Linear Regulators: Theory Of Operation and Compensation

Introduction

The explosive proliferation of battery powered equipment in the past decade has created unique requirements for a voltage regulator that cannot be met by the industry standards like the LM340 or the LM317. These regulators use an NPN Darlington pass transistor (*Figure 1*), and will be referred to in this document as **NPN regulators**. The demand for higher performance is being met by the newer **low-dropout (LDO)** regulators and **quasi-LDO** regulators.

The NPN Regulator

The NPN Darlington pass transistor with PNP driver used in an NPN regulator requires that at least 1.5V to 2.5V be maintained from input-to-output for the device to stay in regulation. This minimum voltage "headroom" (called the **dropout voltage)** is:

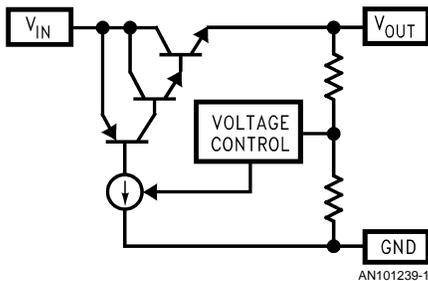$$V_{DROP} = 2V_{BE} + V_{SAT} \text{ (NPN REG)}$$

FIGURE 1. NPN REGULATOR

The LDO Regulator

In the low-dropout (LDO) regulator, the pass transistor is a single PNP transistor (*Figure 2*). The big advantage of the LDO is that the PNP pass transistor can maintain output regulation with very little voltage drop across it:

$$V_{DROP} = V_{SAT} \text{ (LDO REGULATOR)}$$

Full-load dropout voltages < 500 mV are typical. At light loads, dropout voltages can fall as low as 10 to 20 mV.
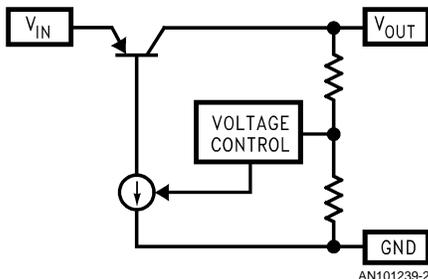

FIGURE 2. PNP LDO REGULATOR

The QUASI-LDO Regulator

Another regulator configuration that is becoming very popular in certain applications (like 5 - 3.3V conversion) is the quasi-LDO regulator (*Figure 3*). The quasi-LDO is so named

because it is "half way" between the NPN Darlington and the true LDO. The pass transistor is made up of a single NPN transistor being driven by a PNP. As a result, the dropout voltage is less than the NPN Darlington regulator, but more than an LDO:
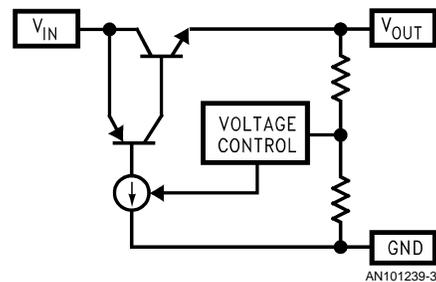
$$V_{DROP} = V_{BE} + V_{SAT}$$

FIGURE 3. QUASI-LDO REGULATOR

Regulator Operation

All three of these regulator types regulate the output voltage to a fixed (constant) value using the same technique (*Figure 4*).

The output voltage is sampled (measured) through a resistive divider which is fed into the inverting input of the error amplifier. The non-inverting input is tied to a reference voltage, which is derived from an internal bandgap reference. The error amplifier will always try to force the voltages at it's input to be equal. To do this, it sources current as required to provide sufficient load current to maintain the output voltage at the regulated value which is given by:

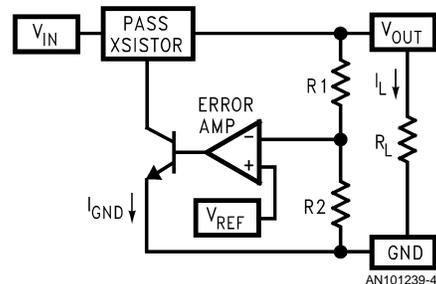$$V_{OUT} = V_{REF} (1 + R1/R2)$$

FIGURE 4. VOLTAGE REGULATOR

Performance Comparison

The primary differences in performance between the NPN, LDO, and quasi-LDO are in the parameters of **dropout voltage** (previously defined) and **ground pin current**. For this analysis, we will define ground pin current (**I$_{GND}$**) as shown in *Figure 4*, neglecting the small IC bias currents which also flow to ground. It can be seen that the value of I$_{GND}$ is the load current I$_L$ divided by the gain of the pass transistor.

The high gain of the Darlington in an NPN regulator means it requires very little drive to source I$_L$, so it's ground pin cur-

rent is very low (typically a few mA). The quasi-LDO also has very good performance, with products like National's LM1085 being able to source more than 3A with less than 10 mA of ground pin current.

The ground pin current of an LDO is typically much higher. At full load current, PNP beta values of 15 - 20 are not unusual, which means the LDO ground pin current can be as high as 7% of the load current.

A big advantage of NPN regulators is that they are unconditionally stable (most require no external capacitors). An LDO does require at least one external capacitor on the output to reduce the loop bandwidth and provide some positive phase shift. Quasi-LDOs typically require some output capacitance, but much less than an LDO and with less restrictive limits on its performance characteristics.

Feedback and Loop Stability

All voltage regulators use a feedback loop to hold the output voltage constant. The feedback signal experiences changes in both gain and phase as it goes through the loop, and the amount of phase shift which has occurred at the unity gain (0 dB) frequency determines stability.

Bode Plots

Understanding stability requires the use of **Bode Plots**, which show the loop gain (in **dB**) plotted as a function of frequency (*Figure 5*). Loop gain and associated terms are defined in the next sections.

Loop gain can be measured on a network analyzer, which injects a low-level sine wave into the feedback path and then measures the gain response while the frequency of the signal is swept from DC up to the frequency where the gain drops well below 0 dB.

Bode plots are convenient tools because they contain all the information necessary to determine if a closed-loop system is stable. However, decoding the information contained in a Bode plot requires understanding the key elements: **loop gain, phase margin, poles and zeros**.
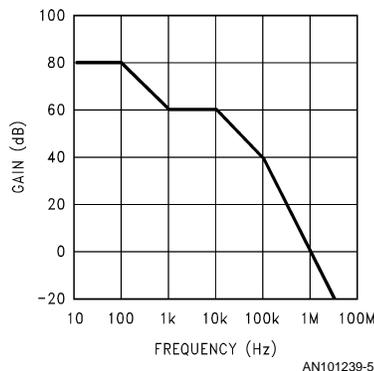
FIGURE 5. TYPICAL BODE PLOT

Loop Gain

Every closed-loop system has a characteristic called **loop gain**. In this analysis of voltage regulators, loop gain will be defined as the magnitude of the voltage gain that the feedback signal experiences as it travels through the loop. The block diagram of the LDO in *Figure 2* will be redrawn to illustrate this concept (*Figure 6*).
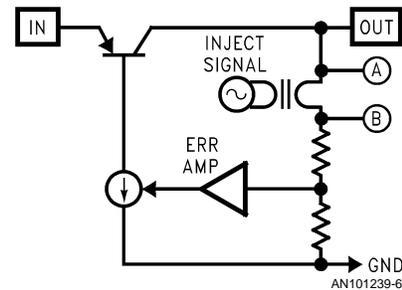
FIGURE 6. LOOP GAIN EXAMPLE

A transformer is used to inject an AC signal into the feedback path between points ″A″ and ″B″. Using this transformer, a small-signal sine wave is used to ″modulate″ the feedback signal. The AC voltages at ″A″ and ″B″ are measured and used to calculate loop gain.

The loop gain is defined as the ratio of the two voltages:

$$\text{Loop Gain} = V_A/V_B$$

it is important to note that the signal starting at the V_B point has a phase shift introduced into it as it travels through the loop (eventually arriving at the V_A point). The amount of phase shift is critical in determining stability.

Feedback

Feedback is used in all voltage regulators to hold the output voltage constant. The output voltage is sampled through a resistive divider (*Figure 6*), and that signal is fed back to one input of the error amplifier. Since the other input of the error amplifier is tied to a reference voltage, the error amplifier will supply current as required to the pass transistor to keep the regulated output at the correct DC voltage.

It is important to note that for a stable loop, **negative feedback** must be used. Negative feedback (sometimes called **degenerative feedback)** is opposite in polarity to the source signal (see *Figure 7*).
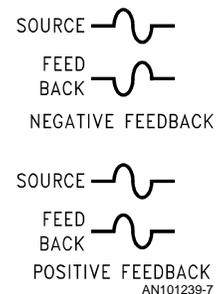
FIGURE 7. FEEDBACK SIGNALS

Because it is opposite in polarity with the source, negative feedback will always cause a response by the loop which opposes any change at the output. This means that if the output voltage tries to rise (or fall), the loop will respond to force it back to the nominal value.

Positive Feedback occurs when the feedback signal has the same polarity as the source signal. In this case, the loop responds in the same direction as any change which occurs at the output. This is clearly unstable, since it does not **cancel out** changes in output voltage, but **amplifies** them.

It should be obvious that no one would intentionally design positive feedback into the loop of a linear regulator, but **negative feedback becomes positive feedback if it experiences a phase shift of 180˚.**

Phase Shift

Phase shift is defined as the total amount of phase change (referred to the starting point) that is introduced into the feedback signal as it goes around the loop. Phase shift (expressed in degrees) is most often measured using a network analyzer.

Ideal negative feedback is 180˚ out of phase wit the source (*Figure 8*), so its ″starting point″ is at −180˚. This ″180˚″ offset can also be seen in *Figure 7*, as the negative feedback waveforms are exactly one half cycle shifted with respect to each other.
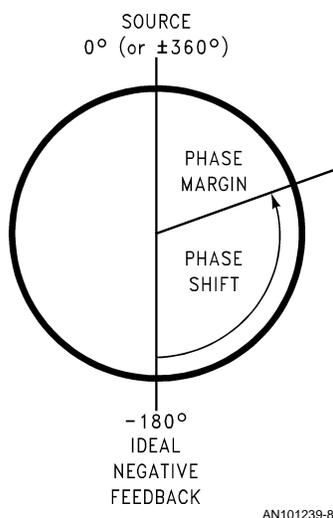

FIGURE 8. PHASE SHIFT MAP

It can be seen that starting at −180˚, an additional phase shift of 180˚ (positive or negative) brings the signal back to zero, **which is in phase with the source signal and would cause the loop to be unstable**.

Phase Margin

Phase margin is defined as the difference (in degrees) between the total phase shift of the feedback signal and −180˚ **at the frequency where the loop gain is equal to 0 dB (unity gain)**. A stable loop typically needs at least 20˚ of phase margin.

Phase shift and phase margin can be calculated using the poles and zeros present in the Bode plot.

Poles

A pole (*Figure 9*) is defined as a point where the slope of the gain curve changes by −20 dB/decade (with reference to the slope of the curve prior to the pole). Note that the effect is additive: each additional pole will increase the negative slope by the factor ″n″ x (−20 dB/decade), where ″n″ is the number of additional poles.
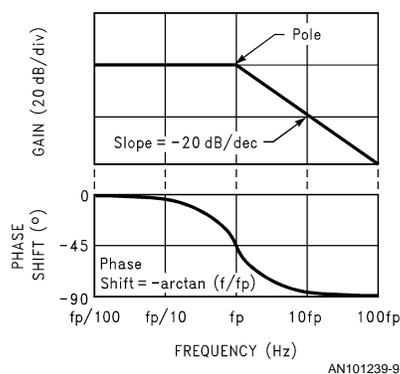

FIGURE 9. POLE GAIN/PHASE PLOT

The phase shift introduced by a single pole is frequency dependent, varying from 0 to −90˚ (with a phase shift added by a pole (frequency). The most important point is that nearly all of the phase shift added by a pole (or zero) occurs within the frequency range one decade above and one decade below the pole (or zero) frequency.

NOTE: a single pole can add only −90˚ of total phase shift, so at least two poles are needed to reach −180˚ (which is where instability can occur).

Zeros

A zero (*Figure 10*) is defined as a point where the gain changes by +20 dB/decade (with respect to the slope prior to the zero). As before, the change in slope is additive with additional zeros.
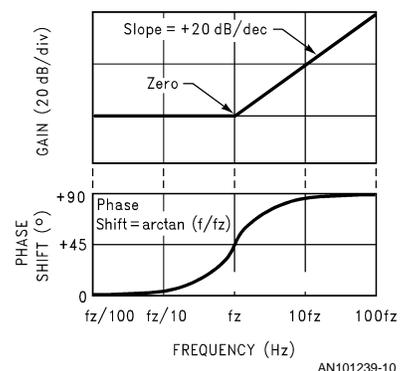

FIGURE 10. ZERO GAIN/PHASE PLOT

The phase shift introduced by a zero varies from 0 to +90˚, with a +45˚ shift occurring at the frequency of the zero.

The most important thing to observe about a zero is that it is an ″**anti-pole**″, which is to say its effects on gain and phase are exactly the opposite of a pole.

This is why zeros are intentionally added to the feedback loops of LDO regulators: they can cancel out of the effect of one of the poles that would cause instability if left uncompensated.

Bode Plot Analysis

A Bode plot which contains three poles and one zero (*Figure 11*) will be analyzed for gain and phase margin.

The DC gain is assumed to be 80 dB, with the first pole occurring at 100 Hz. At that frequency, the slope of the gain curve changes to –20 dB/decade.
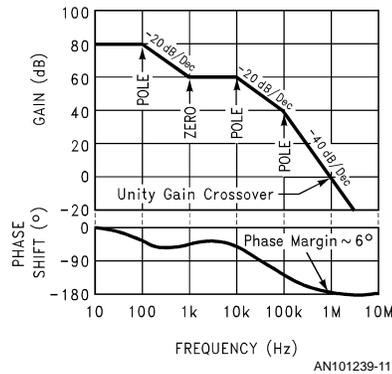

FIGURE 11. BODE PLOT WITH PHASE INFO

The zero at 1 kHz changes the slope back to 0 dB/decade until the second pole at 10 kHz, where the gain curve slope returns to –20 dB/decade.

The third and final pole at 100 kHz changes the gain slope to the final value of –40 dB/decade.

It can also be seen that the unity-gain (0 dB) crossover frequency is 1 MHz. The 0 dB frequency is sometimes referred to as the **loop bandwidth**.

The plot of the phase shift shows how the various poles and zeros contribute their effect on the feedback signal. To produce this plot, the phase shift at each frequency point was calculated based upon summing the contributions of every pole and zero at that frequency. The phase shift at any frequency "f" which is caused by a pole frequency located at frequency "f_p" can be calculated from:

$$\text{Pole Phase Shift} = -\arctan(f/f_p)$$

The phase shift resulting from a zero located at frequency "f_z" can be found using:

$$\text{Zero Phase Shift} = \arctan(f/f_z)$$

Is this loop stable? To answer that question, we need only the phase shift at 0 dB (which is 1 MHz in this case). Finding this does not require complex calculations:

As stated in the previous sections, a pole or zero contributes nearly its full phase shift in the frequency range one decade above and below the center frequency of the pole (or zero). Therefore, the first two poles and the first zero contribute their full phase shifts of –180˚ to +90˚, respectively, resulting in a net phase shift of –90˚.

The final pole is exactly one decade below the 0 dB frequency. Using the formula for Pole Phase Shift, this pole will contribute –84˚ of phase shift @ 1 MHz. Added to the –90˚ from the two previous poles and the zero, the total phase shift is –174˚ (which means the phase margin is 6˚). This loop would either oscillate or ring severely.

NPN Regulator Compensation

The pass transistor of the NPN regulator (see *Figure 1*) is connected in a circuit configuration known as **common collector**. An important characteristic of all common collector circuits is low output impedance, which means the pole from the power stage that it places in the loop gain occurs at a very high frequency.

The NPN regulator uses a technique called **dominant pole compensation** because it has no inherent low-frequency poles. In this case, a capacitor is built into the IC which places a pole in the loop gain at a low frequency (*Figure 12*).
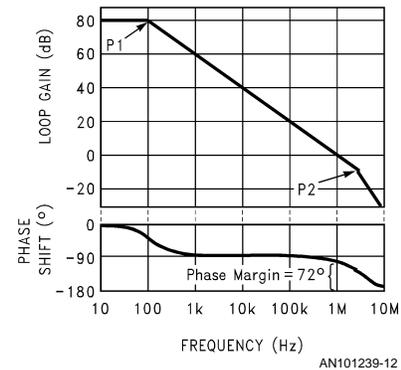

FIGURE 12. BODE PLOT FOR NPN REGULATOR

This **dominant pole** (shown as P1) for a typical NPN regulator is set at about 100 Hz. The 100 Hz pole causes the gain to decrease at a rate of –20 dB/decade until the second pole (P2) which is located at 3 MHz. At that point, the slope of the gain plot changes to –20 dB/decade.

The frequency of P2 is primarily due to the NPN power transistor and associated drive circuitry, so it is sometimes referred to as the **power pole**. Since P2 occurs at a frequency where the loop gain is –10 dB, its contribution to phase shift at the 0 dB frequency (1 MHz) will be small.

To determine stability, it only requires that the phase margin at the 0 dB frequency be calculated:

The first pole (P1) will contribute –90˚ of phase shift, but the second pole (P2) will add only –18˚ of negative phase shift @ 1 MHz (0 dB). This means the total phase shift @ 0 dB is –108˚, which yields a phase margin of 72˚ (which is very stable).

It should also be noted that simple observation would clearly show this loop is stable, since reaching –180˚ of phase shift (the point of instability) would require the full contribution of –90˚ (each) from both poles, and P2 is too high in frequency to contribute significant phase shift at the 0 dB frequency (1 MHz).

LDO Regulator Compensation

The PNP transistor in an LDO regulator (*Figure 2*) is connected in a configuration called **common emitter**, which has a higher output impedance than the common collector configuration in the NPN regulator. This adds an additional low-frequency pole **whose frequency is dependent both on load resistance and output capacitance**. The frequency of this pole (which will be designated P_L for **load pole**) is found from:

$$f(P_L) = 1/(2\pi \times R_{LOAD} \times C_{OUT})$$

The presence of the frequency-variable load pole P_L means that the simple dominant pole compensation method used in the NPN regulator will not work in an LDO unless additional compensation is added. To illustrate why this is true, the loop gain of a 5V/50 mA LDO regulator will be illustrated using these assumptions:

At maximum load current, the load pole (P_L) occurs at a frequency given by:

$$P_L = 1/(2\pi \times R_{LOAD} \times C_{OUT}) = 1/(2\pi \times 100 \times 10^{-5}) = 160 \text{ Hz.}$$

The internal compensation will be assumed to add a fixed pole (**P1**) at 1 kHz.

A 500 kHz power pole (which will be designated **P_{PWR}**) is present due to the PNP power transistor and driver.

The DC gain is assumed to be 80 dB.

R_L = 100Ω (which is the value at maximum load current)

C_{OUT} = 10 µF.

using the conditions stated above, a Bode plot (*Figure 13*) is drawn. It is immediately obvious that this loop is not stable: the two poles P_L and P1 will each contribute –90° of phase shift to reach –180° at the 0 dB frequency (which is about 40 kHz in this example).

To reduce the negative phase shift (and prevent oscillations), a zero must be added to the loop. A zero can contribute as much as +90° of positive phase shift, which will cancel out the effects of one of the two low frequency poles.
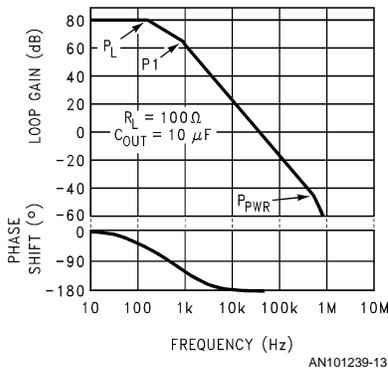
FIGURE 13. LDO GAIN PLOT WITHOUT COMPENSATION

Nearly all monolithic LDO regulators require that this zero be added to the loop, and they derive it from a characteristic that is inherent in the output capacitor: **equivalent series resistance** (usually referred to **ESR**).

LDO Compensation using ESR

Equivalent series resistance (ESR) is a characteristic that is present in every capacitor. It can be modeled electrically as a resistance that is placed in series with the capacitor (*Figure 14*).
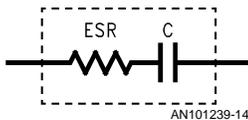
FIGURE 14. CAPACITOR SHOWING ESR

The ESR of the output capacitor puts a zero in the loop gain which can be used to reduce excess negative phase shift.

The frequency where the zero occurs is directly related to the value of the ESR and amount of output capacitance:

$$F_{ZERO} = 1/(2\pi \times C_{OUT} \times ESR)$$

Using the example in the previous section (Bode plot shown in *Figure 13*), we will assume that the value of C_{OUT} = 10 µF and the output capacitor ESR = 1Ω, which means a zero will occur at 16 kHz.

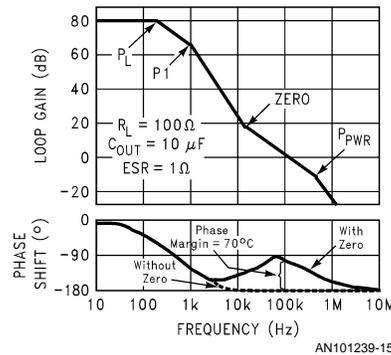*Figure 15* shows how this added zero will change the unstable plot into a stable one:

FIGURE 15. ESR ZERO STABILIZES LDO

The bandwidth of the loop is increased so that the 0 dB crossover frequency moves form 30 kHz to 100 kHz.

The zero adds a total of +81° positive phase shift at 100 kHz (the 0 dB frequency). This will reduce the negative phase shift caused by the poles P_L and P1.

Since the pole P_{PWR} is located at 500 kHz, it adds only –11° of phase shift at 100 kHz.

Summing all poles and zeros, the total phase shift at 0 dB is now –110°. This corresponds to a phase margin of +70°, which is extremely stable.

This illustrates how an output capacitor with the correct value of ESR can generate a zero that stabilizes an LDO.

ESR and Stability

Virtually all LDO regulators require that the ESR of the output capacitor be within a set range to assure regulator stability.

The LDO manufacturer provides a set of curves which define the boundaries of the stable region, plotted as a function of load current (*Figure 16*).
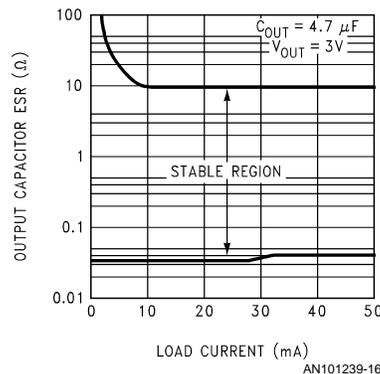
FIGURE 16. ESR RANGE FOR Typical LDO

To explain why these boundaries exist, the effects of low and high ESR on phase margin will be illustrated using the example previously developed.

High ESR

Using the examples developed in the previous sections, we will change the conditions and assume the ESR of the 10 µF output capacitor is increased to 20Ω. This will decrease the frequency of the zero to 800 Hz (*Figure 17*). Reducing the frequency of the zero causes the loop bandwidth to increase, moving the 0 dB crossover frequency from 100 kHz to 2 MHz.

This increase bandwidth means that the pole P_{PWR} occurs at a gain value of +20 dB (compared to –10 dB in *Figure 14*).

Analyzing the plot (*Figure 17*) for phase margin, it can be assumed that the zero cancels out either P1 or P_L. This means the loop has a two-pole response with the low frequency pole contributing –90˚ of phase shift and the high frequency pole P_{PWR} contributing about –76˚ of phase shift.
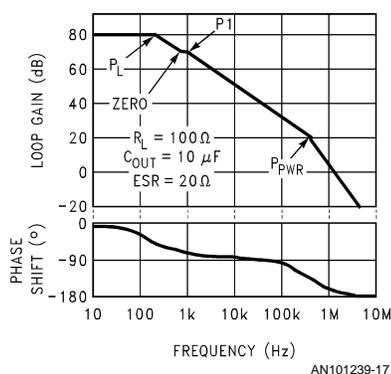

FIGURE 17. HIGH ESR CAUSES UNSTABLE LOOP

Although this appears to leave a phase margin of 14˚ (which **might** be stable), bench test data shows that ESR values > 10 Ω usually cause instability because of phase shifts contributed by other high-frequency poles which are not shown in this simplified model.

Low ESR

An output capacitor with a very low ESR value can cause oscillations for a different reason.

Continuing the example developed in the previous section, we will now reduce the ESR of the 10 µF output capacitor to 50 mΩ, increasing the frequency of the zero to 320 kHz (*Figure 18*).
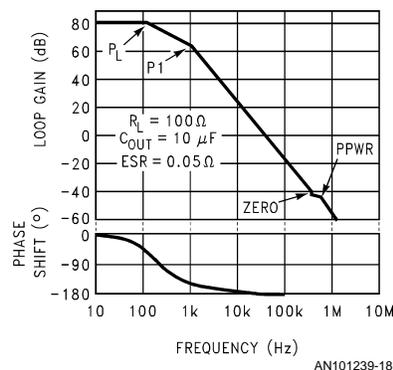

FIGURE 18. LOW ESR CAUSES UNSTABLE LOOP

When the plot is analyzed for phase margin, no calculations are required to see that it is unstable.

The –90˚ phase shift from each of the two poles P1 and P_L will produce a total phase shift of –180˚ at the 0 dB frequency.

For this system to be stable, a zero is needed that would provide positive phase shift before the 0 dB point. However, since the zero is at 320 kHz, it's too far out to do any good (and is cancelled out by P_{PWR}).

Output Capacitor Selection

Since the output capacitor is the user's tool for compensating a monolithic LDO regulator, it must be selected very carefully. Most cases of oscillations in LDO applications are caused by the ESR of the output capacitor being too high or too low.

When selecting an output capacitor for an LDO, a solid tantalum capacitor is usually the best choice (except for parts specifically designed for ceramic capacitors like the LP2985). Tests performed on an AVX 4.7 µF Tantalum showed an ESR of 1.3Ω @ 25˚C, a value that is almost perfectly centered in the stable region (*Figure 16*).

Also very important, the ESR of the AVX capacitor varied less than 2:1 over the temperature range of –40˚C to +125˚C. Aluminium electrolytic capacitors are notorious for exhibiting an exponential increase in ESR at cold temperatures, and are not suitable for use as an LDO output capacitor.

It must be noted that large (≥ 1 µF) ceramic capacitors typically have very low ESR values (< 20 mΩ), and will cause most LDO regulators to oscillate if connected directly to the

output (except the LP2985). A ceramic capacitor can be used if some external resistance is added in series with it to increase the effective ESR. Large value ceramics also have a poor tempco (typically Z5U) which means the capacitance will drop in half as the temperature is increased or decreased to the operating limits.

Quasi-LDO Compensation

When evaluating the quasi-LDO regulator (*Figure 3*) for stability and compensation, it has some of the electrical characteristics of both the LDO and NPN regulator. Since the quasi-LDO uses an NPN pass device, it is in the common-collector configuration which means it's output device node (emitter) looks like a relatively low impedance.

However, because of the base of the NPN is being driven from a high-impedance PNP current source, the regulator output impedance of a quasi-LDO is not as low as the NPN regulator with an NPN Darlington pass device (but is much lower than a true LDO which drives the regulator output off the collector of a PNP).

This means that the troublesome power pole of a quasi-LDO is at a lower frequency than the NPN regulator, so some compensation (output capacitance) is required to make a quasi-LDO stable. Of course, the pole is at a much higher frequency than the LDO, so the quasi-LDO requires less capacitance and the ESR is not as critical.

For example, the LM1085 quasi-LDO rated for 3A of load current requires only 10 µF of Tantalum output capacitance to assure complete stability over all line and load conditions. No ESR graphs are given, since the value of ESR is not critical as it is in an LDO.

Low-ESR Specific LDO's

National Semiconductor does have LDO regulators like the LP2985 and LP2989 which are specifically designed to work with extremely low ESR capacitors like surface-mount ceramics. This type of capacitor can have ESR values as low as 5-10 mΩ, which will cause most typical LDO regulators to oscillate (as demonstrated in *Figure 18*.

To make the LP2985 stable with such low ESR values, an internal zero is built in which takes the place of the ESR zero previously provided by the Tantalum output capacitor. The effect of this is to shift the stable ESR range downward. A typical LDO with no added internal zero might be stable ESR range from about 100 mΩ to 5 Ω (well suited for Tantalums but not ceramics). The stable range for the LP2985 extends down to 3 mΩ, and has an upper limit of about 500 mΩ so it can be used with ceramics.

The reason the upper limit is moved down can be understood by referring to *Figure 15*. As previously stated, the zero is now built into the LDO, so the ESR zero resulting from the output capacitor must stay at a high enough frequency that it does not cause the bandwidth to get too wide where high frequency poles would add enough phase shift to produce oscillations.

The FET Advantage

An LDO regulator can be built using a P-FET as the pass transistor (see *Figure 19*).
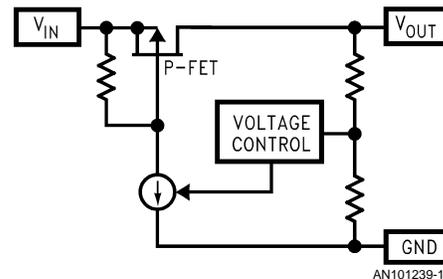

AN101239-19

FIGURE 19. P-FET LDO REGULATOR

To see why using a P-FET LDO would be advantageous, it should be noted that all of the base current required by the power transistor in a PNP LDO (*Figure 2*) flows out of the ground pin and back to the negative input voltage return. Therefore, this base drive current is drawn from the input supply but does not drive the load, so it generates wasted power that must be dissipated within the LDO regulator:

$$\text{PWR (Base Drive)} = V_{IN} \times I_{BASE}$$

The amount of base current required to drive the PNP is equal to the load current divided by the beta (gain) of the PNP, and beta may be as low as 15 - 20 (at rated load current) in some PNP LDO regulators. The wasted power generated by this base drive current is very undesirable (especially in battery-powered applications). Using a P-FET solves this problem, since the Gate drive is very small.

Another advantage of the P-FET LDO is that the dropout voltage can be made very small by adjusting the ON-resistance of the FET. For monolithic regulators, FET power transistors typically will give a lower ON-resistance per unit area than bipolar ONP devices. This allows making higher current regulators in smaller packages.

Notes