Parallel PWM Amplifiers

INTRODUCTION

Paralleling the outputs of two or more amplifiers can be a critical solution to many designs needing extra power. This application note considers the application of that concept to PWM, or class-D, amplifiers. Paralleling PWM amplifiers can increase current in the whole system, lower the effective R_{DSON} of each output element, and deliver much higher power to the load than otherwise possible with a single device.

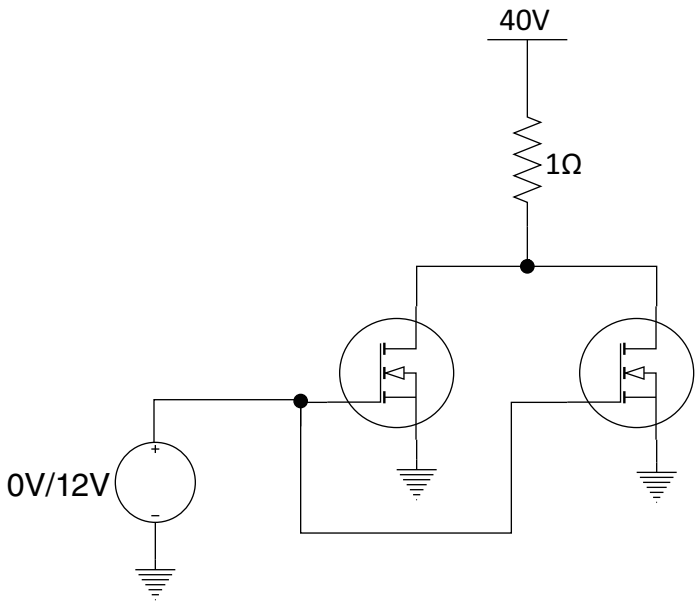
Keep in mind that the process of paralleling PWM amplifiers is not trivial. The length of this application note should serve as an indication of how many factors an engineer must consider before designing such a system. The schematics are large, the PCBs will be crowded, and circuit performance must be de-rated for safe operation. Nonetheless, Apex has investigated this matter thoroughly, and offers the following guide to designing such circuits.

THE BASIC CHALLENGE

What makes this high-power technique so difficult is the fact that each output element has slightly different properties. The wide distribution of R_{DSON}s, turn-off/-on delays, and rise/fall times, even within the same production lot, makes it unwise to connect the outputs directly.

For example, take two 25 A N-channel MOSFETs, one with R_{DSON} = 11 mΩ and the other with R_{DSON} = 25 mΩ, and match them up pin-for-pin. Now put them in the circuit of figure 1 and drive the gates with +12 V. Both FETs turn fully on and start conducting. Ideally, the two FETs share the current equally, receiving roughly 20 A each. In reality, the current is uneven - about 12 A and 28 A. 28 A violates the MOSFET's rating. This can cause permanent damage, create a drain-source fusion, and lead to failure in the rest of the system.

Figure 1: Parallel MOSFETs



Now consider that both FETs have R_{DSON} = 11 mΩ, but the turn-off delay for one FET is 20 ns longer than the other. As the gates are simultaneously driven from +12V to 0V, the faster device will inevitably turn off first. This FET is in the clear; its drain current never exceeded the absolute maximum pulse current of, let's say, 35 A. The slower FET, however, experiences up to 20 ns of 40 A drain current. Realistically, this probably won't damage the device because the fall times of each FET overlap with this time, and 20 ns is a very short time to suffer 40 A. This analysis also doesn't consider source- and drain- inductance, which will limit the current at the time of transition.

Clearly, the kind of stress associated with time-delay differences is far less severe than that of on-resistance differences. But consider the fact that these conditions are occurring at switching frequencies of over 20 kHz. Repetitive stress like this will reduce the lifetime of the MOSFET and lead to an early system death.

These same problems can happen inside an Apex PWM amplifier. Most Apex PWM amplifiers are hybrid devices, so there is no guarantee that the individual MOSFET die used in one unit will match the die of another. If you connect two Apex SA160s pin-for-pin, the part-to-part variations will cause current to share unevenly. Timing delays may be even worse than that of the examples above. Each switching amplifier has internal circuitry that can cause larger variations in turn-off delay time, rise time, etc. Depending on the product, the difference in transition time from the same input signal can be over 100 ns. This problem is not mitigated by inductance because Apex switching amplifiers are specifically designed to *reduce* inductance in these high-current paths.

Timing becomes even more crucial when you consider dead-time. Refer to AN30 section 3.3 "Accuracy – Closed Loop" for more information on dead-time. If the dead times of all paralleled channels do not overlap between the top- and bottom- switching elements, then shoot-through can (and will) destroy the amplifier in a matter of seconds. Apex switching amplifiers with analog inputs have an internally-programmed dead-time that cannot be changed. Switching amplifiers with digital control inputs allow the user full flexibility with setting the dead-time.

PARALLELING THE HALF BRIDGES IN A SINGLE MONOLITHIC CHIP

Monolithic chips are the exception to the rule. Theoretically, two MOSFETs made at the exact same time, under the exact same conditions, should have the exact same properties. That is the rationale behind directly shorting the outputs of the identical channels in one monolithic chip (direct-paralleling).

Note that this is only possible when the gates of each MOSFET can be controlled independently! Some Apex switching products are designed specifically as H-bridges, so when the top-side FET of one channel is closed, the top-side FET of the other channel is automatically open, and vice versa. This not only makes paralleling the channels futile, but it would cause permanent shoot-through if you even tried. Luckily, most Apex monolithic PWM amplifiers allow independent control of each FET.

To reiterate, direct-paralleling of the individual channels of a PWM product can ONLY be done when BOTH of the following are true:

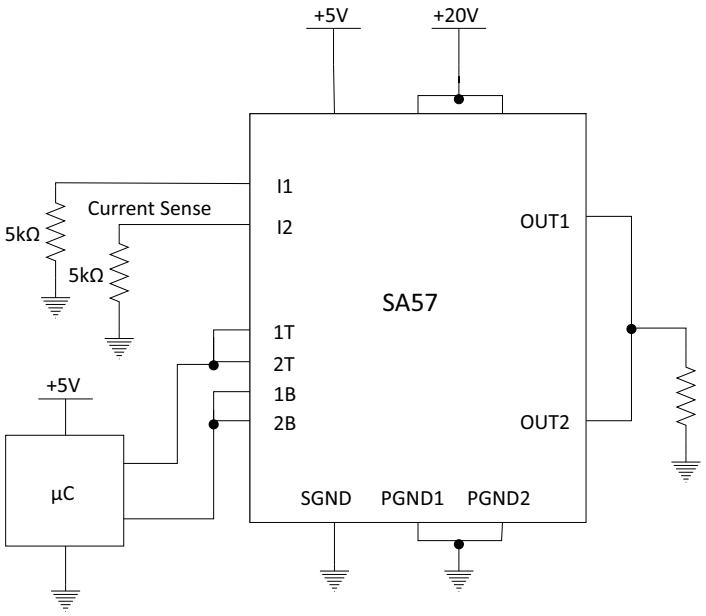
1. The device is a SINGLE monloithic chip (no hybrids!)
2. EACH output element (top and bottom FETs of all half bridges) can be independently controlled.

Apex produces several monolithic ICs for switching applications. SA57 will be used as an example. SA57 is a full H-bridge with complete digital control – all 4 switching elements can be turned on and off inde-

pendently using a microcontroller. This combination of full digital control and identical characteristics for each channel gives the ideal circumstance for a parallel device. The below schematic in figure 2 shows that both half-bridges are shorted together to create a single effective half-bridge with, theoretically, half the R_{DSON} and double the output current capability.

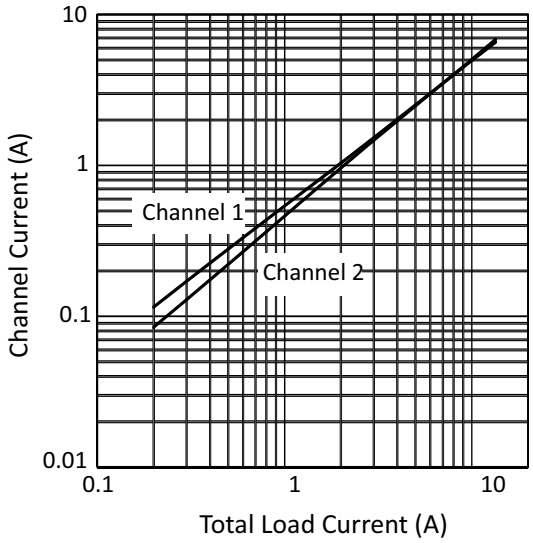
Another great feature of SA57 is its low-loss current sense ability. Using these pins, the current flowing through the top FET of each half-bridge can be monitored. This feature was used to collect the following current-sharing data from the circuit represented in figure 2.

Figure 2: SA57 Direct-Parallel Half-Bridges Schematic



Looking at figure 3, current distribution is clearly not perfect at the low end of the scale. At 200 mA of output current, channel 2 carries only 42% of the total. Fortunately, as current increases, the ratio comes closer to 50%/50% between the two channels. This allowed the SA57, which is rated for 8 A of continuous output current per channel, to perform well beyond that limit – the highest tested current was 13.5 A. Additionally, R_{DSON} of the top FETs in parallel was brought down to 0.19 Ω.

Figure 3: Current Distribution in SA57 Direct-Parallel Half-Bridges



The uneven distribution of current on the low end of the scale is likely due to thermal equilibrium, which is discussed in later sections. This experiment demonstrates that the two top FETs on the same monolithic chip do NOT have identical properties, but they are closer-matched than two discrete FETs selected at random.

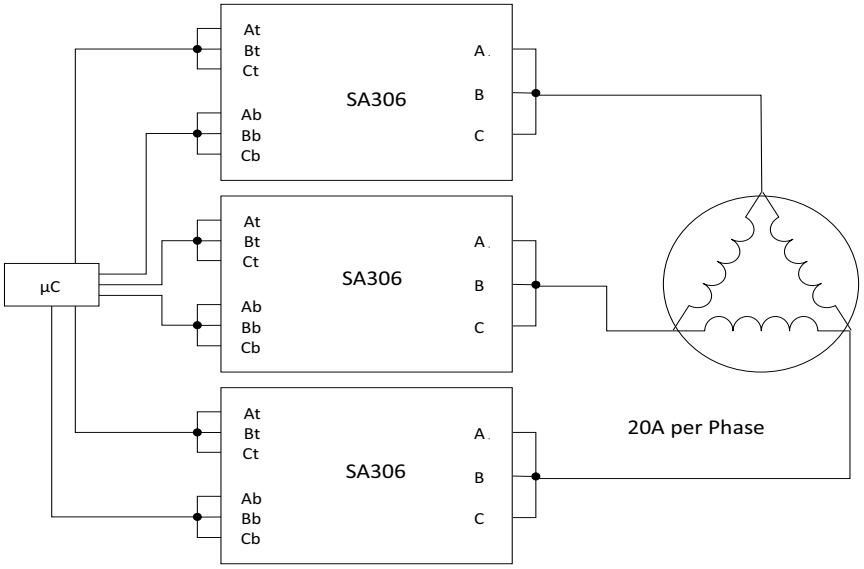
Although this typical unit shows ideal current-sharing abilities at currents above ~2 A, it is important to note that not all units will behave this way. Matching R_{DSON}s and other attributes are NOT guaranteed between channels of a single monolithic chip, so some margin must be made in total current. While this circuit could theoretically handle (8 A * 2) = 16 A, a 2 Ampere margin would protect more units from suffering an early death.

A word of caution: this technique does not address the high power dissipation that accompanies such high currents. At 13.5 A and R_{DSON} = 0.19 Ω, the package dissipates almost 35 W at 100% duty cycle, which is only sustainable when the device is well-cooled. The investigations above were done at low duty cycle, with adequate heatsinking.

Using this direct-paralleling technique, a designer can parallel the two half-bridges in two separate devices, then bridge them together to form the complete H-bridge. Figure 4 shows this concept applied to a 3-phase motor driver: A single SA306 has its 3 channels shorted together. This is done to 3 separate chips, which now become the 3 separate phases driving the motor. As a result, SA306 (an 8 A continuous device) can be used to drive a motor which requires up to 20 A! …With an asterisk: Power dissipation must be managed appropriately. Additionally, motors have a bad habit of drawing more than their fair share of current

during startup, stall, and reversal conditions. For more information, see AN45 “Driving 3-Phase Brushless Motors.”

Figure 4: 3-Phase Parallel System



PARALLELING MULTIPLE PARTS TOGETHER

For all products without independent FET control OR products that are not formed on a single silicon die, three basic rules must be followed:

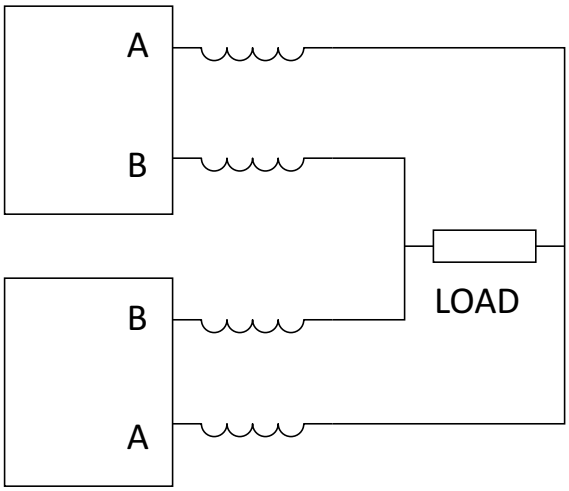
1. DO NOT Connect Outputs Directly Between Amplifiers.

All that talk about different R_{DSON}s, timing delays, and any differences between the circuits will spell out disaster for your paralleled device if the outputs are simply shorted together. There must be some network to share the current dynamically and evenly between the two amplifiers. In addition, safety features must be used to protect the amplifier in case the FETs don’t turn on at the exact same time.

There is an argument here that even unmatched R_{DSON}s will reach thermal equilibrium and share current more evenly than otherwise expected. To illustrate, go back to the example of figure 1. With 28 A of drain current, the 11 mΩ MOSFET will dissipate more power than the 25 mΩ at 12 A. The 11 mΩ FET will get hotter, and with its positive temperature coefficient, it will increase its resistance much more than the 25 mΩ FET will. Putting some numbers to that hypothetical situation ($R_{TH(JA)}$ = 1 °C/W and Temperature Coefficient = 1.5 mΩ/°C), the true current distribution comes closer to 22 A and 18 A after the circuit reaches thermal equilibrium. This is much better than the 28 A and 12 A difference that occurs at cold startup. Just remember – this condition doesn’t take place until the circuit is fully warmed up. Thermal equilibrium should not be relied upon to share current evenly unless the load can be applied gradually to the circuit.

Thermal Equilibrium also does nothing for differing propagation delays. For this, it is best to place some inductance at each output of every paralleled device, as in figure 5.

Figure 5: Isolating Inductors Protect the Amplifiers from Timing Differences



Not only do these inductors serve as a filter to demodulate the PWM signal, but they temporarily limit the current in the event of asynchronous turn-on/-off conditions. To determine the appropriate value of inductance, a first-order filter can be designed using Apex Power Design or the equation in AN30 section 5.2 "How Much Inductance?". For these techniques, design the filter for twice the load resistance; the filter will be duplicated on each amplifier.

The absolute minimum amount of inductance can be calculated with the following equation:

$$L_{minimum} = \frac{V_S \; t_{DIFF}}{I_{PEAK} - \frac{I_{LOAD}}{n}}$$

V_S = Supply Voltage

t_{DIFF} = Difference in propagation delays (to be safe, use $\frac{1}{10 * F_{SWITCHING\_MAX}}$)

I_{PEAK} = Peak current in amplifier's datasheet (de-rate for safety)

I_{LOAD} = Maximum current through load

n = number of paralleled amplifiers

This technique is not only limited to first order filters. 2nd- and 3rd- order filters will work even better to reduce transition currents. See figure 6 for a full schematic of a parallel system with a 2nd-order filter.

In cases where inductance is unfavorable, ferrite beads may be used in their place. A ferrite bead will behave more like a resistor at low frequencies (the frequency of circuit operation), but more like an inductor at higher frequencies (the frequencies associated with switching transients).

Note that none of these techniques guarantee even current sharing in all conditions. Thermal equilibrium forces the currents close to 50%/50% sharing, and a complex filter can bring the design even closer. But ultimately, there will always be some mismatch as long as the output elements in the two different devices have any differences in their characteristics. Therefore, the total current rating of a parallel system will always be LESS than (number of paralleled amplifiers) * (current rating of one amplifier). The necessary margin will depend on the surrounding circuit, the quality of the layout, and the specific mismatch in component characteristics. In circumstances with good circuit layout and a properly designed filter, a margin of 10% of the rated continuous output current *per paralleled device* is sufficient. The following equation calculates the suggested maximum current allowed in a paralleled system:

$$I_{parallel\ max} = n * I_{datasheet\ max} * 0.9$$

Where n = number of paralleled amplifiers.

2. DO NOT Operate with an Analog Control Signal

The reason behind this rule is simple: analog control relies on the internally generated dead-time, which may be too short to synchronize in a parallel amplifier.

Some Apex amplifiers are strictly analog, some are strictly digital, and some are flexible. For those that are flexible, read the datasheet for instructions and schematics. These will often show the amplifier being driven with a DSP or microcontroller. Further explanation of digital mode operation is given in later sections.

The key advantage of a digital control system is that it uses a “disable” pin to set all output elements at high impedance for a set amount of time (dead time). This needs to be done at every transition, and it needs to last long enough for all propagation delays to pass through every amplifier in the system. An example pseudo-code is given below for SA160 operating in digital mode:

```
while(1) {

  //dead time
  disablePin = HIGH;
  delayUs(1);         //delay 1 microsecond

  //transition to high
  inputPin = HIGH;
  disablePin = LOW;
  delayUs(tHigh);     //delay for time ON

  //dead time
  disablePin = HIGH;
  delayUs(1);         //delay 1 microsecond

  //transition to low
  inputPin = LOW;
  disablePin = LOW;
  delayUs(tLow);      //delay for time OFF

}
```

On the subject of choosing an appropriate dead time, enough time needs to be allotted so that all of the propagation delays *and* turn-off transition times in every paralleled device can pass. The dead-time listed in the product datasheet is optimized only for turn-off transition times, and does not consider propagation delays. The total propagation duration is not specified in the datasheet.

A dead time of $\frac{1}{40*F_{SWITCHING\_MAX}}$ is generally safe for any Apex parallel PWM product. Shortening this suggestion is taking a risk that two or more dead-times will "miss" each other and cause shoot-through from one amplifier's top FET into another amplifier's bottom FET. Adding inductance in this path by placing a filter coil just outside of each output (as in figure 5) can mitigate this problem of shoot-through.

3. DO NOT Have More than One Source of Any Signal.

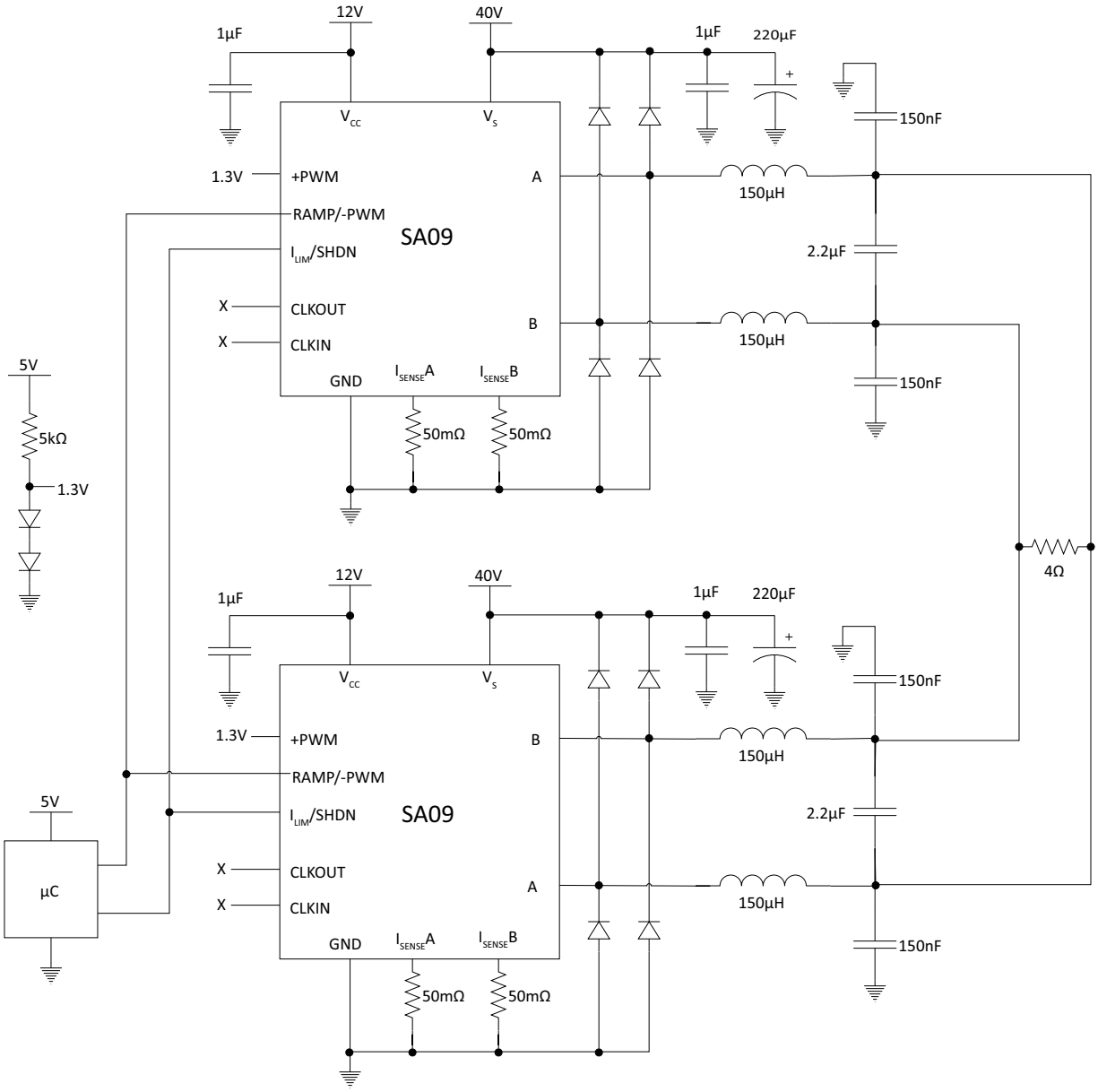
The best thing to do here is physically short all of the input pins together, and all of the disable pins together. Then, drive each of these nodes with the microcontroller as if it was a single amplifier.

It may be tempting to connect each pin individually to the microcontroller and drive them simultaneously through software. Although this would make the circuit more versatile and re-configurable, it adds more potentials for propagation delays and transmission failures.

DESIGN EXAMPLE

In a real application, the above techniques work about as well as the direct-paralleled monolithic chips. Figure 6 shows a parallel SA09 circuit designed to deliver 300 W to a 4 Ω load. At DC, this requires 8.7 A of load current, which exceeds SA09's 5 A continuous current rating AND its 7.5 A peak current rating. Using the equation for maximum paralleled output current, an 8.7 A output is possible by paralleling 2 SA09s.

Figure 6: SA09 300 W Parallel Full Circuit Schematic



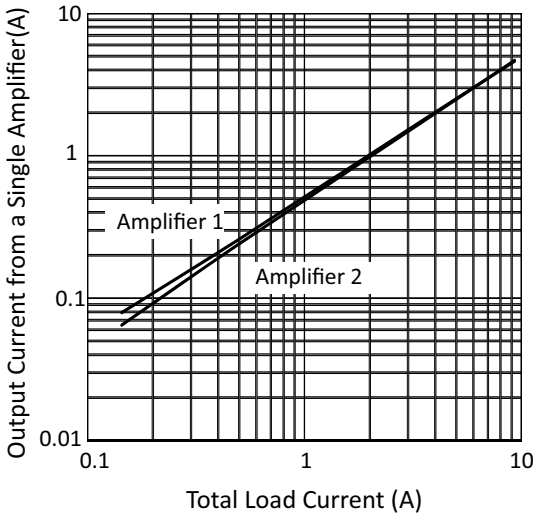
In this example, a 2nd-order filter was designed (for an 8 Ω load with 6 kHz cutoff frequency) using Apex Power Design. This yields inductors of 150 µH and a capacitor of 2.2 µF. The equation for minimum inductance (above) indicates that at least 2.3 µH is required; this is satisfied by the 150 µH from Power Design.

Power Design also suggests a C_{LEG} of 150 nF. The purpose of C_{LEG} is to maintain a stable common mode voltage of the two outputs. This capacitor has very little effect on the overall transfer function. In a parallel design, these components should be duplicated on the output of each paralleled unit, just like the filter.

The inputs for each SA09 are set up for digital inputs. A microcontroller is connected to control the duty cycle, switching frequency, and dead time. Switching frequency is set up for 100 kHz, and the dead time is programmed for 62.5 ns. The inputs are also directly shorted together, satisfying rule #3.
Power supply bypass capacitors and output protection diodes are duplicated on each amplifier. The duplication of these components is critical to prevent local oscillations and output stage failure.

Measuring the current through four 50 mΩ current-sense resistors (one for each half-bridge), the current again shows better than 1% matching at high output power (see figure 7). At the low end of the current scale, the current sharing can be as poor as 40%/60%. As currents increase above 4 A, however, the current sharing improves to within 49%/51%.

Figure 7: Current Distribution in Double-Filter Paralleled SA09 Circuit

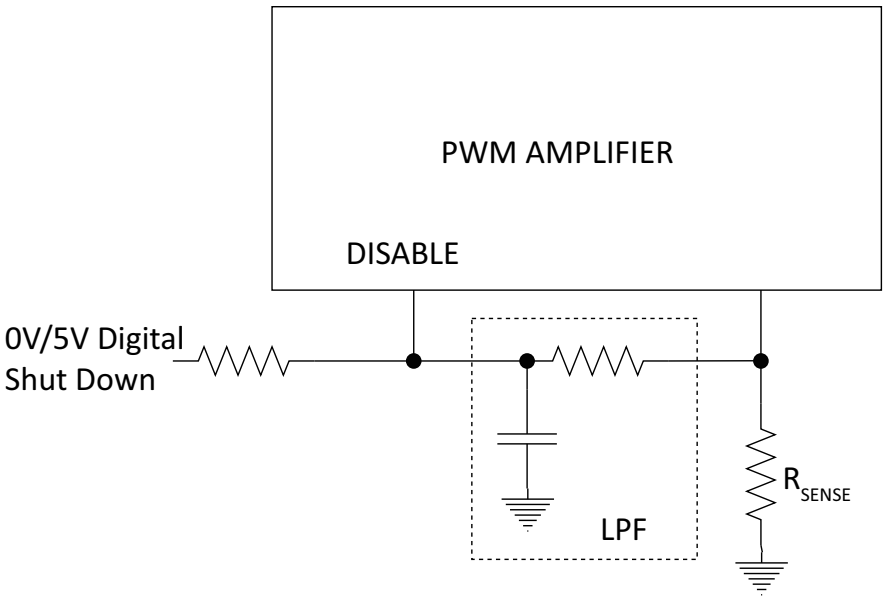


CURRENT LIMIT CONSIDERATIONS

The traditional Apex PWM amplifier brings out the high-current return paths to 1 or 2 “current sense” pins, where a low-value sense resistor can measure current through the bottom switching elements. This technique works well to generate information about load current without large power losses.

The circuit in figure 8 shows a typical connection for the current limit circuit of a singular PWM amplifier. The sense resistor connects to the disable pin (sometimes named I_{LIM} or Shutdown) through a low-pass filter (to prevent switching transients from falsely triggering current limit). One additional resistor connects the disable pin to a 5 V shutdown command that overrides all other analog signals.

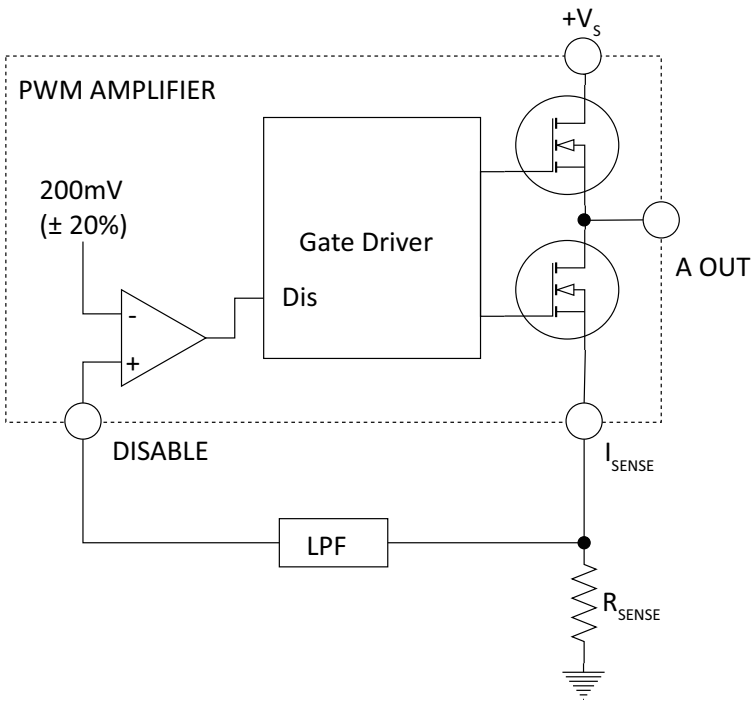
Figure 8: Single-Amplifier Current Limit



This technique will NOT work for a parallel system.

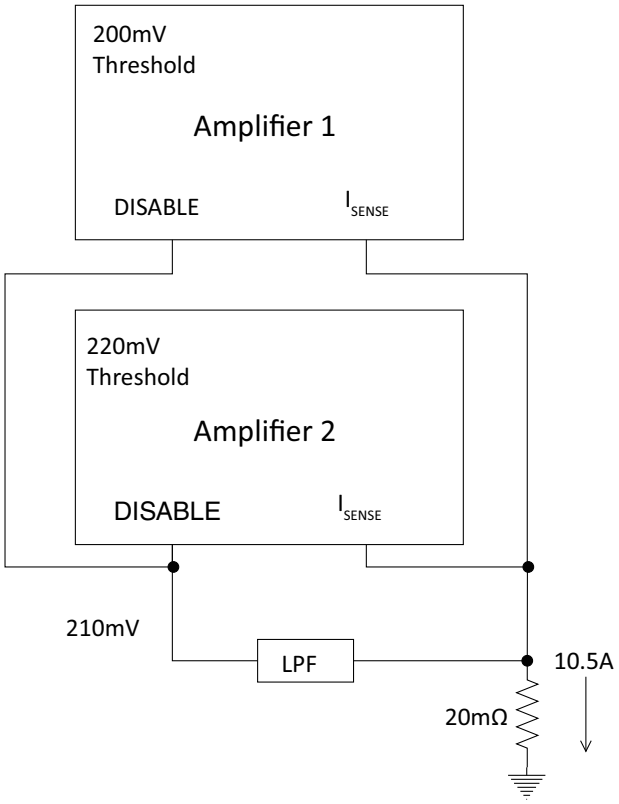
Have a look at the internal schematic of the current limit mechanism (figure 9). There is an internal comparator that compares the sense voltage with some threshold – usually 200 mV. When the sense voltage exceeds this threshold, that signal is level-shifted to 5 V by the comparator, and shutdown is triggered. All outputs go high-impedance until the sense voltage falls back down. The important thing to note is that the voltage at the disable pin never exceeds 200 mV by very much.

Figure 9: Current Limit Internal Circuitry



Now imagine the paralleled system in figure 10. But here's the twist: amplifier 1 has a threshold voltage of 200mV, and amplifier 2 has a threshold voltage of 220mV. The previous section already dictates that the disable pins of all paralleled amplifiers must be shorted together. If this disable "node" is connected to the sense resistor (through the LPF), and the current is high enough to generate 210mV on this node, amplifier 1 goes into shutdown. Amplifier 2 remains in operation, with the full-scale current flowing through it. The voltage at the disable pin stays at 210mV because the shutdown of amplifier 1 does not have much effect on the overall impedance. Amplifier 2 violates its absolute maximum current rating because it never shuts down.

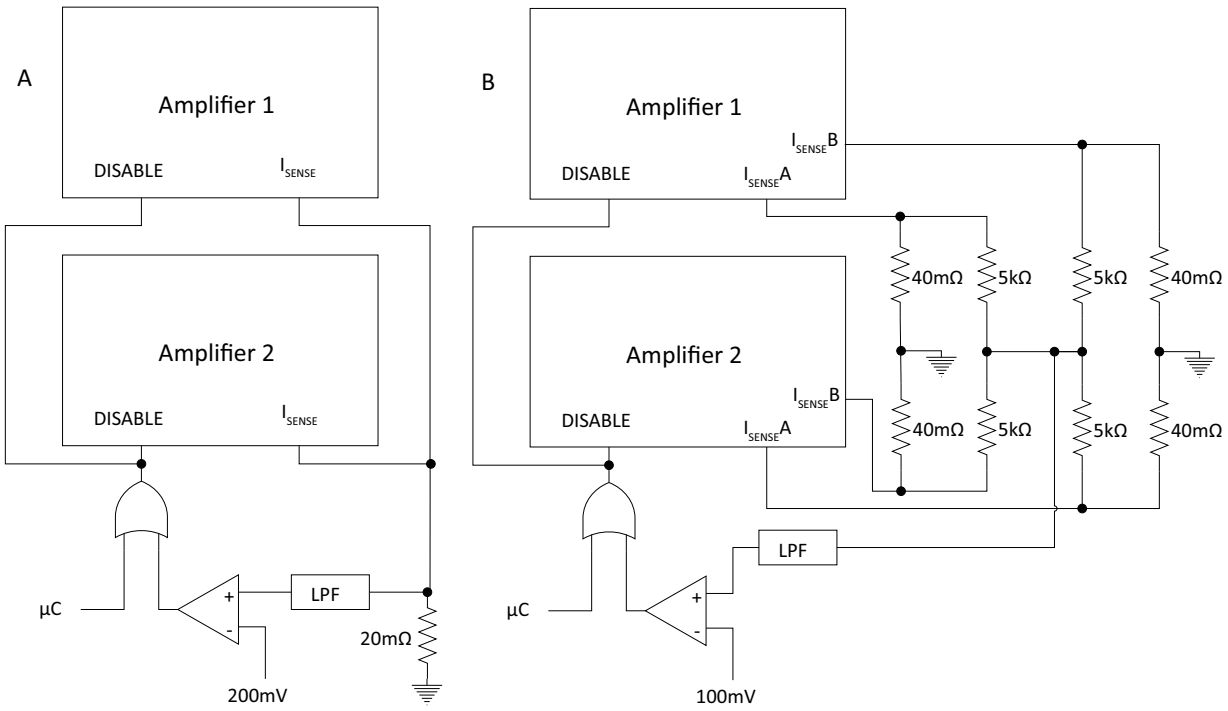
Figure 10: The WRONG way to Limit Current in a Parallel System



Instead, there must be a network that senses current and produces a signal that supersedes any variation in threshold voltages. This can be accomplished by adding a level-shifting comparator to the mix.

The following circuits basically continue the philosophy of digital input mode. The analog voltage after the LPF is converted to a 0V/5V digital signal, just as a microcontroller would provide. This essentially "bypasses" the internal comparator and eliminates the risk of one amplifier shutting down without the other. This concept can also be extended to sense current in each half-bridge independently.

Figure 11: Correct Current Limit Topologies for Parallel Amplifiers. Both are configured for a Current Limit of 10 A total. (A) senses total current in all amplifiers, and (B) senses current in each half-bridge.



SWITCHING FREQUENCY CONSIDERATIONS

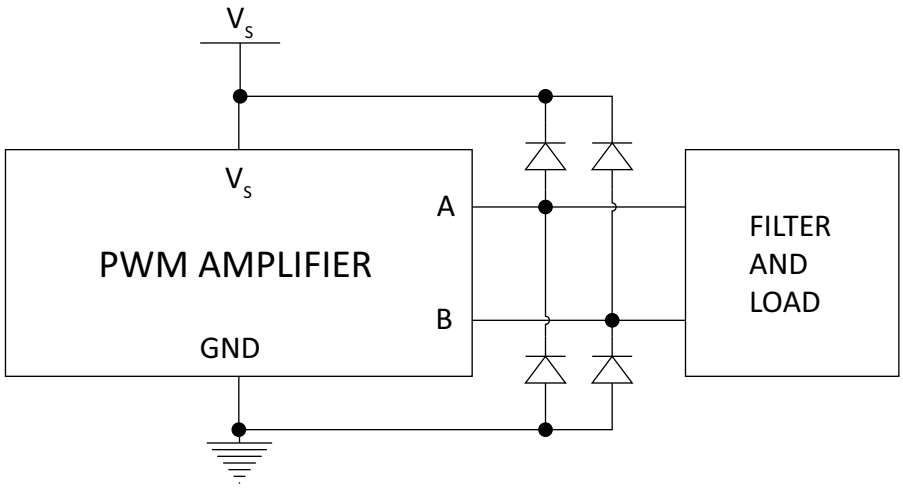
In practice, paralleling amplifiers inevitably causes some MOSFETs to turn on and off at different times. Just as the example in figure 1, this causes excess current to flow into one paralleled FET and violate its current handing ability for a short duration. While this problem can be mitigated with inductance just outside each output, it is also good design practice to minimize the percentage of time during which only one paralleled FET is conducting.

Keeping switching frequencies low can help. For example, a 50ns propagation delay between two paralleled amplifiers, operating at an extremely high F_{SW} of 500kHz, results in 5% of operation time spent in the condition where one paralleled FET is on and the other is off. By reducing the switching frequency to 20kHz, the 50ns propagation delay becomes shorter compared to the switching period, and the fault condition occurs only 0.2% of the time.

EXTERNAL PROTECTION DIODES

External protection diodes are some of the most important considerations when it comes to keeping an Apex switching amplifier alive. While these should always be used, they are even more essential for parallel applications.

Figure 12: External Protection Diodes Must be Used on EACH Amplifier



Recall that the purpose of these diodes is to “snub” inductive voltage spikes so they do not reverse-bias the MOSFETs and damage the parasitic body diode. Even if the load is purely resistive, the inductance in circuit traces and wires is enough to make some considerable voltage spikes when combined with the di/dt associated with Apex switching products.

With two FETs in parallel, it’s likely that inductive loads and filters will force high currents into the body diode with the lowest $V_{FORWARD}$ if no external diodes are provided. This will increase power dissipation in one FET, while the other stays cool (which causes even more mis-match between the two paralleled devices). For this reason, it is crucial to employ a protection diode that has BOTH:

1. A faster reverse recovery time than that of the body diodes. Ultra-fast recovery diodes (such as MUR-type diodes MUR160, MUR440, etc.) are great for this purpose. Schottky Barrier Diodes can perform even faster. Check the datasheet to ensure these diodes have a sufficiently fast reverse recovery, faster than 100 ns.
2. A lower forward voltage than that of the body diodes. All that speed means nothing if the forward voltage is too high for the inductive kick-back to reach it. And remember – forward voltage decreases with temperature. If the conducting FETs are carrying close to their maximum current, you can bet those parasitic body diodes are going to be hot, and their forward voltages will be low. Be sure to consider your operating temperature when you choose these protection diodes.

CIRCUIT LAYOUT

Circuit layout for any high-speed high-power switching circuit can easily dictate if your design performs beautifully, or if it catastrophically explodes in a cathartic rage. AN30 is a great guide for designing PCB lay-

outs for these high-power switching circuits. All of the rules mentioned in AN30 must be entertained in addition to the following guidelines for parallel circuits:

Symmetry is everything! A well-designed parallel PWM amplifier will show repeated elements for all of its amplifiers. Power supply bypass capacitors must be used for EACH amplifier, and they should be in the same position relative to EACH amplifier. The star-point ground should be just as far from one amplifier as it is to any other. For circuits with more than two Apex switching amplifiers, radial symmetry may be needed to accomplish this. Power paths should have the same width and length on each amplifier. If they don't have the same dimensions, the impedance mis-match between amplifiers can be even worse than it already is.

Proximity is another major consideration. All amplifiers need to be as close as possible to each other. Consider the need for matching supply voltages, input signals, and ground references. If the amplifier systems are too far apart, long traces with significant impedances will attenuate signals disproportionately. So the traces must be short and wide - this means a crowded PCB. Once you figure in the size of the filter components, load terminals, bypass capacitors, protection diodes, and the amplifiers themselves, it can be quite a challenge to cram all of these components into an appropriate size. Multiple copper planes can help, but be wary of trace coupling and heat buildup in the high-current planes!

CONFIGURING AN AMPLIFIER IN DIGITAL MODE

The input topologies for Apex switching amplifiers can vary widely. For this reason, changing an analog-mode amplifier into digital-mode is not always the same. The following guidelines give a broad understanding of the process that should be used in tandem with the product datasheet to create a digital-input amplifier.

In analog-mode, an analog voltage is compared with a triangular wave that modulates the input into a proportional PWM duty cycle. In most cases, this input comparator cannot be bypassed, so the *digital input signal must still go into one of the comparator inputs.*

In all cases, the following three pins are reassigned as follows:

1. The triangular wave, or "ramp", will be replaced with the input signal from the microcontroller. This signal tells the amplifier the direction of current (from A OUT to B OUT or vise-versa). See the product datasheet for more information. The amplifier pin is named RAMP, R_{RAMP}IN, or -PWM/RAMP.
2. The analog input voltage will be replaced with a DC reference, typically 1.3 V. Because this pin only connects to an internal comparator input, this DC voltage may have high source impedance (i.e. resistive voltage divider or R-D network). The amplifier pin is named +PWM, +IN, or ANALOG IN.
3. The disable pin will connect to the microcontroller as discussed in previous sections. The amplifier pin is named DISABLE or I_{LIM}/SHDN.

The major difference between amplifier models lies in how to overcome the internally generated ramp signal. Some amplifiers allow for clock detachment, shown in figure 13. Simply omit the typical short between the CLK IN and CLK OUT pins, and the clock will effectively be "detached". Leave the CLK IN and CLK OUT pins unconnected.

Other products, such as SA160, do not allow clock detachment. For these, see figure 14. These products generate the ramp with a high source impedance. A signal with a low source impedance, such as that from a microcontroller, can override the ramp. In other words, no additional considerations are needed other than the three pin-assignments listed above.

Figure 13: Detached Clock Method

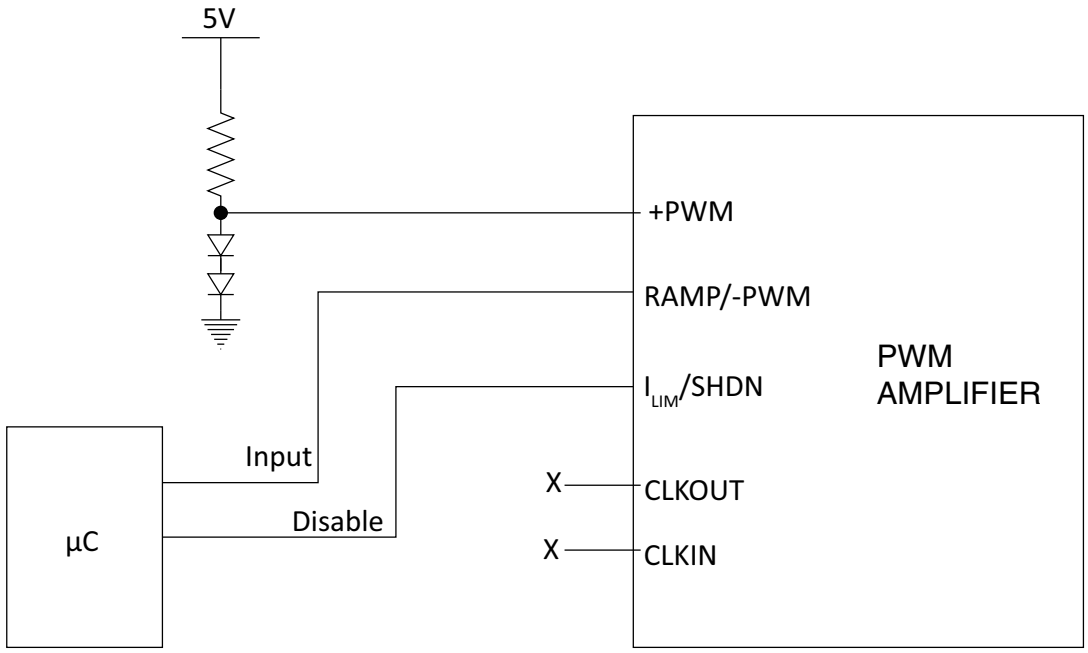
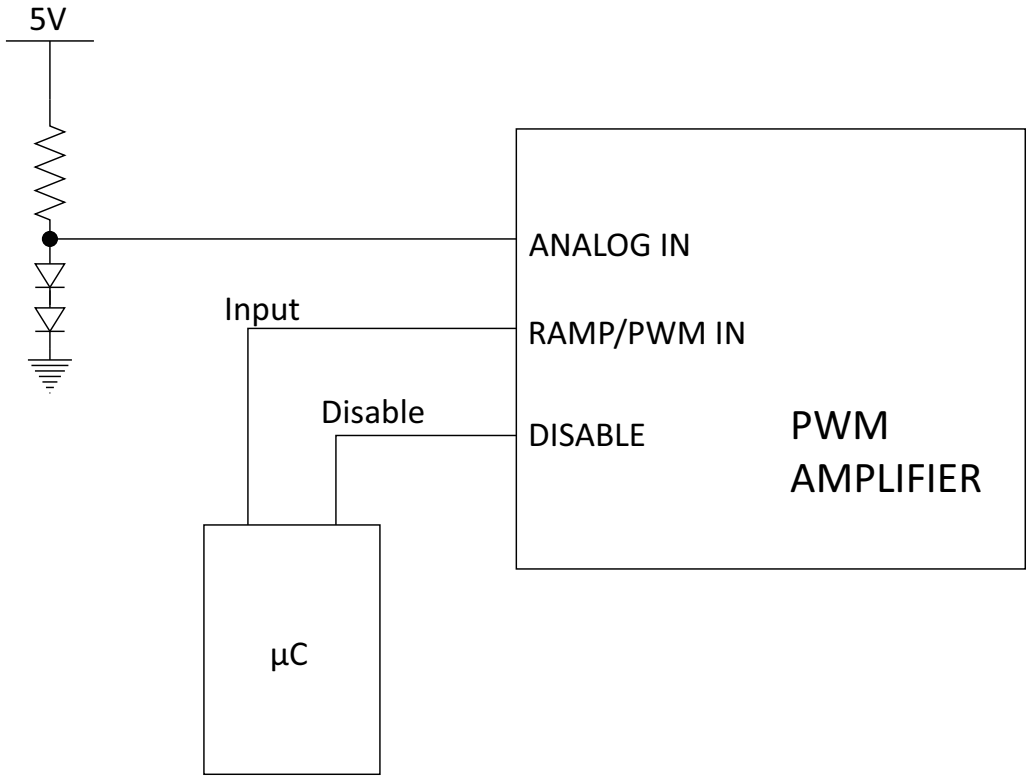


Figure 14: Overridden Clock Method



CONCLUSION

Clearly, the procedures outlined in this note are not for the faint of heart. Be prepared for long simulations, lengthy BOMs, and lots of troubleshooting. Paralleling Apex PWM products is a powerful technique that can multiply currents and increase output power, but it requires diligence.

One last consideration is cost; double the amplifiers means double the price. A single-amplifier system, if one meets the requirements, is almost always more cost-effective than a parallel system. In fact, the SA09 parallel circuit in the design example could have been easily replaced with a single SA160 for less than one fifth of the cost!

These materials are presented to show that PWM products can be safely and reliably paralleled. Paying careful attention to the guidelines listed above, a designer can successfully make Apex PWM amplifiers perform well beyond their limits.