The Miller Effect and Pole Splitting

1.0 Introduction

Engineers frequently design systems to be dominated by a single pole. Aside from being easily analyzed (certainly an extremely attractive property in its own right), such systems also have the highly valuable attribute of being able to tolerate large amounts of negative feedback without stability problems.[1] While it is impossible in practice to build a system that is truly single pole, it is not hard to approximate single pole behavior over a broad enough frequency range to be useful. Consider, for example, the Miller effect: it can increase dramatically the time constant associated with a capacitance that feeds back around an inverting gain stage. Usually, this effect is considered undesirable (because it degrades bandwidth), and we therefore often expend a great deal of design effort to avoid it (through cascoding, for example). However, the Miller effect can also be useful; it can be exploited to make a system's open-loop transfer function approximate simple first-order dynamics over a wide range by creating a dominant pole.

To be confident that the pole created is indeed dominant, though, we must have some way of determining or estimating the location of the next pole. As with open-circuit time constants, we will avoid traditional, rigorous paths to an exact answer. Instead, we'll content ourselves with approximations that convey an intuitive appreciation of the dynamics of a particular two-pole system that recurs with surprising frequency in analog circuit design. It is this intuition that we will emphasize in what follows.

Among the more important insights is that the Miller effect generally makes one pole more dominant while simultaneously making the other one less so. That is, as one pole moves down in frequency, the other moves up in frequency. When this contrary motion (known as *pole splitting*) is an intended consequence, the Miller effect is often renamed Miller compensation. It is a powerful way to force the resulting transfer function to appear first-order over an exceptionally large frequency range[2].

Even if one is uninterested in shaping the frequency response of an amplifier, an understanding of pole splitting is essential to extending to the second order many important insights developed during our study of first-order systems.

2.0 Two-pole Amplifier

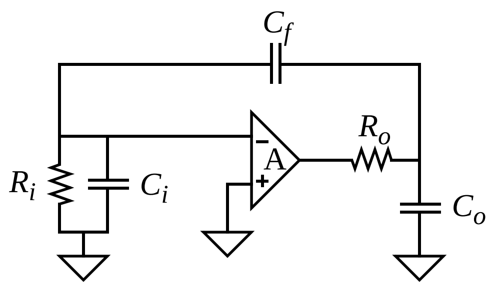
Figure 1 is a model of the system we'll consider. It is a quite general representation of any unilateral two-port linear amplifier with capacitive feedback. For example, this model

1. You may recall that a truly first-order system is unconditionally stable for any amount of purely scalar negative feedback.

2. Pole-splitting occurs nearly any time you couple two dynamic networks together, however, so it isn't unique to Miller compensation.

applies to a common-source amplifier (where the open-circuit gain A corresponds, say, to the product $g_m R_L$), or the second stage of many general-purpose op-amps. Here, R_i and C_i represent the total shunt impedance seen between the input and ground (including contributions from the input source that drives the amplifier, as well as the input impedance of the gain stage itself), R_o models the intrinsic output resistance of the gain stage, C_o represents the total capacitive load on the output, and C_f is the "pole-splitting" compensation capacitor. The gain stage is assumed to have an open-circuit DC gain of $-A$, as seen in the figure.

FIGURE 1. Model of two-pole amplifier (using Thévenin representation of gain stage)



Before proceeding further, let's make a few observations that may help us derive some insight later. First, let's verify that the system is of the second order, despite the presence of three capacitors. Because the capacitors form a loop, KVL tells us that the voltage across any two automatically determines that across the third. Hence, there are actually only two degrees of freedom and the transfer function is therefore indeed of the second order, meaning that the denominator polynomial is a quadratic. Furthermore, such a second-order denominator polynomial can be treated as the product of two first-order factors:

$$(\tau_1 s + 1)(\tau_2 s + 1) = \tau_1 \tau_2 s^2 + (\tau_1 + \tau_2) s + 1 . \quad \text{(EQ 1)}$$

First note that the coefficient of the s^2 term is the product of the pole time constants (but generally *not* of the open-circuit time constants). The coefficient of the s term is simply the sum of the pole time constants. In turn, that sum is precisely equal to the sum of open-circuit time constants, and we have seen that such time constants are readily found, almost by inspection.

To use these observations, we could proceed as follows: First compute an estimate for the dominant pole's time constant through open-circuit time constants, then divide this estimate into the leading coefficient of the denominator polynomial to estimate the other time constant. Of course, implementation of this method requires knowledge of the actual leading coefficient's value. For that, we need an actual transfer function. Fortunately, it isn't terribly hard to derive for this second-order case. In fact, you'll likely be able to write it down with little difficulty after very little practice. Until then, though, you'll have to use standard formal methods (e.g., node equations) to find that the actual transfer function is:

. (EQ 2)

For the purposes of the present problem, we continue to focus only on the denominator. Upon close study of its leading coefficient, a pattern emerges: It consists of the product of the two resistances, as well as the sum of products of the capacitances, taken two at a time. Thus it is actually rather simple to write down *by inspection* an expression for the leading coefficient of the denominator polynomial once the circuit has been reduced to the form shown in Figure 1.

To demonstrate the utility of these observations, first consider the case where the two poles are widely separated. Then one pole necessarily dominates, with a value that is therefore well approximated by open-circuit time constants. After some rearrangement of terms, this pole has an approximate value that may be expressed as:

$$\tau_1 \approx \sum_i \tau_{jo} = R_i(C_i + C_f) + R_o(C_o + C_f) + AR_iC_f \ . \qquad \text{(EQ 3)}$$

Written in this fashion, one can immediately see that the dominant pole’s frequency is generally the result of combined contributions from the input and output ports. It is thus misleading, and potentially dangerous, to speak of an “input” or “output” pole, or to speak of associating poles with nodes, although such language is frequently encountered in the literature. Feel free to use such terminology, but only with the awareness that there can be many cases where it doesn’t make much sense to do so. An example of where it might make sense is when AR_iC_f is large enough to swamp out the other terms. The estimate of the dominant time constant then becomes

$$\tau_1 \approx AR_iC_f. \qquad \text{(EQ 4)}$$

Because the time constant depends only on the input resistance in this instance, it would be reasonable to speak of an input pole here.

Having estimated the dominant pole’s time constant, simple division into the leading coefficient gives us the other:

$$\tau_2 \approx \frac{R_iR_o[C_oC_i + C_oC_f + C_iC_f]}{R_i(C_i + C_f) + R_o(C_o + C_f) + AR_iC_f}. \qquad \text{(EQ 5)}$$

Before proceeding further, note a few general trends: As the product AC_f increases, one pole (the “input” pole) moves down in frequency (i.e., becomes more dominant), while the other pole (the “output” pole?) moves upward in frequency. That is, the poles move oppositely, leading to the term *pole splitting* to describe the action. In the limit of very large AC_f (relative to the other time constants in Eqn. 3 and in the denominator of Eqn. 5) the pole locations become

(EQ 6)

and

$$\tau_2 \approx \frac{R_o[C_oC_i + C_oC_f + C_iC_f]}{AC_f}. \qquad \text{(EQ 7)}$$

Note that the product of the two estimated time constants is independent of A, meaning that a doubling of one time constant through a gain increase is accompanied by a halving of the other. This conclusion is not an artifact of our assumptions. Inspection of the original, exact transfer function reveals that the leading coefficient is independent of A, so that the product of the pole frequencies is *always* independent of A.

Further note that one estimated time constant depends only on the input resistance, while the other depends only on the output resistance. Hence, in this limit of large AC_f, it is again perhaps legitimate to speak of separate input and output poles.

As a closing comment on these time constants, it is often the case that C_oC_i is small compared to the other terms in the brackets of Eqn. 7. Then the estimate of the non-dominant pole simplifies further, to

$$\tau_2 \approx \frac{R_o[C_o + C_i]}{A}. \qquad \text{(EQ 8)}$$

2.1 Discussion

Using the exact second-order transfer function in conjunction with open-circuit time constant estimates, we have derived approximate expressions for the pole locations. We've deduced that the product of the pole frequencies is always independent of the open-circuit gain A. Furthermore, in the limit of large AC_f, we've found that one may safely speak of separate input and output poles, although the intuitive value of doing so in general remains dubious. Be that as it may, it is frequently instructive to provide alternative derivations, since traversing different analytical paths can often confer valuable additional insights. Here, let us attempt to deduce the limiting behavior without resorting to use of the complete second-order transfer function.

Consider first the special case in which the feedback capacitance, C_f, is zero. We can find the two poles by inspection: There is a pole at $-1/R_iC_i$ and another at $-1/R_oC_o$. Because the input and output ports are clearly isolated from each other here, we can speak confidently of separate input and output poles. From this point on, though, it becomes increasingly less clear that we can continue to do so.

As C_f increases from zero, the Miller effect works its magic, causing an increase in the capacitive load as seen from the viewpoint of the driving source (whose Thévenin equivalent resistance is R_i). In the limit where the Miller multiplication dominates, the effective capacitance grows to a value that is approximately A times as large as C_f, leading to an "input" pole whose time constant is

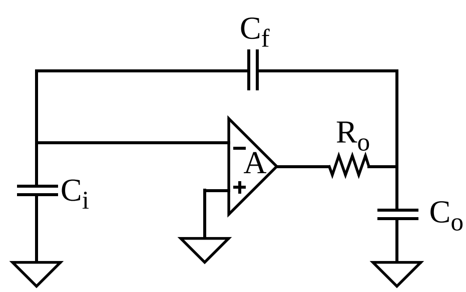
$$\tau_1 \approx AR_iC_f. \qquad \text{(EQ 9)}$$

If we persist in the view that there is also an "output" pole, then we might imagine that it involves the output resistance and/or the output capacitance. This assumption is somewhat shaky, and its invocation here is partially the result of already knowing the correct answer. But let's press on.

We can find the effective resistance facing C_o by applying a test voltage source and computing the ratio of v_{test} to i_{test} as is traditional in these sorts of cases. To simplify (considerably) the derivation, assume that the non-dominant pole is at a frequency high compared with $1/R_iC_i$. In that case, we may neglect R_i.

With that simplification, the system appears as follows:

FIGURE 2. Cheesy model for estimating frequency of "output" pole



Finding the effective resistance facing C_o is then found with the usual procedure (i.e., applying a test voltage, measuring the response current, etc.):

$$R_{out} \approx \frac{R_o}{1 + A_{eff}} = \frac{R_o}{1 + A\left(\frac{C_f}{C_f + C_i}\right)} = \frac{R_o(C_f + C_i)}{(1 + A)C_f + C_i} \qquad \text{(EQ 10)}$$

From this last equation, we see that R_{out} decreases from a value of R_o when $C_f = 0$ (corresponding to the case of no feedback capacitor), to a value of $R_o/(1+A)$ (corresponding to the case of an infinitely large feedback capacitor). The time constant associated with C_o therefore changes by a similar ratio, moving to higher frequency as the feedback capacitor increases.

In the limit of very large A and C_f, the output resistance approaches R_o/A. Furthermore the output and input capacitors are essentially shorted together in that limit, leading to an output pole time constant of

$$\tau_o \approx \frac{R_o(C_o + C_i)}{A}, \qquad \text{(EQ 11)}$$

in accord with the limiting estimate for the non-dominant pole derived earlier (Eqn. 7). This “derivation” thus confirms that pole splitting can be viewed as resulting from a lowering of the input pole frequency (because of Miller multiplication of feedback capacitance), and an increase in the output pole frequency (because of the feedback-induced reduction in output resistance). However, the large number of approximations (several seemingly quite arbitrary in nature) required to arrive at these insights using the second approach should leave the reader somewhat unsure about the intuitive value of the input and output pole viewpoint.

3.0 Summary and Final Comments

As we’ve seen, a Miller feedback capacitor can create a dominant pole while simultaneously making the non-dominant pole even less dominant. As a result, a feedback amplifier that uses this type of compensation will appear single-pole over a wide frequency range, making it particularly easy to use without worrying about stability problems. Just as valuable (arguably, even more valuable) is that the analysis of such a circuit is made trivial because there is only one significant pole to worry about.

The analysis presented in this brief note offers a method for determining the location of the second pole, so that the limits of validity of the first-order approximation can be quantified with reasonable accuracy. In the worst case, the assumption of a dominant pole is in error by a factor of two because the open-circuit time constant estimate of the dominant pole is precisely equal to the pole time constant sum. The worst case error thus occurs when the two poles are equal in frequency. Because the leading coefficient in the second-order denominator polynomial is the product of the pole time constants, a factor of two error in estimating one leads to a factor of two error in the other. This one-octave error bound is often sufficiently tight that robust circuits can be designed even if the approximations are used somewhat outside of their strict domains of validity. Perhaps more important, such circuit design can proceed in the absence of potentially misleading notions about input and output poles.

