

An Adaptive On-Chip Voltage Regulation Technique for Low-Power Applications

Nicola Dragone^{*}, Akshay Aggarwal, L. Richard Carley

Department of Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA 15213

^{*} The author is also with STMicroelectronics

Abstract

In this paper we present a completely on-chip voltage regulation technique which promises to adjust the degree of voltage regulation in a digital logic chip in the face of process induced delay variations so as to minimize energy dissipation while always guaranteeing the target operating frequency. For this purpose the delay of a critical path replica of the circuit being regulated is constantly compared with the target delay to provide the regulator with the information needed to select the optimum voltage levels. The proposed solution is even more attractive in that no external components are required. Based on this scheme, a completely on-chip voltage regulator has been fabricated in a commercial 0.5 μ m CMOS process and used to generate the inner rail voltages for a DSP multiplier-accumulator (MAC) implemented in mixed swing QuadRail. Measured results indicate that the voltages generated by the regulator offer a very high degree of load regulation thus verifying the fast response time of the on-chip output buffer.

I. INTRODUCTION

Voltage scaling represents one of the most effective means of power reduction because of its quadratic influence on dynamic power consumption. However lower supply voltages also lead to reduced drive currents and increased gate delays and an indiscriminated voltage scaling would end up degrading the overall system performance. On the other hand, in a complex circuit, only few groups of gates determine the maximum operating frequency. For this reason, a Multiple Supply Voltage (MSV) approach can be used to limit power dissipation by operating only non-critical gates at reduced voltage while guaranteeing that the delay specifications will be met.

A great source of concern with such a methodology, which

clearly aims at trading delay slack for power reduction, is represented by manufacturing and operating range variations. In fact, although in a conventional design employing a single power supply voltage only few critical paths suffer from process variations, in a MSV solution the number of such critical paths can increase dramatically since, in principle, maximum power savings can be achieved when each path is operated at zero delay slack.

However, the crucial problem in a MSV technique is the very generation of multiple voltages. In spite of their high efficiency, switching regulators making use of off-chip passive components may not be the best solution, especially when many voltage levels are required. A more attractive solution is represented by completely on-chip series regulators. In complex systems (i.e. SoC), we could even think of adopting various on-chip regulators for specific modules thereby maximizing the power savings. In addition, much of the savings realized from voltage scaling is a dramatic reduction of short-circuit currents which scale roughly as power supply voltage cubed. 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 systems integration point of view but also from that of minimizing the total implementation cost.

In this paper we propose a realistic IC implementation of the regulation scheme presented in [1],[2] and we demonstrate its effectiveness on a testchip fabricated using a commercial 0.5 μ m bulk CMOS process. The proposed regulator is based on a delay servo loop. Using the critical path replica circuit and an appropriate servo loop, the voltage at the output of the adaptive voltage regulator is adjusted so as to minimize power dissipation while guaranteeing that the delay requirements of the logic block are always satisfied. Some margin (10-15%) is included in the critical path replica to account for the physical differences and mismatch between the critical path in the actual block of logic and the critical path replica used within the regulator circuit.

In this paper we use the servo-loop to regulate the voltage levels for a 16*16+36-bit DSP MAC implemented in mixed swing QuadRail technique as described in [3]. However the design of the regulator is very general purpose and could be used to regulate the voltage levels for any digital block of considerable size.

The remainder of the paper is organized as follows. In section II a block description of the regulation scheme with top

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ISLPED '00, Rapallo, Italy.

Copyright 2000 ACM 1-58113-190-9/00/0007... \$5.00.

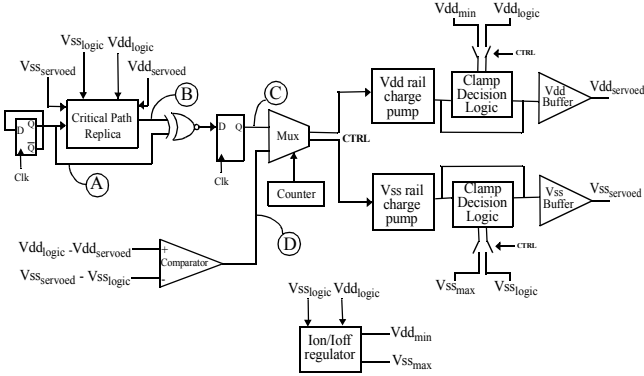


Figure 1. Delay servoing based regulation scheme

level functionality is presented. Section III shows a detailed analysis of the measured results. Finally we offer our conclusions in Section IV.

II. REGULATION SCHEME

In figure 1 a block diagram description of the regulation loop is presented. It is comprised of various blocks which perform the three basic steps of this voltage generation process:

- prediction of the worst-case delay of the block being regulated (on-chip delay measurement)
- generation of the optimum voltage level (voltage dialing)
- generation of the actual power supply voltage (voltage buffering) to feed the load circuit.

The ultimate regulator has the form of a Delay Locked Loop that generates the output voltage such that the target speed is met under all process, temperature and voltage variations.

A detailed understanding of the functionality of each block could be obtained through [1]. However for sake of brevity we discuss only the top level functionality here.

Delay servoing is based on the performance of the system's critical path replica (CPR) which has some slack (10-15%) built in as safety margin. The critical path replica can be extracted from the actual circuit using a static timing analyzer and extra stages should be included to compensate mainly for the fact that the loads due to long interconnects are not reproduced. As shown in figure 1, the CPR is fed in input with the signal A which toggles at $f_{clk}/2$ where f_{clk} is the target clock frequency. The side inputs in the CPR are set such that the signal can propagate through the entire path (B). The comparison between A and B performed through an Exor or Exnor gate, depending on the polarity of the transitions, generates a logic signal C which tells the servo-loop whether the circuit is fast enough. If this is the case, the power supply voltage can be reduced allowing for power savings; if not, the power supply has to be increased to maintain the correct functionality of the circuit. In our particular case, the regulator generates two voltage levels, $V_{dd_servoed}$ and $V_{ss_servoed}$, from the main external power supplies V_{dd_logic} and V_{ss_logic} since the target application implement-

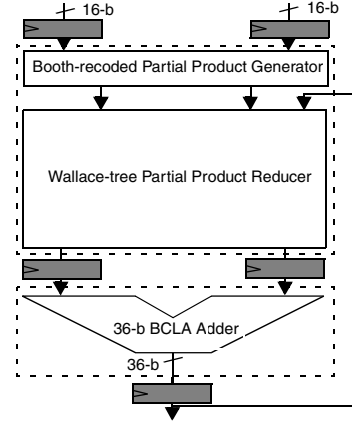


Figure 2. 16*16+36-bit MAC architecture.

ed as QuadRail requires four power levels [4]. The signal C acts as a control signal to the charge pumps that appropriately adjust the voltage levels: if the CPR delay doesn't meet the maximum permitted timing specifications $V_{dd_servoed}$ is increased and $V_{ss_servoed}$ is reduced increasing the swing of the logic and improving its performance. Vice versa, the swing is reduced (i.e. $V_{dd_servoed}$ is lowered and $V_{ss_servoed}$ is increased) and power saving is achieved. In the actual implementation of the servo-loop the signal C is multiplexed with D , where the latter represents the result of the comparison between $V_{dd_logic} - V_{dd_servoed}$ with $V_{ss_servoed} - V_{ss_logic}$. This operation is important in that it allows to keep the generated inner rails always centered around a middle value (i.e. $(V_{dd_logic} - V_{ss_logic})/2$). Once every N cycles, where N can be fixed using a counter, the control is passed from the signal C to the signal D and in this particular clock cycle both the generated rails tend to move in the same direction (i.e. $V_{dd_servoed}$ and $V_{ss_servoed}$ are both increased or decreased) to reach the optimal centering. The compression or expansion of the inner rails is performed by two capacitive ratio based charge pumps as described in [1]. The size of the voltage steps from the charge pump is chosen as a trade off between the resolution at the output voltage and the time it takes for the regulator output to reach its steady state value. The range of voltage regulation is delimited between V_{dd_logic} and V_{dd_min} for $V_{dd_servoed}$ and between V_{ss_max} and V_{ss_logic} for $V_{ss_servoed}$. V_{dd_min} and V_{ss_max} are selected using the same kind of circuit described in [3]. In that application a series regulator (Ion/Ioff ratio based regulator) is used to select the minimum voltage swing for the inner rails in a mixed swing QuadRail that could guarantee a fixed ratio between the average dynamic current in a MOS transistor (Ion) and its leakage current when it is shut off (Ioff). Other circuit design styles may also wish to set a minimum on the regulated output voltage, V_{dd_min} , for other reasons (e.g., to maintain adequate noise margins). This function of "confining" the output voltage between V_{dd} and V_{dd_min} is done by a

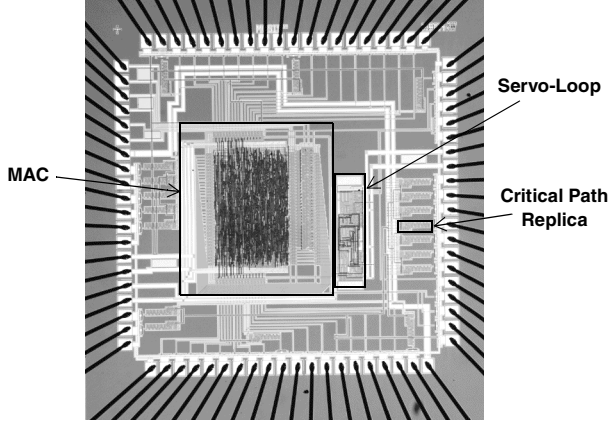


Figure 3. Die microphotograph.

comparator and some logic circuitry as in figure 1. The last components in this servo-loop are the output buffers used to generate the actual voltage to drive the circuit [1].

III. TARGET APPLICATION

The load circuit that we chose for our voltage regulator is a $16 \times 16 + 36$ Bit MAC. The MAC consists of an overlapped bit-pair Booth-recoded, (3,2) counter-based Wallace tree 16×16 -bit multiplier and a 36-bit Block Carry Lookahead final accumulator (BCLA), with a single pipeline stage between the multiplier and accumulator for enhanced throughput (figure 2). The Wallace tree multiplier is the most power-critical MAC component, consuming 75% of total power. This is due to the substantial interconnect capacitances driven by the 28-transistor-based (3,2) counters within the Wallace tree. In order to lower the multiplier power, it is realized in the mixed-swing QuadRail technique utilizing four distinct power rails [3]. The outer two rails ($V_{dd_{logic}}$ and $V_{ss_{logic}}$) are fixed and supplied externally whereas the inner two rails ($V_{dd_{servoed}}$ and $V_{ss_{servoed}}$) are generated by the on-chip regulator. These two inner voltages are adaptively adjusted by the regulator in the face of process variations or target operating frequency so as to exploit the delay slack which exists between the multiplier (power critical block) and the BCLA (delay critical block) and thus obtain additional power savings.

IV. TESTCHIP

The regulation scheme described in the previous paragraphs has been realized in a commercial $0.5\mu\text{m}$ bulk-CMOS process, as shown in figure 3. Along with the fundamental blocks used to validate the methodology (i.e. MAC and Servo-Loop), several dummy critical path replicas have been included for an evaluation of the impact of process variations on both intra and inter-chip gate delays. The measured waveforms in figure 4 depict the degree of adaptive regulation which the proposed servo-loop is capable of providing. In this particular set of experiments the operating frequency is

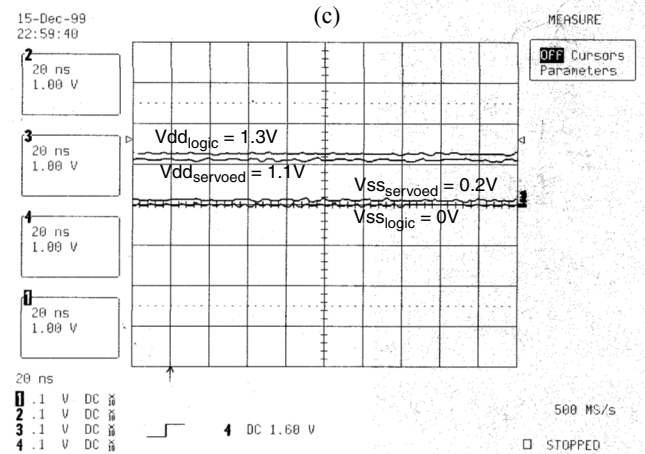
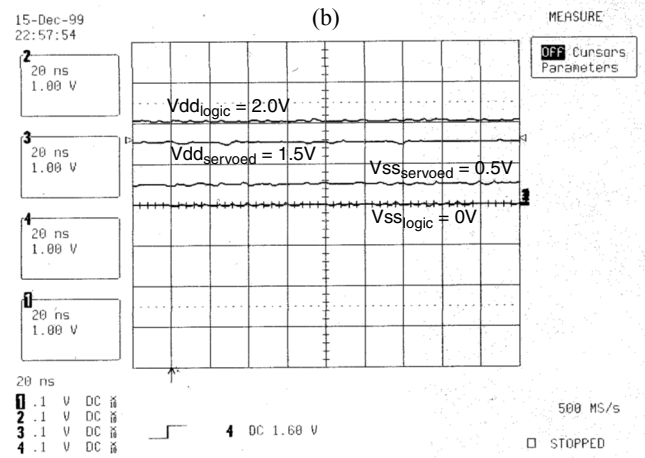
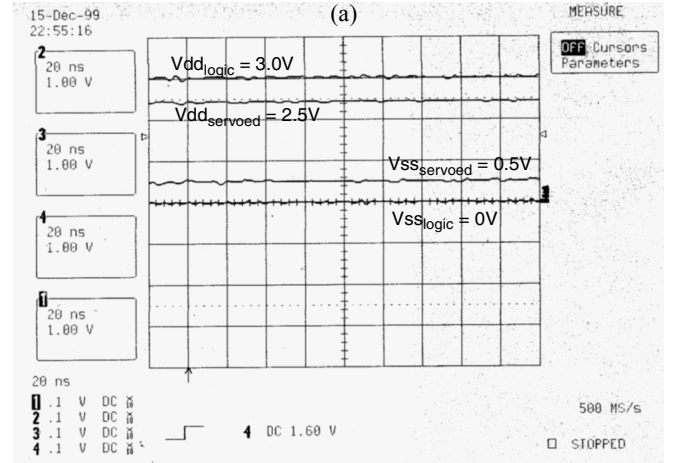


Figure 4. Regulated rails: measured waveforms.

kept fixed (20Mhz) and the external voltage is progressively decreased. If the external power supply voltage is 3V (a), $V_{dd_{servoed}}$ is equal to 2.5V and $V_{ss_{servoed}}$ is equal to 0.5V. In this case the output voltages are determined by the inner bounds $V_{dd_{min}}$, $V_{ss_{max}}$ (power optimum rails fixed by the series regulator). If the external voltage is lowered to 2V (b), $V_{dd_{servoed}}$ is equal to 1.5V and $V_{ss_{servoed}}$ is equal to 0.5V.

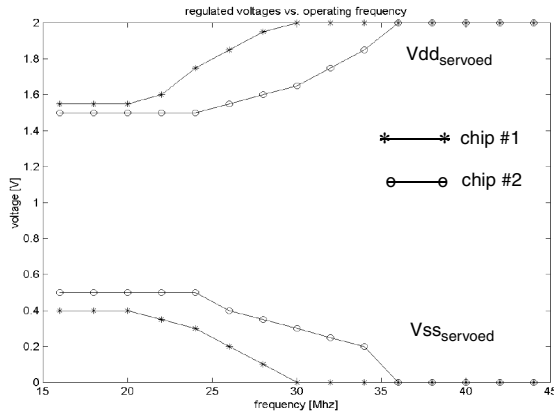


Figure 5. Regulated voltages vs. operating frequency

Even for this value of external supply voltage, the process is fast enough to guarantee the correct functionality of the circuit and the output voltages of the regulator are again fixed by the Ion/Ioff based regulator. In the last step of this experiment (c) the power supply is further reduced down to 1.3V and $V_{dd_servoed}$ is regulated to 1.1V and $V_{ss_servoed}$ to 0.2V. Since the external power supply is quite low, the servo-loop overrides the inner voltage bounds as dictated by the series regulator and instead generates the levels required to meet the target speed. Thus in summary the voltage regulator always generates the power levels so as to minimize the power consumption of the load circuit while *always* guaranteeing target speed.

A complementary set of experiments was performed employing a fixed supply voltage of 2V and progressively increasing the operative frequency. The results for two sampled chips are shown in figure 5. We can see how the regulator tends to expand the two inner rails, thereby guaranteeing the circuit performance as the target operating frequency is ramped up.

The regulator shows a high degree of voltage regulation as is clearly visible from the extremely low ripple on measured waveforms in figure 4. The negative feedback circuit used in the output buffer [1] and the use of two large decoupling capacitors (30pf) helps us in achieving this high degree of regulation. Nevertheless, even with such large passive components the area of the regulator is only about 12% the area of the MAC.

In table 1 the power dissipation of the regulator is compared with the power consumption of the regulated MAC. In particular we show here the values corresponding to different operating conditions (different external power supply voltages and target operating frequency). At higher frequencies (50Mhz) the power consumption of the regulator is about 25-35% of the power consumption of the MAC depending on the external voltage. If the operating frequency is decreased the power consumption tends to increase and eventually the regulator ends up consuming more than the MAC when the

frequency is as low as 10 Mhz. This is mainly due to the fact that the power consumption of the regulator is frequency independent since only few blocks comprising it are directly driven by the main clock. Most of the blocks are analog circuits biased at a certain current and consuming a fixed amount of power.

Such a study clearly shows that this technique is very effective for lower voltages and high operating frequencies where the fixed power in the regulator is more than offset by the high switching power in the regulated block.

Table 1: Power dissipation

operating conditions	MAC (μ W)	servo-loop(μ W)
3V@50MHz	22.8	8.0
3V@25MHz	10.0	7.6
3V@10MHz	5.4	7.5
2V@50MHz	9.1	2.4
2V@25MHz	5.3	2.3
2V@10MHz	1.9	2.1

In figure 6 the power dissipation in the regulator is broken down to give an approximate estimation of the contribution of the different elements. The elements which contribute most to the total power consumption of the regulator are the output buffers. The clamp decision logic and the bias network contribute to a fair amount of energy dissipation, too. From the data on the power consumption few conclusions can be drawn. First of all, in order to offset the intrinsic power consumption of the regulator high frequencies should be targeted since the power consumption of the servo-loop is basically frequency independent. If we were to consider a block consuming more power than the MAC, the very same regulator could be used and only an up sizing of the output buffers and decoupling capacitances should be needed.

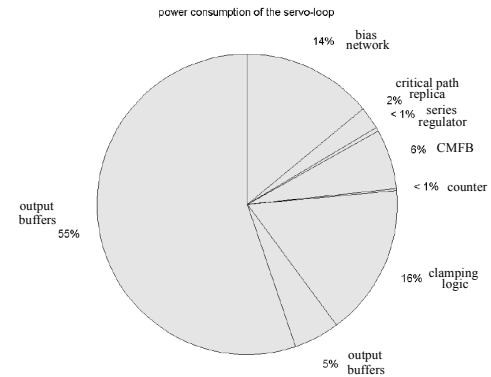


Figure 6. Power contribution of different elements in the regulator.

As previously mentioned, 19 critical path replicas have been included in the testchip. Measurements of their delay on the 25 chips available is summarized in figure 7. The delay distribution shows a maximum variance of 7ns. Though this variance value is lower than that obtained through simulations of the critical path using HSPICE (using ‘fast’ and ‘slow’ MOS models, the difference in delay is 11ns) but it is still a significant amount. On the other hand an analysis of the delays of groups of critical path replicas on the same chip and in close proximity, showed that this variance is reduced to 1-2ns. The global on-chip variance extended to all the 19 CPRs was around 3-4ns. From these observations, we can

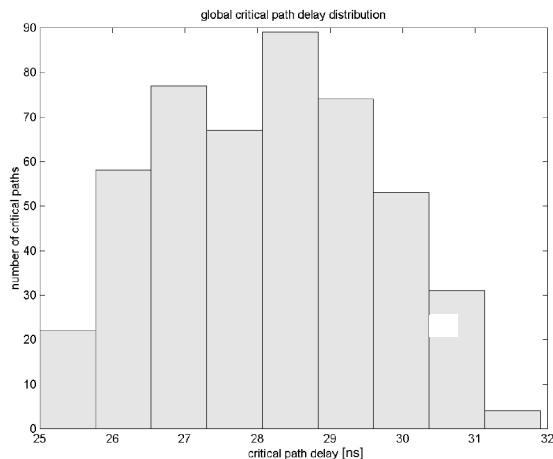


Figure 7. Critical path replicas delay distribution.

conclude that an on-chip measurement aiming at predicting the system’s performance is extremely valuable to track process fluctuations and is even more efficient if the CPR is placed in close proximity to the circuit being regulated

V. CONCLUSIONS

We have proposed a completely on-chip voltage regulation scheme based on delay-servoing which, in the face of manufacturing and operating range variations, adjusts the voltage level so as to minimize the power consumption while always guaranteeing the target clock frequency. We have also described the measured results from a testchip fabricated in a 0.5µm CMOS process to prove the effectiveness of this technique. An extremely small ripple on regulated power rails demonstrates the high degree of load regulation this servo-loop can provide. An analysis of the power consumption of the regulator indicates that in order to offset the power consumption of the regulation block, high operating frequencies should be targeted. Further improvements are being considered on the components that most contribute to the total power dissipation, such as the output buffers and clamp decision logic.

VI. REFERENCES

- [1] A.Aggarwal, and L.R.Carley, “A Completely On-Chip Voltage Regulation Technique for Low Power Digital Circuits,” Proc. ISLPED, Jul. 1999.
- [2] L.R.Carley, A.Aggarwal, R.Krishnamurthy, “Decreasing Low-Voltage Manufacturing-Induced Delay Variations with Adaptive Mixed-Voltage-Swing Circuits,” Proc. ISLPED, Aug 1998.
- [3] R.K.Krishnamurthy, H. Schmit, and L.R.Carley, “A Low-power 16-bit MAC using Series-Regulated Mixed-Swing Techniques,” CICC, May 1998.
- [4] L.R.Carley, “QuadRail: A Design Methodology for Ultra Low Power Integrated Circuits,” Proc. IEEE IWLDP, April 1994.