

A Completely On-Chip Voltage Regulation Technique For Low Power Digital Circuits

L. Richard Carley
Carnegie Mellon University
Dept. of Electrical and Comp. Eng.
Pittsburgh, PA 15213 USA
1-412-268-3597
carley@ece.cmu.edu

Akshay Aggarwal,
Lucent Technologies
1247 S. Cedar Crest Blvd.
Allentown, PA 18103 USA
akshay.aggarwal@lucent.com

Abstract

This paper describes a completely on-chip voltage regulation technique for locally generating an adaptive low voltage power supply rail from a given higher voltage power supply without requiring any external component. The on-chip regulator, based on delay servoing, primarily comprises of a critical path replica, charge pump and a high performance voltage buffer which is the most critical component of the design. Simulation results in 0.5 μ m CMOS process demonstrate that the buffer offers a low DC output impedance, a high degree of voltage regulation (output ripple of 12% of Vdd) and a superior line regulation (upto the maximum clock frequency of 50MHz) even under strongly varying load conditions. The regulator response for a typical worst case load exhibits a maximum voltage fluctuation of 4% of Vdd with a reasonably fast response time.

1. Introduction

The use of multiple supply voltages on a digital logic chip can lower the power dissipation by allowing the operation of non-critical gates at reduced voltages [1]. Further, maximum power savings in these non-critical path gates can be achieved while guaranteeing that delay specifications will be met, even in the face of manufacturing and operating range variations, by adjusting the degree of voltage regulation in response to an on-chip measurement of the current speed of the device. In principle, maximizing the overall power savings for a complete system would require a different optimal power supply voltage for every block. In order to approach this goal, we have chosen to implement completely on-chip series voltage regulators, in spite of the fact that their efficiency can be lower than that of switching regulators making use of off-chip inductors.

The first component of an on-chip voltage supply is the method for predicting the worst-case delay of a block. Two standard approaches to implementing this function are (1) build a replica of the critical path circuitry, or (2) take the actual logic block out of service and apply the worst-case pattern sequence in order to

measure the delay exactly. Although (2) is more accurate, it imposes significant constraints on how the logic block can be used at the system level. Therefore, (1) will be used in this paper and enough margin will be incorporated into the critical path replica circuit to account for a worst-case mismatch between the actual logic block replica circuit and the replica circuit.

Using the critical path replica circuit and an appropriate servo loop, the voltage at the output of the adaptive voltage regulator is then adjusted so as to minimize power dissipation while guaranteeing that the delay requirements of the logic block are always satisfied. There has been a great deal of prior work on the design of efficient on-chip power supplies for digital logic; e.g., The Digital Power Supply Controller [2], Variable Voltage Supply Scheme (VS) [3], and the Digital PWM Power Controller [4]. All four of these examples are high efficiency DC-to-DC switching converters requiring off-chip components (L,C). In contrast, we describe a *completely* on-chip voltage regulation technique which locally generates the low voltage power supply rails from the given higher voltage power supply rails without requiring *any* external components. The switching DC-DC converter function is replaced by a series pass transistor that is part of a high performance voltage buffer that is able to provide a high degree of voltage regulation under strongly varying load conditions. In addition the proposed series regulator incorporates a novel sleep-mode control feature whereby the standby power for multiple supply systems without level converters (e.g., [5]) is minimized. This concept has been proposed by us earlier [9] and in the present paper a complete description of the entire circuitry involved along with detailed simulation results is presented.

Due to the power loss in the DC series path, the dynamic power savings in a completely on-chip regulator based system instead of being near quadratic in nature tends to be linear. However, the nearly cubic reduction in the short-circuit power is still retained, and hence the series loss doesn't cause as severe reduction in the power savings as compared to a switching regulator as one might expect. In addition, there are switching losses and overheads associated with switching regulators; their typical efficiencies are 85%-95%. For applications in which a 3V power supply is being regulated down to 2V using a series regulator, the efficiency is still about 65%. And, as mentioned above, it is difficult to provide 10 different power supply voltages when external components are needed for each voltage regulator. However, with a series regulator scheme, any number of on-chip power supply voltages could be generated right where they are needed (saving on wiring). For example, in a four power rail based methodology like *QuadRail* [5] where different units of the system could be operating at different internal swings, if one were to use a single switching regulator for economic reasons, the power savings in the system might be undertapped. On the other hand one could use various on chip regulators for specific modules thereby maximizing the power savings. The additional power savings so obtained might actually more than compensate for the series loss. This promises to be very attractive not only from the system integration point of view but also from the point of view of minimizing the total implementation cost.

2. Servo Loop Scheme

2.1 Voltage Regulation Techniques

The three scenarios which broadly represent the variety of ways in which the load configuration might occur for a regulated circuit are shown in Fig. 1. In Fig. 1(a) where the regulator needs to *source*

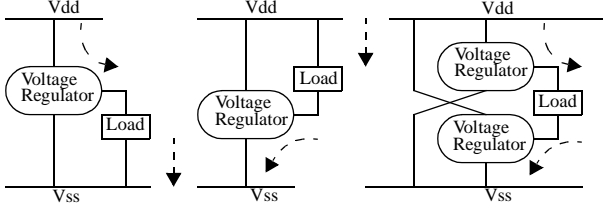


Figure 1: (a) Upper Rail Regulator (b) Lower Rail Regulator (c) Dual Rail Regulator

the varying load current and hence derives the output voltage from the upper rail may be termed as Upper Rail Regulator. Such a scheme is implemented in [3] using a switching DC-DC converter. An interesting combination of 1(a) and 1(b) is seen in 1(c) where one would require two independent regulators to generate the two inner rails. An important feature of any such on chip regulator would be to dynamically source/sink varying load current without suffering a large DC drop across the source/sink as this would limit the available output voltage. This would require high performance VCVS buffers. In fact as we shall soon see, the design of such a buffer is the most challenging part of an on-chip regulator.

2.2 Output Buffers

There are a few essential requirements that an output buffer driving a dynamically varying heavy load must satisfy. (1)Low output impedance for all load demands. (2)High degree of load regulation. (3)Superior line regulation (high PSRR) and (4)Fast response time. The buffer presented here satisfies all of these requirements. Fig. 2 shows our buffer circuit which we use in an Upper Rail Regulator. The buffer topology that we propose is a two stage symmetric CMOS OTA. The dynamically varying load current has been modeled as I_{load} . C_d represents the large load capacitance and the on chip decoupling capacitor. Transistors M3,M4 and M7 mirror the signal current to the outer leg formed by M6 and M8 and the summation is done at node V_{int} . Because of the particular topology that we have, the node V_{int} is able to swing all the way from $V_{ss}+(V_{eff})_{M8}$ to $V_{dd}-(V_{eff})_{M6}$ before the devices enter the linear region. For commonly used values of V_{eff} node V_{int} has almost rail-to-rail swing. This is a significant advantage in this kind of an application as the gate of the output transistor will need to modulate heavily in order to source widely varying load currents. M9 forms the series lossy path between V_{dd} and the regulated output voltage and thus is a critical component of the design. M9 needs to be large in size so that it has a relatively low V_{eff} and thus offers a low DC output impedance. Under extreme loading conditions when M9 is forced into linear region, its gate is able to go all the way to $V_{ss}+(V_{eff})_{M8}$ (if required) thus satisfying the load demand. Since the buffer is desired to operate for V_{dd} 's as low as 1.5V this high swinging requirement at the V_{int} node becomes imperative. M10 is used for improved PSRR performance [6]. However, a more subtle advantage of M10 comes during load regulation. When V_{out} goes down rapidly in response to a sudden increase in load current, the compensation capacitor C_c causes node V_c to fall at nearly the same rate. Due to this current in M10 increases rapidly thus discharging V_{int} at a faster rate, which prevents the output voltage from dropping further.

The response of the buffer for an input voltage step is shown in Fig. 3. The output voltage can swing all the way to V_{dd} when there is no loading. With a heavy loading condition, there is a small DC drop across M9 and hence the swing is limited to about 90% of

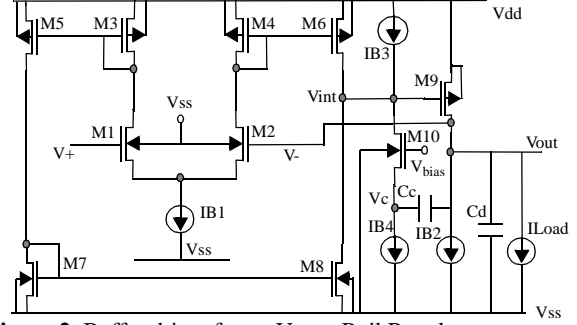


Figure 2: Buffer driver for an Upper Rail Regulator

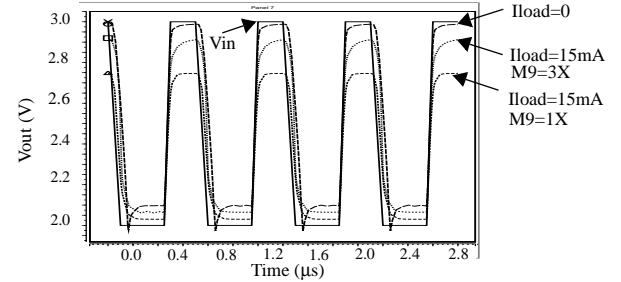


Figure 3: Buffer Transfer Characteristics. ($V_{dd}=3V$)

V_{dd} . However, if M9 were to be made 3X, the available output swing is more than 95% of V_{dd} . As can be seen the response of the buffer is quite fast. This circuit was tested with the output load stimulus that is the worst case scenario for a large digital circuit where *every* circuit switches at the *same* time thus causing large current glitches. The peak current (12mA) value is about 75 times the bias current in the output stage. Under these circumstances, the peak-to-peak ripple in the output load voltage is about 12% of V_{dd} . As the glitch width increases (different circuits switching at different times), the output ripple decreases at a faster rate. The buffer has a very good high frequency PSRR performance as can be seen from Table 1. At 50MHz (f_{clk} for $V_{dd}=2V$ is 25MHz) for a target output voltage of 1.75V and a heavy output load of 10mA, the noise coupling from V_{dd} to the output is only around a factor of 0.1. For a typical maximum peak ripple of 0.2V on V_{dd} , noise at the output, V_{out} , is only around 20mV.

$V_o(V) / I_{load}(mA)$	PSRR For Different Frequencies (dB)			
	1kHz	1MHz	10MHz	50MHz
1.75 / 0.0	-32.4	-32.0	-25.0	-16.0
1.75 / 10.0	-18.0	-18.0	-16.1	-10.0

Table 1: V_{dd} PSRR for the Buffer ($V_{dd}=2V$)

2.3 Delay Servoing

Having realized some of the important design issues for the buffer we now concentrate on the voltage signal that the buffer receives as an input. Any Servo Loop based regulator could have control signals from several possible sources; specifically (1) A Leakage Current Sensing Loop (e.g., [7]) (2) A Delay Locked Loop (e.g. [2],[3]). Since the main goal for any design is to meet the target speed under all process/temperature/voltage variations, a DLL based regulator is more common; hence, we adopt this approach.

Delay Servoing is done based on the performance of the system's critical path replica (CPR) which has some slack as a safety margin. The dynamic input to the CPR is $f_{clk}/2$ (Fig. 4) where f_{clk} is the target clock frequency. The logic values for the other static inputs are such that every rising/falling edge of the dynamic input is equivalent to applying the worst case test sequence which ensures toggling of every node in the CPR. If the correct output

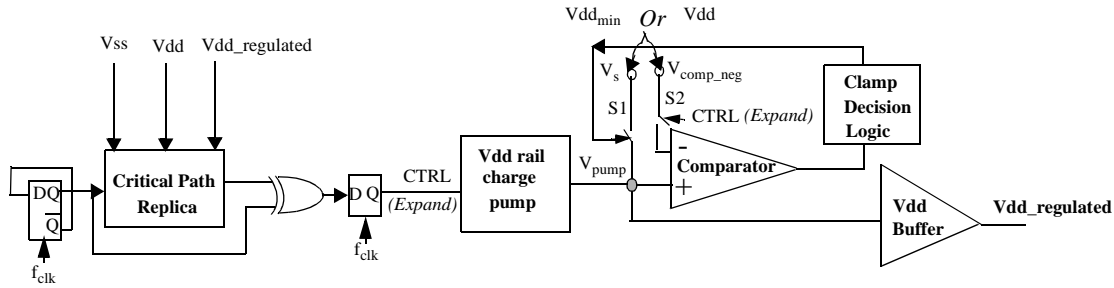


Figure 4: Implementation of the Delay Servoing based Completely On-chip Voltage Regulator

transition of the CPR is of the same polarity (inverse polarity) as its input transition, an Exclusive-OR (Exclusive-NOR) of CPR's output with its input would give a logic 0 or a logic 1 respectively depending on the fact whether it meets the target speed or not. The output of the ExOR gate acts as a control signal to the charge pump which appropriately adjusts the regulated voltage. In our scheme of regulation, the initial value for the output regulated voltage is V_{dd} and this value is decreased adaptively until the point where the CPR fails to meet the target speed. By doing this we ensure that the Servo Loop locks on to the *first* false state (local minima). Note that if the loop is not guaranteed to lock on to the first false state, one would require to implement a system like [3] where the CPR is given some excess clock cycles for its output to reach the correct value before a comparison is made. Also in our scheme since we decide to use only a single CPR to reduce the power overhead, the output voltage has a ripple about the locked value due to the absence of a "dead zone". However, by controlling the resolution of the charge pump this ripple can be kept well below the load induced variations.

3. Servo Loop Response

The response of the Servo Loop is tested by loading it with a large digital circuit that has a high switching activity. In this case we used a parallel combination of 16 subcircuits (Booth Encoder + Booth Multiplexer + 6 Series connected Carry Save Adder's) each of which represents the deepest slice of a 16×16 bit Wallace Tree Multiplier. The dynamic current demand of this circuit is shown in Fig. 5. As it can be seen, the load consists of large current spikes (about 8.5mA) with extremely small rise and fall times and thus poses severe demands on the load regulation capability of the loop. From the response of the Servo Loop in Fig. 6 it can be seen that as per the process corner the regulated voltage is such that the load circuit consumes minimum possible power while always being guaranteed to operate at the target clock frequency [9]. The maximum ripple (at 3 μ S) in the regulated voltage due to load variation is about 80mV which is about 4% of V_{dd} .

The output reaches its final value within 110 clock cycles for the FNFP (fast) case and within 60 cycles for the SNSP (slow) case. Note that since the output reaches its final value starting from V_{dd} , the load circuit is always guaranteed to meet the target speed even

during this transition period. The maximum ripple in the SNSP case in the locked state is about 75mV. Note that by decreasing the resolution of the charge pump this ripple value can be made as low as desired at the expense of longer time for the output to reach its final value.

4. Conclusions

We have proposed a *completely* on-chip voltage regulation scheme which minimizes the power consumption in a digital circuit operating on multiple voltages while always guaranteeing target delay. The output buffer has been shown to exhibit good performance even for a strongly varying load. Simulation results for the regulator for a worst case load demonstrate a high degree of voltage regulation (output ripple of 4% of V_{dd}). Finally, the absence of any off-chip component makes this technique very attractive for multiple use in a complex system where every module might need a different optimal power supply voltage for maximum overall power savings. Presently efforts are ongoing to get the circuit fabricated in 0.5 μ m CMOS process.

5. Acknowledgements

This work was funded in part by DARPA under Order A564, NSF under Grant MIP9408457, and SRC under Contract 068.007.

6. References

- [1] K.Usami, M.Horowitz, *Proc. ISLPD*, April 1995, pp. 3-8.
- [2] Gu-YeonWei, M.Horowitz, *Proc. ISLPED*, August 1996, pp. 313-317.
- [3] T. Kurado et al, *Proc. IEEE JSSC*, March 1998, pp.454-462.
- [4] A.Dancy,A.Chandrakasan, *Proc. ISLPED*, August 1998, pp. 191-196.
- [5] R.Krishnamurthy,Schmit, Carley, *Proc. CICC*, May 1998, pp. 499-502
- [6] B.K.Ahuja, *Proc. IEEE JSSC*, December 1993, pp. 629-633.
- [7] T.Kobayashi, T.Sakurai, *Proc. CICC*, May 1994, pp. 271-274.
- [8] J.B.Burr, J. Shott, *Proc. ISSCC*, February 1994, pp. 84-85
- [9] L.R.Carley, A.Aggarwal, R.Krishnamurthy, *Proc. ISLPED*, Aug 1998, pp 106-108.

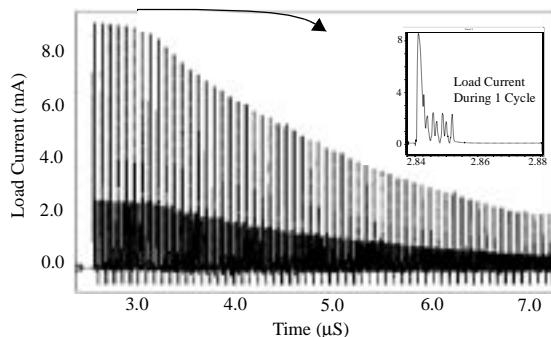


Figure 5: Current Demand for the Load Circuit

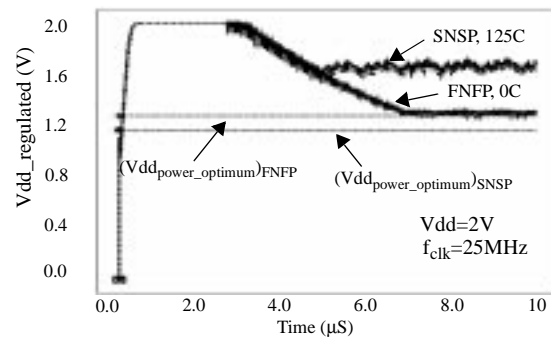


Figure 6: Servo Loop Response