

Reducing Processor and Server Power with Digital Power Conversion

Edward Herbert
September 17, 2006
Canton, CT 06019
<http://digitalpowerconversion.com/>

Reducing Processor and Server Power with Digital Power Conversion

- *The greatest waste of power in a processor or server is the loss due to leakage current when the processor core is idle. By turning off the voltage to the processor core, this power is saved.*
- *A processor or server core cannot be turned off unless it can resume **fast enough** to meet system requirements. For this reason, most processors and servers have power applied continuously.*
- *Only a **true digital power converter** can turn on and off fast enough so that power can be removed whenever the core is idle.*
- *Only a **true digital power converter** can provide individual core voltage control (including on-off) for multi-core processors.*

The [switched current power converter with switched charge circuits](#) is the only *truly digital power converter* topology. The current and voltage are applied and removed very quickly and accurately, as fast as solid-state switches can change state.

- **Turn on:** 0 V, 0 A to any VID, any load: **< 2 us.**
- **Load dump:** Any VID, any load to 0 V, 0 A: **< 1 us.**
- **Voltage transition:** Any VID, any load to any other VID, any load: **< 1 us.**
- **Unconditionally stable:** Voltage regulation is by hysteretic control. Voltage transition is by open loop charge transfer.

[*These transition times are from [SPICE simulations](#) for a switched current power converter with switched charge circuits, using the interposer impedance model of Intel® VR 10.2. The **totem-pole power converter for processors** is expected to be at least **an order of magnitude faster**.*]

Voltage and clock manipulation saves power, maybe, but turning off the core is the only way to save significant power. See "[Do Voltage and Clock Switching Really Save Energy?](#)" (<http://digitalpowerconversion.com>).

Analog Power Conversion:

Present power supplies for processors are analog. Some power supplies are called "digital" because they have a data bus for parameter input and status reporting. Others have a digital processor controlling the analog power conversion function (usually a multi-phase buck converter). These are inherently too slow to enable fast turn-on and turn-off of the processor or server core voltage.

Problems and solutions:

- **The problem:** Present analog power supplies change current by applying voltage to an inductor and waiting for the current to ramp, a very slow process.

The solution: The switched current power converter changes current in digital increments, with digital switching, nearly instantaneously.

- **The problem:** Present analog power supplies increase voltage by increasing the current in an inductor to provide more charge to a capacitor, a very slow process. Decreasing the voltage is even more problematical.

The solution: The switched charge circuits step the output voltage digitally by adding charge to the output capacitors in precise increments, resulting in very precise and very fast steps in the output voltage, up or down. While it takes finite time to transfer the charge, it is orders of magnitude faster than present analog power supplies. The voltage can be reduced to zero just as fast, even with a simultaneous load dump.

- **The problem:** A power converter with a large output capacitance cannot provide the higher voltage needed to drive a high di/dt through the parasitic inductance of the interposer.

The Solution: The switched current power converter uses a current source to the drive voltage to overcome the parasitic inductance

- Using the "Totem-pole power converter for processors", there is no di/dt in the interposer.
- Using the switched current power converter with switched charge circuits as an external VRD, and using total charge measurement, a large but controlled voltage drives the di/dt through the interposer inductance. The effects of the parasitic inductance are largely attenuated.

- **The problem:** With a power converter that is a voltage source, the energy in the power distribution changes wildly with large di/dt as the load current changes. The energy change is $\frac{1}{2} (I_1^2 - I_2^2) L$, where L is the parasitic inductance.

The solution: Using the totem-pole power converter, the energy in the power distribution bus hardly changes at all with load changes, as the current in the bus is constant. The bus voltage changes with load, but the energy change is small.

- **The problem:** Present power supplies use voltage and current measurement as their control inputs.
 - Voltage measurement is not suitable as a control input for a fast power converter, as it takes too long to settle down following a transient.
 - Current measurement is even slower, and the circuits are complicated.

The solution: The switched current power converter is controlled by measuring the [total charge](#) on the distributed capacitors. No current measurement is used.

- If the total charge is constant, the output voltage is also constant, with a slight decrease with increasing load (much less than presently used).
- Maintaining constant charge forces the source current and the load current to be equal.

Digital Power Conversion

- **Digital current control for voltage regulation:** To regulate voltage, the current is switched digitally in small increments. The current changes in response to load changes as fast as MOSFETs can switch on or off.
- **Digital charge transfer for voltage step change:** To step the voltage, charge is transferred digitally. While it takes finite time to transfer the charge, it is orders of magnitude faster than analog voltage transitions.
- **Individual core voltage control for multi-core processors:** Unlike a multi-phase buck converter, there is no penalty and great benefit in dividing the current to be controlled among many smaller switches. One benefit is that the voltage of the cores in a multi-core processor can be individually controlled. While more voltage measurement circuits are needed, the power switching is simplified as the currents are much smaller. A master voltage reference is used.
- **Extremely fast and unconditionally stable:** Voltage regulation is by hysteretic control. Voltage stepping is by open loop charge transfer. As long as the output capacitor is above its critical value, the control is unconditionally stable
- **Output capacitance is very much smaller:** The output capacitor must be large enough so that the voltage does not overshoot during the propagation delay of the digital logic. Because the control is so fast, this is very much smaller than the capacitor needed for an analog control.

Power Control in the Processor or Server

Parasitic inductance: The **parasitic impedance** of the interposer for a processor or server is significant. SPICE models using the impedance model from Intel® VR 10.2 suggests that a break frequency above 5 MHz is not possible with an external power converter. If, however, part of the power converter is on the processor or server die, or in the same package, the **parasitic inductance** of the interposer has a stabilizing influence, and **it becomes part of the solution**.

Adding losses to save power???: It is counterintuitive that including part of the power converter, with its losses, in the processor or server package actually saves power, but that is the case. This is because *in this location* the power converter can be **fast enough** to turn on and turn off in **a fraction of a microsecond**. Each core can be turned on when needed and turned off when it is not. Most servers operate at a low duty-cycle most of the time. See: [Taurus: A Taxonomy of the Actual Utilization of Real UNIX and Windows Servers](#), David G Heap, Principal IT Consultant, IBM Enterprise Server Group, Somers, NY, USA (pdf, 7.9M).

US Patent 7,098,638, "**Totem-Pole Power Converter for Processors**," shows that the switching cells for a totem-pole power converter are very similar to a totem-pole digital driver, so the power control can be easily implemented on the processor or server die. Alternatively, the power control can be on a separate power co-processor adjacent to the processor or server die.

Dividing the current: Present multi-phase buck converters use four or six phases or perhaps a few more. There are advantages to having more phases, but the penalty in complexity of adding more phases to a buck converter is very high, with diminishing returns. That is not the case with the switched current power converter.

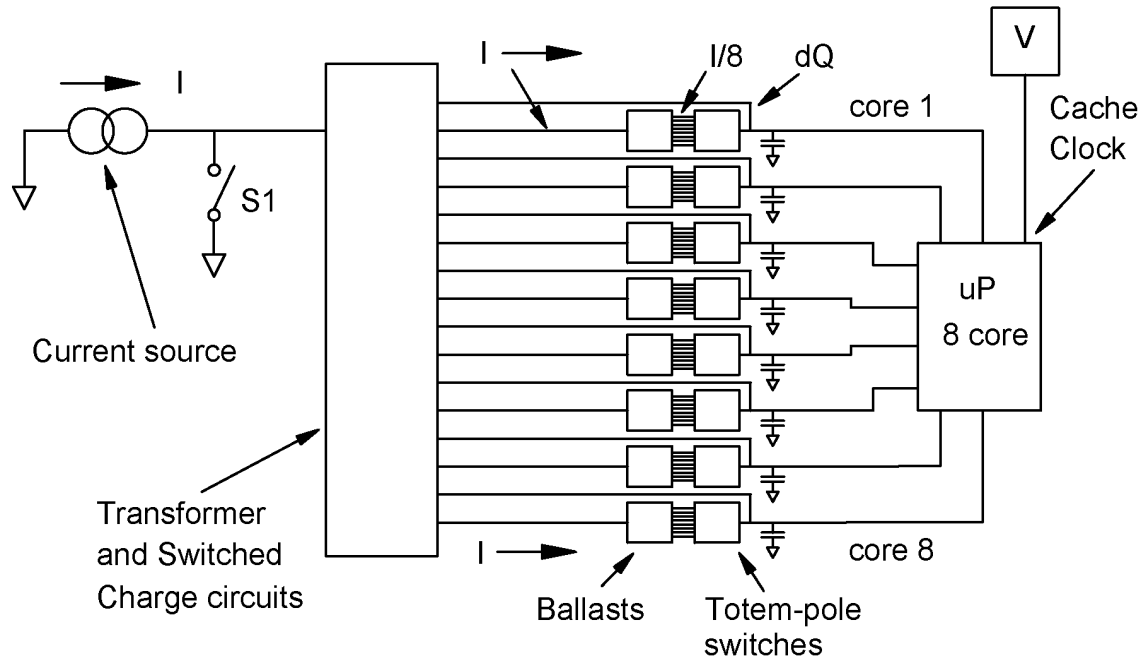
The total silicon area for controlling a certain maximum current is largely determined by the current density on the die. If divided, there may be more switches, but each is smaller, so that even a very large division of the current costs little in silicon area. That is true of the switch drivers as well. There may be more drivers, but each is smaller. As the switches and their drivers get smaller, they also are much faster and they can be driven directly with digital logic circuits with much less buffering.

As the current increments are reduced (greater division) and the switches become faster (smaller switches), the dynamic response gets faster and the size of the output capacitor can be reduced. In the limit, the current can be divided among the available pins in the processor package, which is a large number. Dividing the current first on a per core basis and then again on a per pin basis may be optimum.

Individual core voltage control: Individual voltage control for the cores of the processor actually simplifies the power switching circuits, as the individual currents are much smaller. A voltage measurement is needed for each core, but one master reference voltage is sufficient.

Circuit Example:

Power Converter for 8 Core Processor



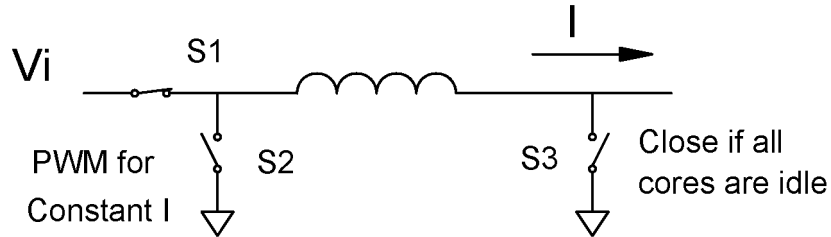
Application example: As an example, let us consider a power converter for an **eight-core processor** for a game console or a video display driver. Its task is to refresh the display at an 85 Hz rate, or every 11.8 ms. Let's further assume that the eight cores are somewhat interactive, occasionally having to wait for data from another core. Once their task is finished, they can be turned off. The core voltage goes to 0 V, and the wasted power due to leakage currents is saved.

The 11.8 ms cycle starts by turning on all of the processor cores and outputting the screen refresh data from the previous cycle. Then, each core begins its current task, generating the next screen refresh data. While it is active, the core voltage will be maximum (turbo mode) per the VID input. If the task is less intensive, or if a core is waiting for data that will arrive quickly, its voltage can be reduced (normal mode) to save power. When its task is over, or if the wait state will be longer, the core voltage can go to zero, completely eliminating the wasted power of the leakage current.

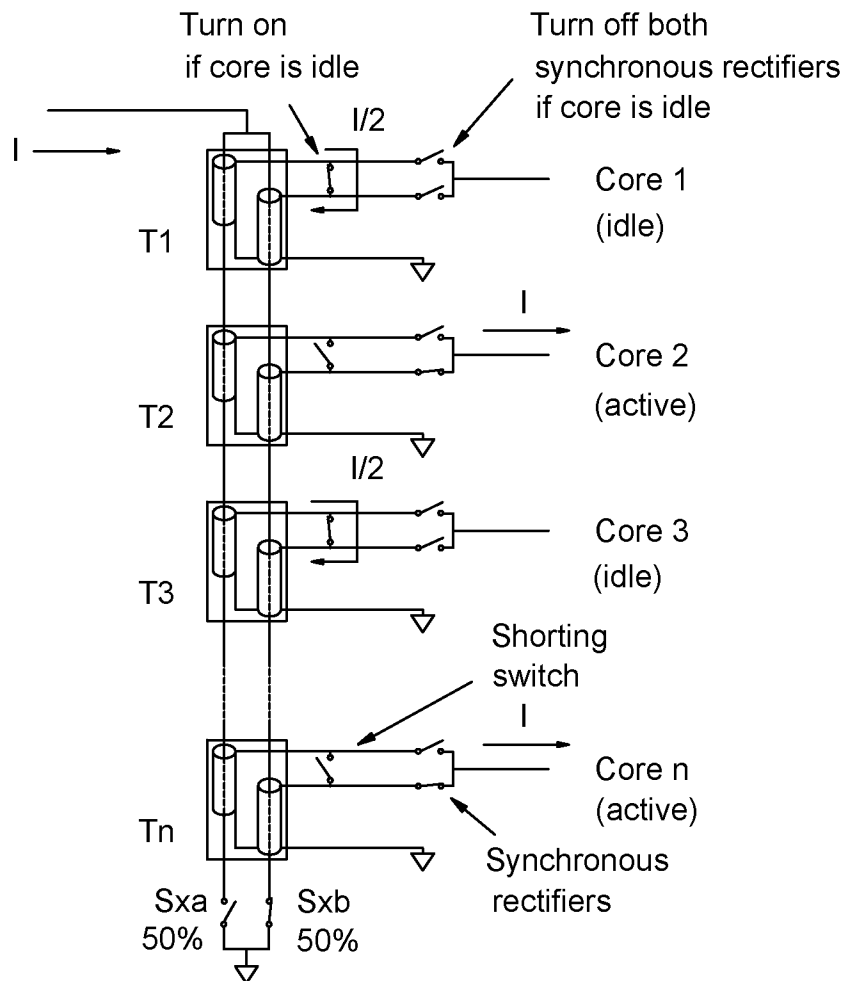
When a core is active and its voltage is being regulated, there are losses in the on-board power converter. However, once a core's task is done, the current source feeding that voltage channel can be turned off at the transformer. **The power loss for that channel goes to zero** in the processor package. When all of the cores have completed their tasks, the current can be turned off at the transformer, eliminating the losses there as well.

Many processors and servers are idle for 80 percent of the time or more. The small penalty of incorporating part of the power converter in the processor or server package

may increase the losses 10 percent or so, *when on*, but *if the processor or server is idle for more than 10 percent of the time, there is a net power savings*. If the processor or server is idle for 80 percent of the time, *the net power reduction is over 70 percent*.



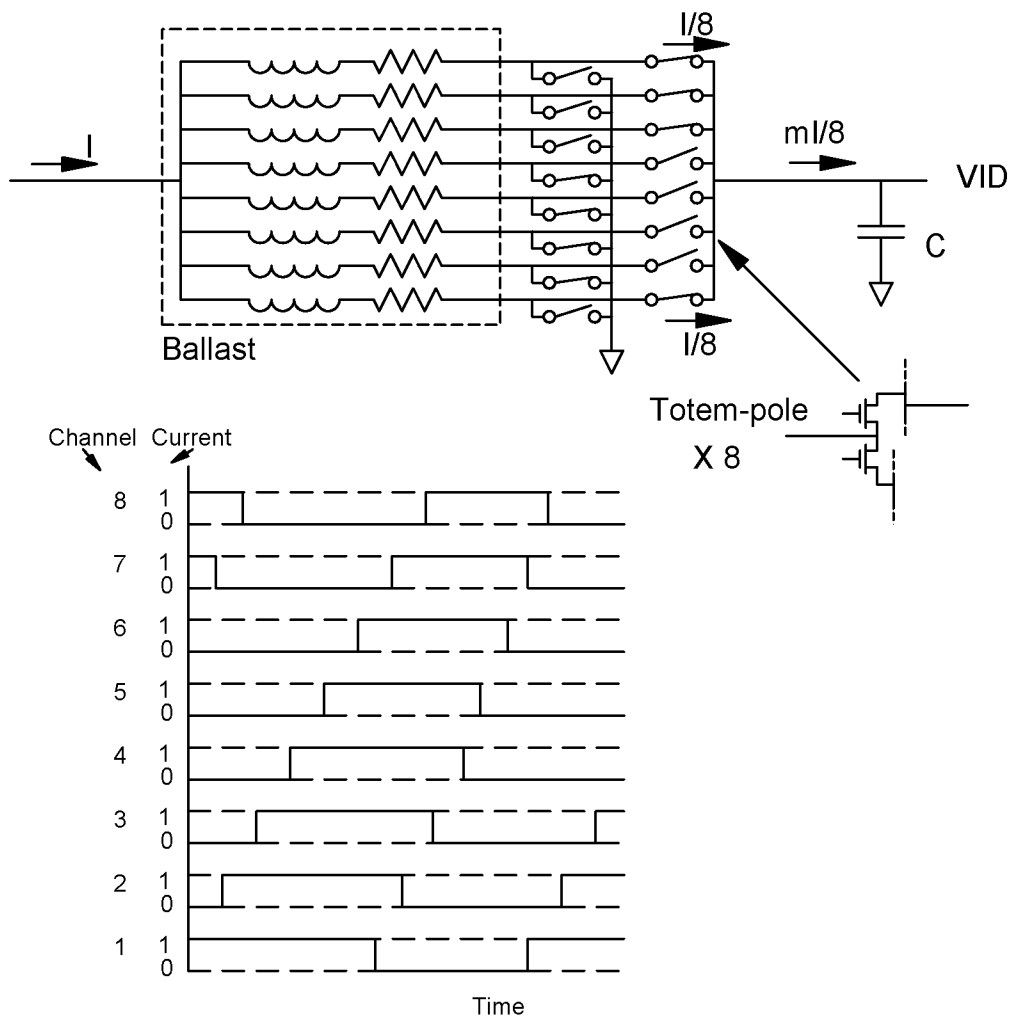
Power source: The power source for the switched current power converter is a constant dc current source. ("Constant current" does not mean that the current never varies, as it may be changed for different modes of operation). The constant current source may be a buck converter, controlled for constant current, with no output capacitor. It can be powered from a non-isolated source, like a 48 V dc bus.



If no current is needed at the output (all cores are idle), the dc current source may be shorted to return. Then, the only losses are those due to the dc resistance of the buck inductor and the switch.

Multiple currents: The primary constant current source feeds the primary of a multi-section matrix transformer with 100 percent duty-cycle push-pull switching. The rectified outputs provide multiple parallel constant current sources for the switched current power converter. In our example, there are eight outputs, one for each core.

If a particular secondary current is not needed at the output (its core is idle), that section of the matrix transformer is turned off by turning off the synchronous rectifier and shorting the secondary winding.



Further division of the currents: Present multi-phase buck converters use four or six phases or perhaps a few more. There are advantages to having more phases, but the penalty in complexity of adding more phases to a buck converter is very high, with diminishing returns. That is not the case with the switched current power converter.

The total silicon area for controlling a certain maximum load is largely determined by the current density on the die. If divided, there may be more switches, but each is smaller, so that even a very large division of the current costs little in silicon area. That is true of the switch drivers as well. There may be more drivers, but each is smaller. As the switches and their drivers get smaller, they also are much faster and they can be driven directly with digital logic circuits with much less buffering.

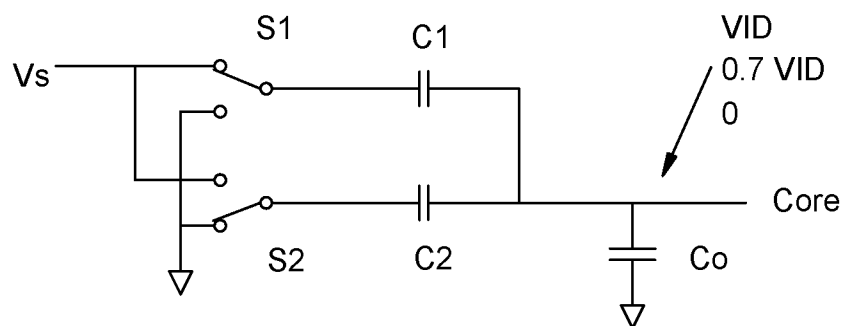
The recently issued patent, *"Totem-Pole Power Converter for Processors,"* (Link: [US 7,098,638](#)), shows how to use passive ballasting to divide a constant current into a large number of smaller constant currents. The ballasting is very small parallel inductors, and the self-inductance of parallel circuit board traces is sufficient for most applications. In our example, there are 64 small totem-pole current switches, eight for each of the eight cores. A housekeeping supply maintains the cache and clock.

Unlike in a multi-phase buck converter, the currents are not measured. If the resistances are nominally equal, and the timing of the totem-pole switches maintains approximately equal duty-cycles on the parallel lines, the currents will divide well enough. The timing diagram shows that if more current is needed, the switch that has been at ground the longest will switch its current increment to the load. If less current is needed, the switch that has been connected to the load the longest will switch to ground. This arrangement tends to correct any current imbalance.

Multiple output voltages: By dividing the current into a very large number of small constant currents, a number of individually controlled output voltages is easily accommodated. This actually simplifies the power switching because the current is smaller. As an example, the eight-core processor or server may have independent voltage control of the eight cores. Each core may be turned off when idle.

Switched charge circuits: Switched charge circuits can be a binary digital array, and if designed that way, any VID voltage can be applied. Changing the digital VID input causes the voltage to switch to the new VID in under 1 μ s.

For our example, a simplified design has two charge switches so that there are three states, VID_{max} (Turbo mode), $0.7 VID_{max}$ (normal mode) and 0 volts (off).



V_s may be in the order of 12 V, and it can be varied to control the size of the voltage steps to have the correct VID_{max} for the processor. If S_1 is switched from ground to V_s , a

fixed charge is added to the output capacitor C_o very quickly, producing a very fast and accurate step in the voltage, for example, from 0 to $0.7 V_{ID_{max}}$. If the switch S_2 is then switched to V_s , the output steps again, to $V_{ID_{max}}$. If both switches were switched to V_s initially, the voltage steps directly to $V_{ID_{max}}$.

If both switches are returned to ground, the charge is removed just as fast, and the voltage returns to 0 V. The switched charge circuits and the charge transfer capacitors C_1 and C_2 may be located with the matrix transformer, and a single line per channel can transfer the charge.

The switched charge circuit has no ability to *regulate* the output voltage. It only provides fast and accurate steps. Thereafter, the voltage regulation is maintained by slight modulation of the current control. Accordingly, the voltage reference for the current control must be changed simultaneously with the voltage step command.

There is some loss of energy in transferring charge. See [Switched Current Power Converters \(pdf\)](#) page 41. This loss is directly proportional to the size of the output capacitor C_o , so it is desirable to minimize this capacitor. The small increments of the switched current and the high speed of the totem-pole switches allow this capacitor to be very small.