. An obvious sign of robustness is uninterrupted tracking for at least as long as the specification requires it, and preferably for times 50%-100% longer. The record of e_c should be compared with the output range of the actuator driver.[34] In a well designed system, there is range to spare; one ad-hoc rule would be too make sure that e_c is within less than one third of the available range, most of the time.

[34] which is ideally just a little narrower than the corresponding actuator range.

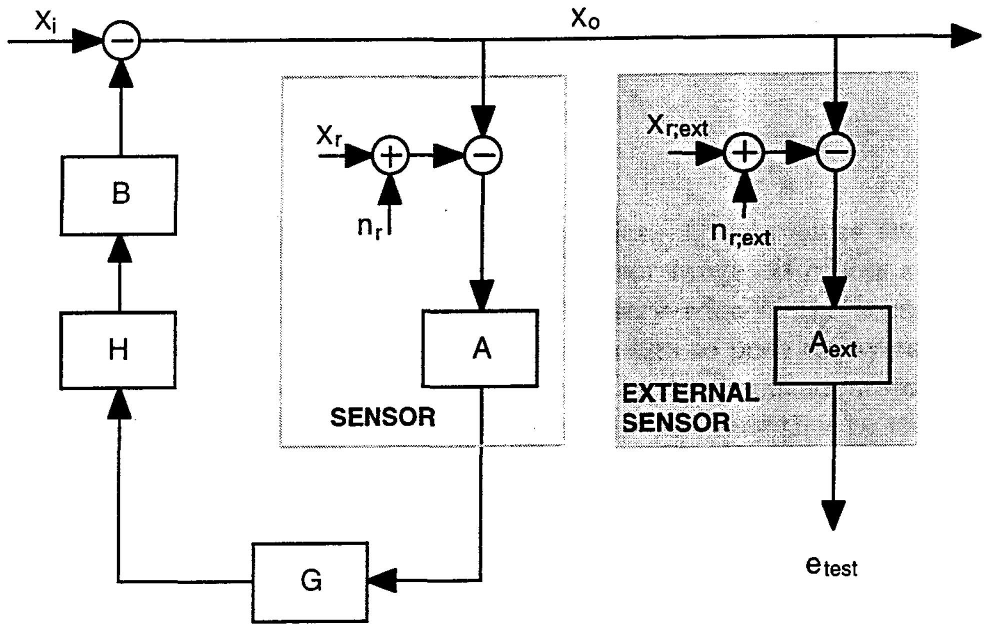


Figure 5.9: Use of an external sensor (shaded) for testing tracking performance.

5.7 Lock Acquisition Efficiency

Testing the lock acquisition arrangement consists of tacking the system out of lock and letting it re-acquire a number of times. If the system re-acquires on its own within a reasonably short timespan,[35] every time, the design is successful. In order to make the test meaningful, it is desirable that, every time the system is taken out of lock, a different initial condition be established, so that the entire range of initial conditions likely to be encountered in practice is covered. For example, in the case of the frequency-stabilized laser, every time lock is interrupted one would manually tune the frequency of the laser to a different point in the interval between two resonances of the reference resonator. Frequently encountered causes for difficulty in lock acquisition are:

1. Unfavorable initial conditions. For example, in the case of the camera trying to acquire an airplane, the target search pattern may be so fast, that when the presence of the aircraft in the field of view is detected, the FCS does not have enough power, i. e. gain, to stop the search pattern before the target is lost. Possible remedies are:

[35] "reasonable" needs to be defined for each concrete case.

- Selection of a slower search pattern.
- Increasing the open loop gain L.

2. The presence of offsets in the electronics. Offsets cause the system to lock away from the center of the sensor range. If offsets are large, the system may tend to track at the very edge of the sensor range and even small perturbations may be enough to push the system "over the edge" and cause it to loose lock. Electronic offsets are often the result of changes in operating temperature and may thus be the reason for intermitent lock acquisition difficulties. The remedy is to carefully analyze the specifications of the amplifiers used in the design of the electronics and select those with sufficiently low temperature dependent offsets.

5.8 Refining the system

After carrying out the tests described in Sections 5.6,5.7, the information needed to make the system perform to specification is available. The areas most likely to need improvement arelisted below.

1. The open-loop gain L may require:
   - An increase in magnitude at certain frequencies, for adequate tracking and easier lock acquisition. Methods for increasing L are discussed in Section 6.
   - An increase in phase margin, if "ringing" around the unity gain frequency is observed, and if the measured open-loop gain shows less than 30^o phase margin.
2. If the range covered by the free-runing variable x_i is almost as wide or wider than the actuator range, the actuator selection will have to be revisited. It is desirable that the actuator range be at least e. g. three times the range of x_i.
3. In order to ensure smooth and efficient lock acquisition,[36] it may be necessary to:

[36] as discussed earlier, lock acquisition is an issue only if the range of x_i exceeds the range of the sensor

- Modify the search pattern.
- Redesign the electronics in order to reduce temperature-dependent DC offsets.
- Increase the open-loop gain at certian frequency ranges.

Once all the changes are implemented, the system needs to be put again through the tests discussed in Sections 5.6,5.7.

6 Increasing Loop Gain

6.1 The Need for High Bandwidth

6.2 Parallel Control Paths

6.2.1 The Need for Parallel Control Paths

6.2.2 Parallel Control Paths: System Stability

6.2.3 Bypass Configuration

6.2.4 Simultaneous Use of Several Actuators

7 Case Study: Laser Frequency Stabilization

by Serge Dubovitsky