

A Voltage Overscaled Low-Power Digital Filter IC

Rajamohana Hegde and Naresh R. Shanbhag

Abstract—In this brief, we present an integrated circuit implementation of a low-power digital filter in 0.35- μm 3.3-V CMOS process. The low-power technique combines voltage overscaling (VOS) and algorithmic noise tolerance (ANT) to push the limits of energy efficiency beyond that achievable by voltage scaling alone. VOS refers to scaling the supply voltage beyond the limit imposed by the throughput constraints. ANT is an algorithmic level error-control technique that is employed to restore the algorithmic performance degradation in terms of output signal-to-noise ratio (SNR) caused by VOS. Measured results indicate 40%–67% reduction in energy dissipation over optimally voltage-scaled systems with less than 1-dB loss in SNR for a wide range of filter bandwidths ($0.05f_s$ – $0.25f_s$, where f_s is the sampling frequency).

Index Terms—Error control, low power, voltage overscaling.

I. INTRODUCTION

SUPPLY voltage scaling has proved to be an effective technique [1] for designing low-power systems, in general, and digital signal processing (DSP) and communications systems, in particular. However, a reduction in supply voltage results in an increase in the circuit delay τ_d given by

$$\tau_d = \frac{C_L V_{dd}}{\beta(V_{dd} - V_t)^\alpha} \quad (1)$$

where α is the velocity saturation index, β is the device transconductance, C_L is the load capacitance, V_{dd} is the supply voltage, and V_t is the device threshold voltage. Low-power design techniques where throughput is traded off for energy efficiency via voltage scaling in either static [2] or dynamic fashion [3] have been proposed. For all known techniques, the extent of voltage scaling (and hence, power savings) is limited by the throughput requirements of the application. We propose a combination of *voltage overscaling* (VOS) and *algorithmic noise tolerance* (ANT) to overcome this limit. The proposed technique is particularly suitable for the design of low-power DSP and communications systems, as described next.

Fig. 1 illustrates the key idea. The input $x(n)$ arrives at a fixed sample rate f_s . This puts an upper bound equal to the sample period $T_s = 1/f_s$ on the critical path delay T_{cp} of the digital filter. The supply voltage at which $T_{cp} = T_s$ is denoted by $V_{dd-crit}$, which is also the lower bound on the supply voltage V_{dd} . Present-day voltage scaling operates in the range $V_{dd} \geq V_{dd-crit}$. On the other hand, *VOS implies operation in*

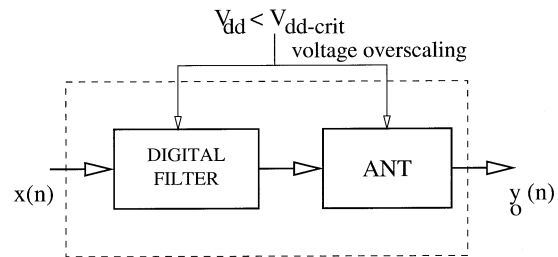


Fig. 1. Principle of voltage overscaling.

the range $V_{dd} < V_{dd-crit}$, resulting in greater energy savings but also leading to intermittent errors at the output whenever the critical and other longer paths are excited. In the context of DSP and communication systems, these errors lead to a degradation in the signal-to-noise ratio (SNR). Thus, voltage overscaled systems need to include some form of error-control to restore the SNR degradation. The error-control block should also have relatively low complexity in order for appreciable energy savings to occur. It turns out that one can employ ANT [4] techniques to design low-complexity error-control blocks for DSP and communication systems.

In this brief, we describe measured results quantifying the energy savings of a prototype digital filter IC designed and fully tested in a 0.35- μm 3.3-V CMOS process. Measured results indicate 40%–67% reduction in energy dissipation over optimally voltage-scaled present day systems with less than 1-dB loss in SNR for a wide range of filter bandwidths ($0.05f_s$ – $0.25f_s$, where f_s is the sampling frequency). The rest of this brief is organized as follows. In Section II, we describe the design issues related to the proposed scheme and the error-control scheme for digital filtering. The architecture of the digital filter and other implementation details are presented in Section III, followed by measured results in Section IV.

II. ALGORITHMIC NOISE TOLERANCE (ANT)

The effectiveness of the proposed scheme depends on the following:

- 1) the magnitude of the errors and the error frequency (the rate at which the errors occur);
- 2) the effectiveness of the error-control scheme in detecting and correcting the errors;
- 3) the error-control overhead in terms of additional power dissipation, area, and design time.

The magnitude of the errors and the frequency at which they occur depends primarily on the architecture employed, the extent to which the supply voltage is overscaled, and the input statistics. For example, in Fig. 2, the path delay distribution (histogram of the evaluation time) for all possible input combinations for an 8-bit ripple-carry adder is shown. It can be seen that more than 95% of the possible input combinations are evaluated

Manuscript received November 19, 2002; revised October 6, 2003. This work was supported by the National Science Foundation under Grant CCR 00-00987.

R. Hegde is with Intersymbol Communications Inc., Champaign, IL 61821 USA (e-mail: raj@intersymbol.com).

N. R. Shanbhag is with the Electrical and Computer Engineering Department and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801 USA (e-mail: shanbhag@mail.icims.csl.uiuc.edu).

Digital Object Identifier 10.1109/JSSC.2003.821775

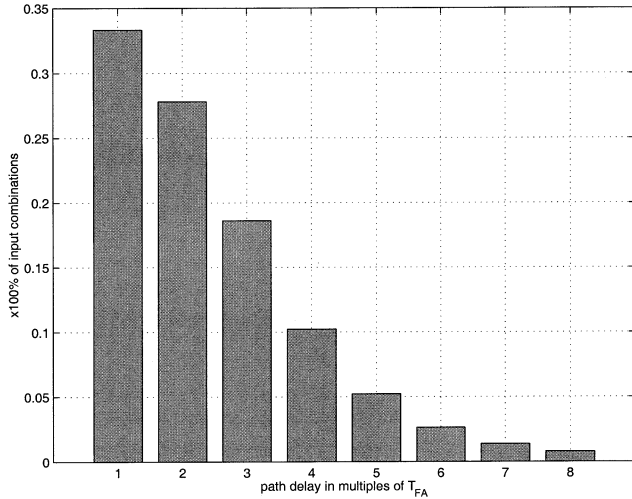


Fig. 2. Path delay distribution over all input combinations for an 8-bit ripple-carry adder.

in $5T_{FA}$ time units, where T_{FA} is the delay of a 1-bit full adder. With a uniform input distribution, reducing V_{dd} such that the new full-adder delay $T'_{FA} = (5/8)T_{FA}$ leads to an error probability of 0.05. Note that the error probability can be reduced further if the input distribution is such that the critical paths are excited less often.

At the algorithmic level, several noise-tolerance techniques can be developed to provide error control. The effectiveness of the particular technique depends on the type of errors it can detect, the accuracy of error correction, and the overhead involved in terms of additional power dissipation and area. The objective here is to achieve a reduction in *total* energy dissipation while meeting the algorithmic specifications in terms of SNR. In this section, we first present a simple error-control technique that involves little overhead but whose effectiveness is also limited. We refer to this technique as the difference-based technique. We then present the more general technique based on linear prediction.

A. Difference-Based Error-Control Technique

This error-control technique is suitable for low-pass digital filters with a relatively narrow passband. This technique can be shown to be a special case of the more general predictive error-control scheme presented in the next subsection.

The output $y(n)$ of an N -tap filter is given by

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (2)$$

where $h(k)$ denotes the k th coefficient of the filter, $x(n-k)$ denotes the input at the $n-k$ th instant, and N is the length of the filter. The difference in consecutive samples of the filter output is given by $y_d(n) = y(n) - y(n-1)$. Let $\hat{y}(n)$ denote the filter output when the filter implementation is unreliable, with $\hat{y}(n) = y(n) + y_{err}(n)$, where $y_{err}(n)$ denotes the soft errors in the filter output. In the context of soft DSP, note that $y_{err}(n)$ is nonzero only when the input pattern is such that longer paths in the filter implementation are excited. Assuming that the past

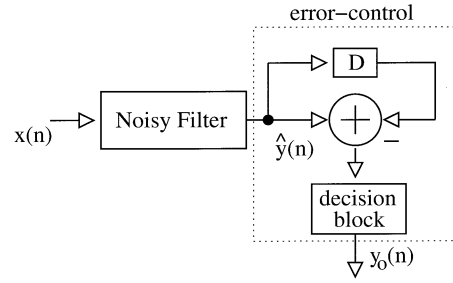


Fig. 3. Difference-based ANT technique for filtering.

output is noiseless, i.e., $y_{err}(n-1) = 0$, we have $\hat{y}_d(n) = y_d(n) + y_{err}(n)$, where $\hat{y}_d(n)$ is the difference in the filter output in presence of errors. The difference-based error-control scheme illustrated in Fig. 3 is as follows.

- Compute $\hat{y}_d(n) = \hat{y}(n) - \hat{y}(n-1)$.
- Error detection: If $|\hat{y}_d(n)| \geq E_{th}$, an error is declared. The choice of E_{th} is described later.
- Error correction: If an error is declared, $y_o(n) = \hat{y}(n-1)$; else $y_o(n) = \hat{y}(n)$.

If an error is detected, the past output sample is taken to be the estimate for the current output sample. For smaller filter bandwidths, the variance of $y_d(n)$ is an order of magnitude lower than that of $y(n)$. Therefore, if the magnitude of the error in filter output $|y_{err}(n)|$ is large (this is the case in soft DSP as the errors occurring mainly in the MSBs), then $|\hat{y}_d(n)|$ will be large. The value of E_{th} is chosen such that $|\hat{y}_d(n)| < E_{th}$ when $y_{err}(n) = 0$ (in the absence of error) and $|\hat{y}_d(n)| > E_{th}$ when $y_{err}(n) \neq 0$ (in the presence of error). As the variance in $y_d(n)$ increases with bandwidth, the effectiveness of the above approach in performing error detection deteriorates. Hence for larger bandwidths, a more sophisticated technique such as the prediction-based technique is employed.

For the difference-based error-control technique for low-pass filters, note that the past sample of the filter output is used as the estimate of the current sample. Hence, as shown in Fig. 3, the overhead involved is just an adder, a delay element, and a comparator for the decision block.

B. Prediction-Based Error-Control Technique

A general prediction-based scheme is shown in Fig. 4. In this technique, a low-complexity linear forward predictor is employed to get an estimate of the current sample of the filter output based on its past samples. The prediction error $e_p(n)$ is usually small in the absence of errors in the filter output. A large error in the filter output due to excessive voltage reduction leads to an increase in the magnitude of prediction error which makes it easy to detect errors. Error correction is performed by employing the predicted sample as the actual filter output. A more detailed description of this scheme is provided in [4].

Note that the difference-based scheme is a special case of the prediction-based scheme, where the predictor has a single tap with unity gain. The prediction-based ANT algorithm can provide effective error-control if the following three conditions are satisfied: 1) errors in the filter output are spaced apart by at least N_p samples, where N_p is the length of the predictor; 2) the error magnitudes are large; and 3) the error-control block is

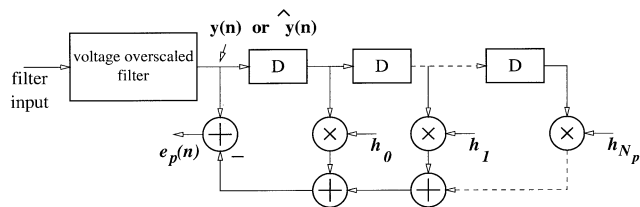


Fig. 4. Prediction-based ANT technique for filtering.

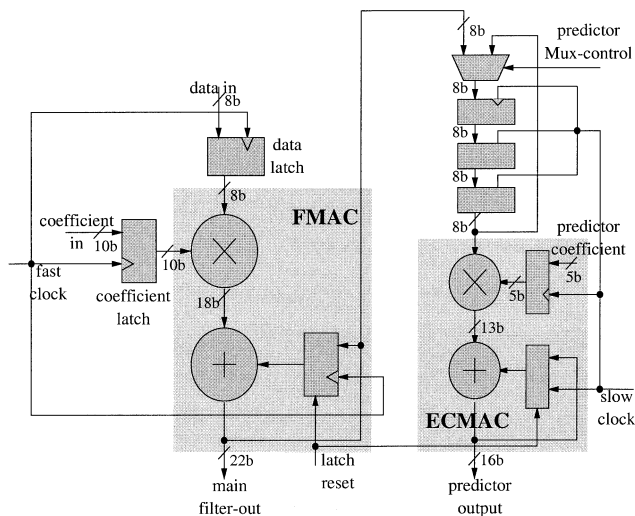


Fig. 5. Chip architecture.

error free. These three conditions can be met in practice by employing an architecture which has the following properties: 1) it is delay-imbalanced; 2) it employs arithmetic units that compute in an LSB-first manner; and 3) the error-control block has a shorter critical path than that of the filter. All three properties are present in an architecture that employs the commonly used ripple-carry adders, array multipliers, and the prediction-based ANT technique.

III. CHIP ARCHITECTURE

The architecture of the prototype VOS digital filter IC is shown in Fig. 5. The IC has two multiply-accumulate (MAC) circuits, one for the digital FIR filter (FMAC) and the other for the predictor in the error control block (ECMAC), as shown in Fig. 5. The FMAC consists of a 10-b \times 8-b unsigned array multiplier followed by a modified two's complement 22-b accumulator. The ECMAC is architecturally similar to the FMAC but has a 5-b \times 8-b array multiplier followed by a 16-b accumulator. The smaller precisions of the ECMAC translate to a lesser area and power overhead and also a shorter critical path delay.

The coefficient as well as data inputs to the FMAC and the coefficient inputs to the ECMAC are fed externally and are latched. Note that the clock signal that drives the ECMAC (*slow clock*) latches is slower than that for the FMAC (*fast clock*). This is because the length of the predictor is substantially shorter than that of the FIR filter. For a filter length of $N = 32$ taps, the maximum predictor length is $N_p = 4$ taps. For the experiments reported in this brief, the *fast clock* frequency was eight times that

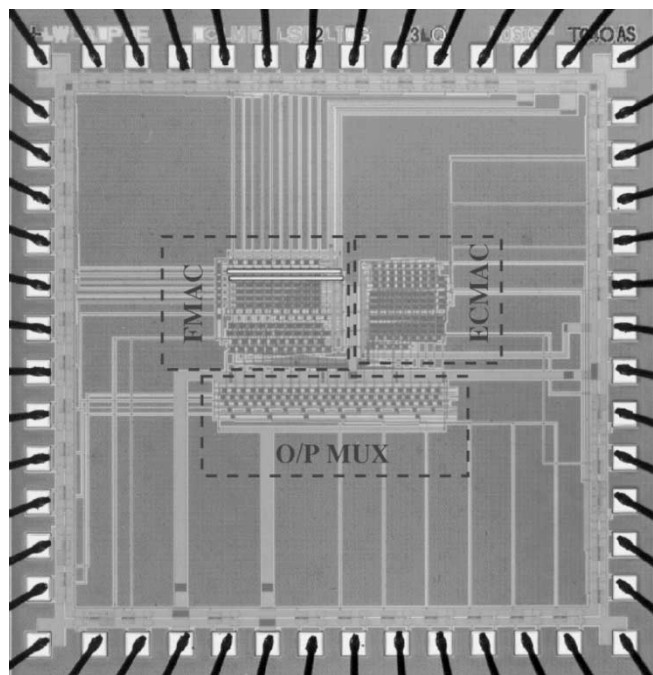


Fig. 6. Chip microphotograph.

of the *slow clock*. This ensures correct operation of the ECMAC at reduced supply voltages due to the slower clock and hence enables VOS for both FMAC as well as ECMAC.

The FMAC output feeds into the ECMAC which employs the past FMAC output samples to statistically predict the present FMAC output. All circuit structures are implemented for most part in static CMOS except for the latches which employ the C²MOS style. In this implementation, the latches employed in both the coefficient and the data registered have been designed to operate properly at subcritical supply voltages. This prevents errors occurring in the latches during subcritical operation. The latches are also powered through a separate supply pin. This provides the option of not overscaling the supply voltage to the latches. In a datapath-intensive circuit, the additional power dissipation due to nonoverscaled latches would be marginal. The full adders in both the FMAC and the ECMAC are implemented in the classic 24-transistor symmetric configuration [6].

IV. EXPERIMENTAL RESULTS

A prototype VOS digital filter IC (Fig. 6) has been designed and fully tested in a 0.35- μ m 3.3-V CMOS process. The transistor complexity of the FMAC is 3872, and that of the ECMAC is 2423.

Measurements in the laboratory have shown that the critical supply voltage for FMAC at a clock frequency of 88 MHz is $V_{dd,crit,FMAC} = 3.55$ V. The worst case power consumed at this frequency and voltage is 105.47 mW. Due to the lower precision and fewer taps in the predictor in Fig. 5, the ECMAC is clocked at a frequency of 11 MHz, which is eight times slower than the main MAC. Thus, the ECMAC has $V_{dd,crit,ECMAC} = 2.25$ V.

In the VOS mode, the FMAC and ECMAC are operated at the same voltage which lies in the subcritical range 2.32 V $<$ $V_{dd,FMAC}$ $<$ $V_{dd,crit,FMAC}$ at a fixed clock frequency of 88 MHz. Thus, the FMAC has intermittent errors at

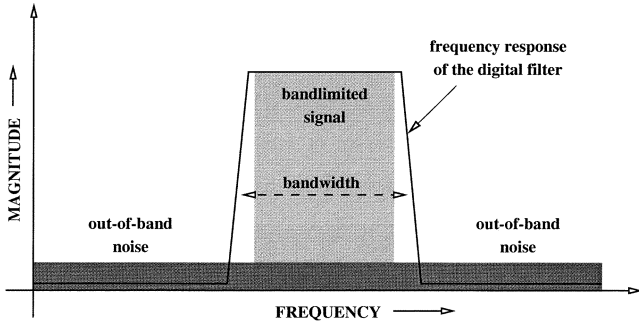


Fig. 7. Frequency selective filtering.

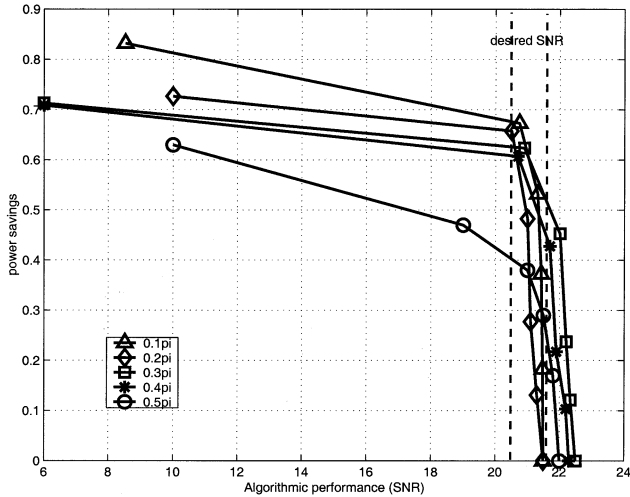


Fig. 8. Measured results: Algorithmic performance versus power savings.

the output while the EMAC is error free. The error frequency in FMAC depends on how often the critical paths in the circuit are excited. As the FMAC employs the array architecture, the longer paths form a very small fraction of the total number of possible paths due to the delay imbalance. Hence, the error frequency increases gracefully with reduction in supply voltage.

The algorithmic performance in terms of SNR is measured in the context of a frequency selective filtering application as illustrated in Fig. 7. Note that the signal is band limited and the noise is wideband. The digital filter is employed to suppress the out-of-band noise and improve the SNR. The filter bandwidth is set to be the same as the signal bandwidth. This ensures maximum out-of-band noise rejection. The coefficients of the filter determine its frequency response. For the implementation presented in this brief, the filter coefficients are fed externally. This enables measurements for any bandwidth.

The plot of SNR versus measured power savings for filter bandwidths ranging from $0.05f_s - 0.25f_s$ is shown in Fig. 8. The power savings is given by

$$PS = \frac{P_{\text{ref}} - P_{\text{vos}}}{P_{\text{ref}}} \quad (3)$$

where P_{vos} is the power at overscaled supply voltage and P_{ref} is the power measured with respect to a reference filter operating at $V_{\text{dd,crit,FMAC}} = 3.55$ V. The reference filter is referred to as an *optimally scaled present-day system* because this filter operates error free, and hence, does not require ANT techniques. The power consumed by the ECMAC is included in the power calculations for the VOS filter. Observe from Fig. 8 that power savings in the range 40%–67% is achieved over a present-day voltage scaling scenario with a degradation of less than 1 dB in output SNR for a 32-tap FIR filter for bandwidths in the range $0.05f_s - 0.25f_s$ at a clock speed of 88 MHz. As expected, an increase in bandwidth reduces power savings due to the reduction in output correlation.

V. CONCLUSION

We have presented an IC implementation of a low-power digital filter. Voltage overscaling in combination with algorithmic noise tolerance [4] (referred to as soft DSP) is the low-power technique being employed. We have demonstrated experimentally that the proposed low-power technique provides up to 67% additional energy savings over an optimally voltage-scaled present-day system, i.e., a system that operates with a supply voltage equal to $V_{\text{dd,crit}}$. Note that variations in process and temperature could either speed up or slow down the datapaths. In such cases, either the supply voltage or the complexity of the error-control technique can be adjusted adaptively by monitoring the output SNR.

This technique opens up new areas of research in low-power design, such as the investigation of 1) arithmetic and filter architectures that favor soft DSP and 2) ANT techniques for other commonly employed DSP blocks such as fast Fourier transforms, discrete cosine transforms, and infinite impulse response and adaptive filters. On a broader scale, in addition to low-power design, ANT techniques can pave the way for achieving the dual goals of reliability and energy efficiency of integrated circuits in current and future technologies by combating deep-submicron noise and process nonidealities at the algorithmic level.

REFERENCES

- [1] A. Chandrakasan and R. W. Brodersen, "Minimizing power consumption in digital CMOS circuits," *Proc. IEEE*, vol. 83, pp. 498–523, Apr. 1995.
- [2] R. Gonzalez, B. M. Gordon, and M. A. Horowitz, "Supply and threshold voltage scaling for low power CMOS," *IEEE J. Solid-State Circuits*, vol. 32, pp. 1210–1216, Aug. 1997.
- [3] A. Sinha and A. Chandrakasan, "Energy efficient filtering using adaptive precision and variable voltage," in *Proc. 1999 ASIC/SOC Conf.*, pp. 327–331.
- [4] R. Hegde and N. R. Shanbhag, "Soft digital signal processing," *IEEE Trans. VLSI*, vol. 9, pp. 813–823, Dec. 2001.
- [5] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: Principles, Algorithms and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
- [6] J. M. Rabaey, *Digital Integrated Circuits: A Design Perspective*. Englewood Cliffs, NJ: Prentice-Hall, 1996.