Audio Engineering Society

Convention Paper 5695

Presented at the 113th Convention
2002 October 5–8 Los Angeles, California, USA

Audio Power Amplifier Output Stage Protection

Eric Mendenhall[1]

[1] *Gibson Labs, Redondo Beach, CA, USA*
Correspondence should be addressed to eric.mendenhall@gibson.com

ABSTRACT

This paper reviews a progression of circuits used for protecting bipolar power transistors in the output stages of audio power amplifiers. Design oriented methods of determining the protection locus are shown in a mathematical and graphical procedure. The circuits are then expanded from their standard configurations to allow for transient excursion beyond steady state limits, and thermally dependent protection limits, to better match the protection limits to the actual output stage capability. This allows the protection scheme to prevent output stage failure in the least restrictive way.

A new method is shown for achieving a junction temperature estimation system without the use of a multiplier.

INTRODUCTION

A class B power amplifier output stage is typically made from bipolar power transistors, which are characterized by their safe operating area, SOA. The SOA limits are determined by bond wire limits, power dissipation limits, breakdown voltage, and second breakdown limits. (see Fig. 1)

For analysis and design, amplifier output voltage waveforms and load impedance magnitude and phase are used to place the load lines on the same V_{CE}-I_{CE} plane as the SOA. If there are conditions under which the SOA will be exceeded, protection circuitry should be activated to avoid output device failure. Since SOA is a function of temperature and time, the protection circuits can be tailored to take these variables into account.

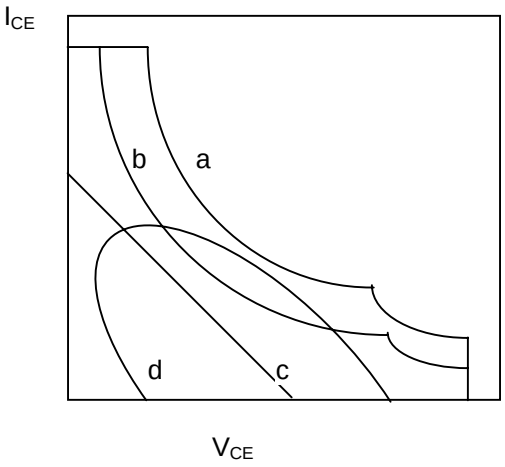

Fig 1. SOA and load lines.
(a) 25C steady state SOA
(b) 100C steady state SOA
(c) resistive load line
(d) reactive load line

TIME AND TEMPERATURE INVARIANT CIRCUITS

Simple Current Limiter

The use of the circuit in Fig. 2 will give a protection locus as shown in Fig. 3. The protection threshold is:

$$I_{CE} = (V_{BE}/R_E)(R_1+R_2)/R_2 \quad (1)$$

We can see this circuit, although very simple, cannot be optimized to adequately protect the output device and preserve normal operation of the amplifier.

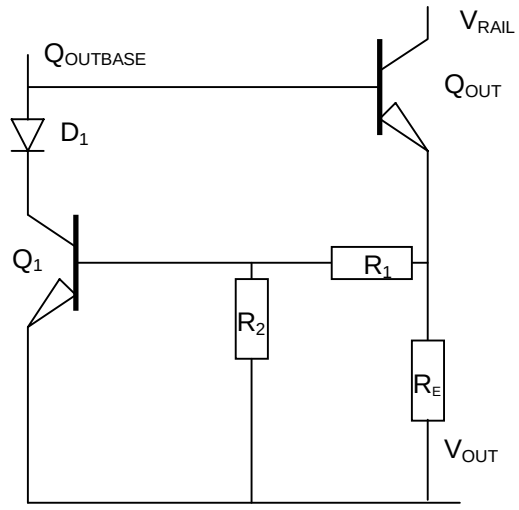

Fig. 2. Simple Current Limiter.

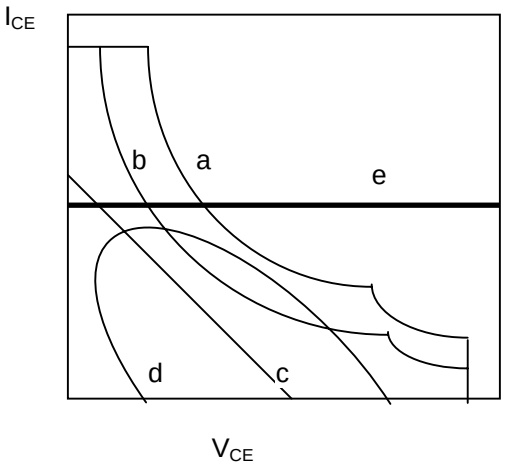

Fig. 3. Simple Current Limiter
(a) 25C steady state SOA
(b) 100C steady state SOA
(c) resistive load line
(d) reactive load line
(e) protection limit

Single Slope Limiter

The use of the circuit in Fig. 4 will give a protection locus as shown in Fig. 5. This circuit is better than a simple current limit, and can be used when high voltage transistors are used with lower voltage rails. Under these conditions, the second breakdown limits will not come into play.
The protection threshold is:

$$I_{CE} = \left[V_{BE} - \frac{V_{CE}}{R_3} \frac{(R_1R_2)}{(R_1+R_2)} \right] \frac{(R_1+R_2)}{(R_2R_E)} \qquad (2)$$

This is derived by superposition with several simplifying assumptions. We assume R_3 is large with respect to R_1 and R_2, and that R_E is small with respect to R_1, R_2, and R_3. We also assume V_{RAIL} is large compared to the saturation voltage of Q_{OUT} and the voltage across the emitter resistor, R_E.

Design Procedure

First R_2 is chosen to be a value that will scale all the rest of the resistors. Next we determine the Y intercept (V_{CE} = 0, $I_{CE} = I_{LIM1}$) by setting $V_{CE} = 0$ in Eq. (2) and solving for R_1.

$$R_1 = \frac{R_2(I_{LIM1}R_E - V_{BE})}{V_{BE}} \qquad (3)$$

At this point, V_{OUT} is equal (ideally) to V_{RAIL} and there is no current flow in R_3, so the equivalent circuit is the same as the simple current limit of Fig. 2.

Next R_3 is determined by the X intercept ($V_{CE} = V_{LIM1}$, $I_{CE} = 0$). At this point, There is no voltage across R_E, and R_1 is in parallel with R_2. Set $I_{CE} = 0$ in Eq. (2) and solve for R_3.

$$R_3 = \frac{V_{LIM1}(R_1R_2)}{V_{BE}(R_1+R_2)} \qquad (4)$$

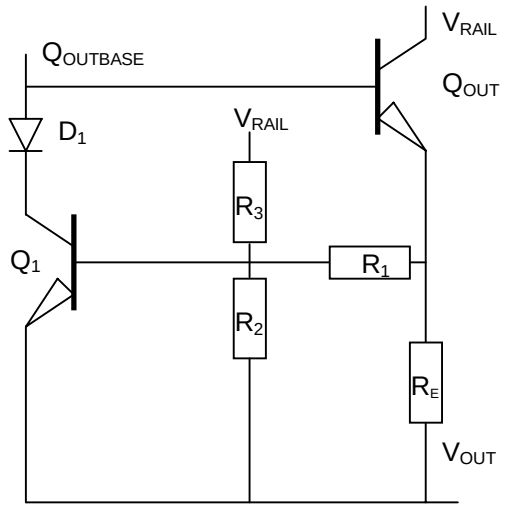

Fig. 4. Single Slope Limiter.

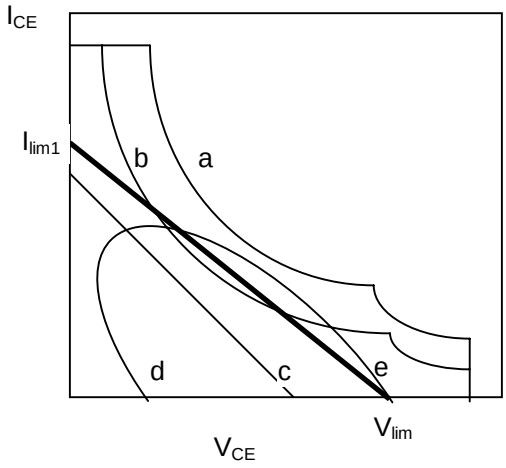

Fig. 5. Single Slope Limiter.
(a) 25C steady state SOA
(b) 100C steady state SOA
(c) resistive load line
(d) reactive load line
(e) protection limit

Two Slope Limiter

The use of the circuit in Fig. 6 will give a protection locus as shown in Fig. 7. The protection threshold is:

For $V_{CE} > V_{RAIL}$:

$$I_{CE} = \left[V_{BE} - \frac{V_{CE}}{R_3} \frac{(R_1R_2)}{(R_1+R_2)} \right] \frac{(R_1+R_2)}{(R_2R_E)} \qquad (5)$$

For $V_{CE} < V_{RAIL}$:

$$I_{CE} = \frac{V_{BE}(R_1+R_2)}{(R_2R_E)} + \frac{(V_{RAIL}-V_{CE})R_1}{(R_4R_E)} - \frac{V_{CE}R_1}{(R_3R_E)} \qquad (6)$$

This circuit matches the protection limits more closely to the actual device capability. The circuit is analyzed in parts, first with D_2 OFF, then with D_2 ON. With D_2 OFF, we degenerate to the single slope limiter of Fig. 4. The diode will come into play when the output voltage is above ground, that is, when V_{CE} is less than or equal to V_{RAIL}. Fig. 8 shows the two lines, where the single slope limit of I_{LIM1} is used for analysis, but is overtaken by the new limit I_{LIM2} after the breakpoint.

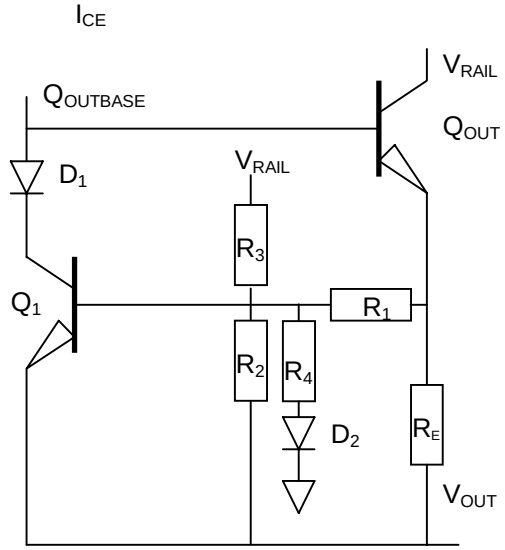

Fig. 6. Two Slope Limiter.

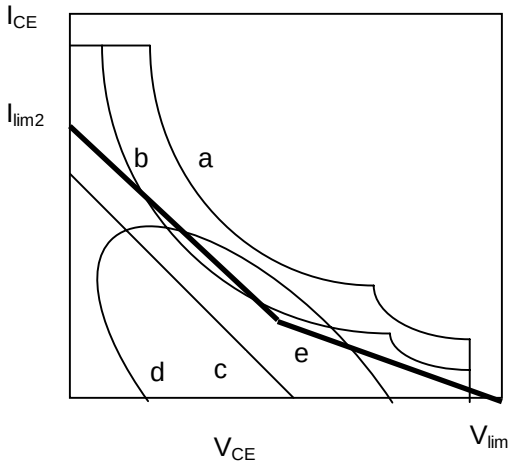

Fig. 7. Two Slope Limiter.
(a) 25C steady state SOA
(b) 100C steady state SOA
(c) resistive load line
(d) reactive load line
(e) protection limit

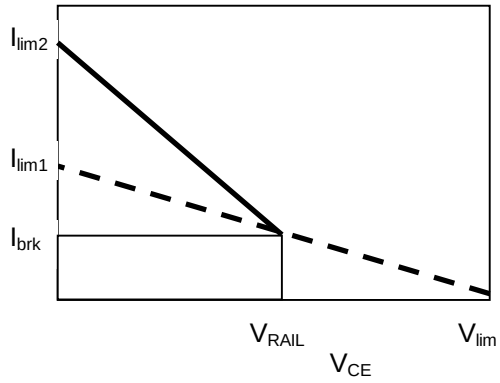

Fig. 8. Two Slope Limiter.

Design procedure begins with a single slope equivalent circuit, but instead of choosing I_{LIM1}, I_{BRK} is used as a starting point. We write the equation of the dashed line as

$$I_{CE} = (-I_{LIM1} / V_{LIM1})V_{CE} + I_{LIM1} \qquad (7)$$

Solving for I_{LIM1} , setting $V_{CE} = V_{RAIL}$ and $I_{CE} = I_{BRK}$:

$$I_{LIM1} = \frac{I_{BRK} V_{LIM1}}{(V_{LIM1} - V_{RAIL})} \qquad (8)$$

So after choosing I_{BRK}, we find I_{LIM1} and proceed with the single slope design method until R_1, R_2, and R_3 are defined.
R_4 is determined next. The only point to determine is the Y intercept ($V_{CE} = 0$, $I_{CE} = I_{LIM2}$).
Set $V_{CE} = 0$ in Eq. (6) and solve for R4:

$$R_4 = \frac{V_{RAIL}R_1R_2}{I_{LIM2}R_ER_2 - V_{BE}(R_1+R_2)} \quad (9)$$

EXTENSION TO MULTIPLE BREAKPOINTS

Four Slope Limiter

The use of the two slope circuit as a piecewise linear approximation to a constant power curve is better than a single slope approximation; allowing more breakpoints would result in an even more accurate approximation. Fig. 9 shows a limiter modified for four slope operation.

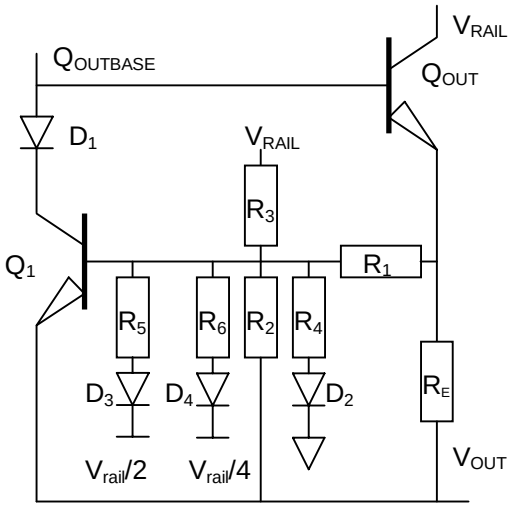

Fig. 9. Four Slope Limiter.

This requires three breakpoints, one of which is set at ground as before. The other two are set by using reference voltages in between ground and the main rail voltages, some or all of which may be already present in a class G or H design. These are shown in this example as being set at V_{RAIL} / 2 and V_{RAIL} / 4 for convenience, but in practice they could be set wherever the constant power curve is best fit. The area beneath D_2's breakpoint V_{RAIL} is fit with a single slope, while the area above it makes better use of the additional breakpoints. The extra references allow for the breakpoints as shown in Fig. 10 at a cost of two resistors and two diodes. It can be seen that the four slope limiter has a protection characteristic that can be optimized to simulate device characteristics accurately.

An additional benefit is the ability to tailor the protection slope in the second breakdown region, where allowable device dissipation is less than that determined by the thermally limited region. A least-squares method could be used to determine the optimum curve fitting, but a graphical approach is sufficient for most purposes.

Design Procedure is much the same as for a two slope design, starting with the rightmost portion of the curve with D_2, D_3, and D_4 OFF, then proceeding to add breakpoints. Details are left as an exercise for those interested.

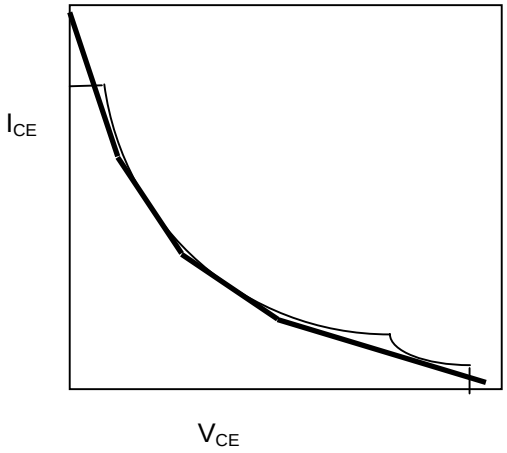

Fig. 10. Four Slope Limiter.

Nonlinear Override

The protection curve in Fig. 10 is best matched in the constant power region by setting the Y intercept I_{lim2} to exceed the bond wire limitations, shown by the abrupt truncation of the curve. In order to further improve the circuit, we can add a maximum current limit with a suitable reduced time constant by adding D_5, D_6, and R_7 shown in the circuit of Fig. 11.

In this circuit, the effective value of R_1 is reduced by paralleling R_7 via D_5 and D_6 when a maximum drop across R_E is exceeded. D_5 and D_6 are simply constant drops; the string may be replaced by a zener, or a different number of diodes.

Although the four slope curve also exceeds the breakdown voltage limit, no nonlinear override should be necessary, as the power supply rail voltages are chosen such that this limit cannot be exceeded.

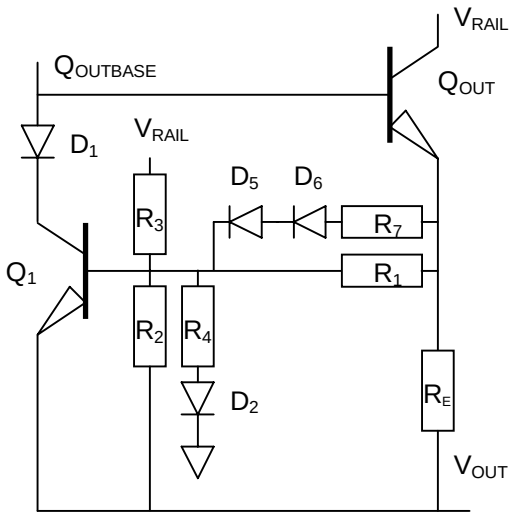

Fig. 11. Two Slope Limiter with Nonlinear Override

TEMPERATURE VARIANT CIRCUITS

Two Slope Current Limit

For simplicity, I will address temperature variations with respect to the two slope limiter.

The SOA of a transistor is a function of temperature, as shown in Fig. 1, curves a and b. It follows that the protection circuitry will be most effective when its limits are set as a function of temperature as well. The first effect to consider is the negative temperature coefficient of Q_1's V_{BE}, which works in the proper direction, tightening limits as Q_1's temperature is increased.

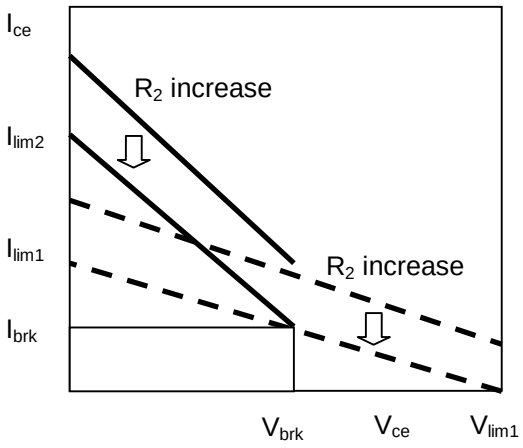

Fig. 12. Temperature Variance with Thermistor.

This can be used most effectively when Q_1 is thermally coupled to the output devices. If they are not coupled, then the protection circuit's limits will still be a function of temperature, but it may be ambient temperature that they are tied to.

The change in V_{BE} is not significant enough to provide a swing in the protection limits that would be large enough to accurately track Q_{OUT}'s SOA versus temperature.

A thermistor can be used to enhance the thermal tracking of the limiter if it is configured to track Q_{OUT}'s case temperature. The thermistor can either replace R_2 or work in series or parallel, depending on the type. The effects of this are shown in Fig. 12.

TIME VARIANT CIRCUITS

Two Slope Current Limit

For simplicity, I will address time variations with respect to the two slope limiter.

The SOA limits used above are all steady state limits. Since most of the SOA is thermally limited, the steady state SOA may be exceeded on a transient basis as long as the average power does not exceed the SOA limits. This is due to the transient thermal impedance of the device itself acting as a capacitance in the electrical analog shown in Fig. 13, modeled with a single RC network, although multiple time constants can be applied for a more accurate model.

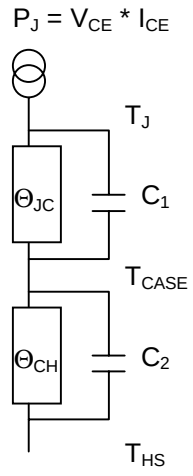

Fig. 13. Thermal-Electrical Analog.

Since most transistors used in high power audio amplifiers have a transient thermal impedance response curve with a 10-90% rise time on the order of 200mS, and most audio amplifiers have a corner

frequency around 20Hz (25mS half cycles), we are justified in treating the total thermal impedance from junction to case as an effective averaging circuit. This means we can exceed the SOA by a factor of 2 in the thermally limited region, as long as we have the 50% duty cycle inherent in a class B design.

Allowing the limiter to use a thermal equivalent RC network will let transients pass while leaving the steady state limits intact. The capacitances should be chosen to match the output device's transient thermal impedance 10-90% rise time, although this will just give a simple approximation to a complex system.

Fig. 14 shows the two slope limiter with the RC networks applied. C_2 could be placed in Miller configuration, but time constants will vary with beta. Other configurations could allow a smaller time constant in the second breakdown region.

In this configuration, it is obvious that the circuit is behaving as an analog of the power device in Fig. 13, but instead of computing instantaneous dissipation with a true multiplier, we have a two slope piecewise linear summing approximation. The use of a linear multiplier with a thermal analog RC circuit and temperature depedent limits is the basis of the protection circuit in [3]. The multiplier circuit allows a close match of device capability and protection limits in the thermally limited region, but can be difficult to implement in practice, requiring two pairs of matched transistors to create the multipliers along with surrounding circuitry. Since the output of the multiplier is a high impedance, additional active circuitry is needed to apply the RC networks, and bias supplies are needed. The bias supplies are not an additional cost in the case of a grounded output, where this circuit has typically been applied. A conventional topology would require output referenced bias supplies.

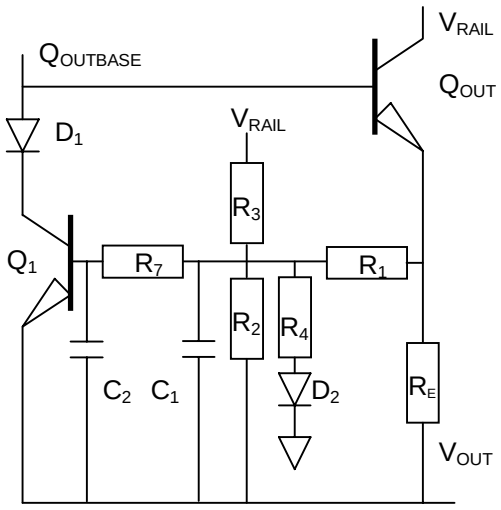

Fig. 14. Two Slope Limiter with Time Constants.

The circuit shown behaves such that average output voltage on R_2 is a weighted sum of inputs, rather than the product, which means that the threshold of activation can be matched to that of a multiplier but the approximation becomes increasingly less accurate as the inputs increase or decrease. For example, if at a given point V_{CE} and I_{CE} contributed equally to the voltage on R_2, then if each were halved the voltage on R_2 would be halved while the true power was reduced to one fourth. This is a large error, but here we are concerned mostly with accuracy near the limits and getting improved performance from a simple circuit rather than making the leap to a multiplier topology.

Protection Activation

When the limiter circuit must take protective action, Q_1 is turned on, drawing away drive current from the output device. This action would result in benign clipping into a resistive load. Since a highly reactive load has a more intrusive load line, it is more likely that the protection will be activated under reactive conditions than resistive conditions.

A capacitive load will accept a reduction in drive current by simply slowing down the rate of charge or discharge, causing distortion that may be worse than the resistive case, but still relatively benign.

A highly inductive load, however, will react to an attempted reduction in drive current, negative dI/dt, with a reversal of the voltage across it. In our half-circuit shown, a positive output voltage will produce a positive output current into an inductor, but a negative dI/dt will force the output voltage toward the negative rail, which increases V_{CE} on the output transistor, driving the protection circuit even harder. This regenerative effect causes the output voltage to clamp to the negative rail, a very audible and well known distortion mechanism.

Since the protection circuit is concerned with both V_{CE} and I_{CE}, we could enable the circuit to reduce either one in response to an unsafe condition. Reduction of the V_{CE} can be achieved by allowing rail voltages to fall, easily achieved in a regulated switching power supply. Since the protection circuit is at the output voltage potential, a level translation circuit is needed. An optocoupler can be used, as shown in Fig. 15. The output of the optocoupler, not shown, would be tied to the power supply voltage regulation circuit to reduce output voltage when triggered. Since the reaction time of the voltage reduction is not likely to be fast enough to adequately protect under all conditions, such as short circuit current limiting, the local action of Q_1 cannot be removed. An additional diode D_5 is added to increase Q_1's local protection threshold, which allows the optocoupler to be activated first and Q_1 to be activated as a last resort.

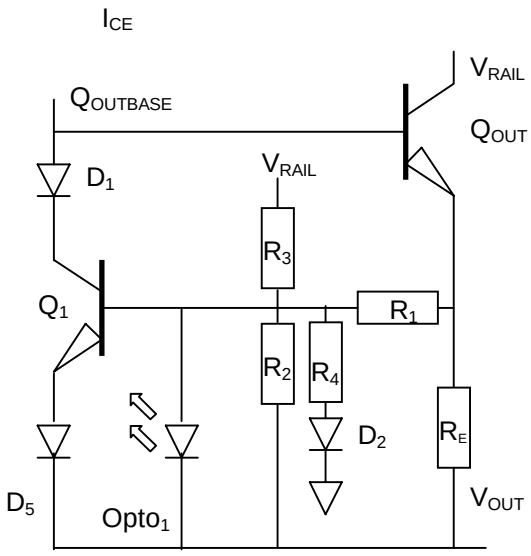

Fig. 15. Optocoupler Activation for Supply Voltage Reduction

Regardless of the type of protection mechanism used, the primary goal should be to avoid output stage failure without affecting normal operation. If a properly designed protection circuit is being activated with normal load, program, and ambient conditions then the problem is not with the protection circuit but with the power stage design, indicating inadequate silicon or heatsinking. If the protection circuit is designed with limits based on the hottest possible operation and is then activated at lower temperatures, then it is not actually protecting the output devices but is causing distortion with no benefit.

CONCLUSIONS

Several protection circuits for the output devices of audio power amplifiers have been reviewed with a design-oriented approach. A more accurate method of protection with four linear slopes has been shown. A nonlinear override circuit has been shown.

Time and temperature invariant circuits have been shown to be less than ideal; methods for adding time and temperature variance to these circuits have been shown. The time and temperature variant circuits have been shown to be a piecewise linear approximation to the multiplier-based junction temperature estimator approach, and can be designed to allow protection limits that track the actual device capability as the multiplier circuit does.

Several of the ideas and circuits presented in this paper are in the process of United States and international patent protection.

REFERENCES

[1] Eric Mendenhall, “Computer Aided Design and Analysis of Class B and Class H Power Amplifier Output Stages,” Audio Engineering Society Preprint 4329 (1996)
[2] Bill Roehr, Bryce Shiner “Transient Thermal Resistance – General Data and its Use,” Motorola Application Note AN-569
[3] Gerald R. Stanley “Thermal Protection Circuit for the Die of a Transistor,” United States Patent 4,644,288 (1982)
[4] Analog Devices “Nonlinear Circuits Handbook,” (1976)