# Design of Completion Detection Circuits for Self-timed Systems Operating in Subthreshold Regime

Omer Can Akgun, Yusuf Leblebici, Eric A. Vittoz
Swiss Federal Institute of Technology (EPFL)
Microelectronic Systems Laboratory (LSM)
Lausanne, CH-1015, Switzerland
E-mail:{omercan.akgun, yusuf.leblebici, eric.vittoz}@epfl.ch

***Abstract*— In this paper implementation of a novel completion detection method for self-timed, asynchronous subthreshold circuits is presented. By employing the self-timed operation principle, substantial speed gains in the operation of the asynchronous pipelines can be realized. The completion detection system is very simple, consisting of a sensor transistor, a very basic AC-coupled amplifier and a monostable multivibrator. The proposed method can be easily integrated into the CMOS design flow. The advantages of the proposed completion detection system is shown through simulations on an 8-bit ripple carry adder in a standard 0.18$\mu m$ CMOS process operating at 400mV supply voltage.**

## I. Introduction

Power density and power consumption of complex digital systems has become a major concern during the recent years, both due to thermal concerns and due to limited battery lifetime in mobile applications. Any significant reduction in power dissipation can only be achieved by lowering the operating voltage of the circuits. This would be possible by relaxing the constraints of classical strong-inversion operation of MOSFETs, and by accepting the notion that transistors can (and will) be operated well below threshold, in the subthreshold (weak-inversion) regime.

In subthreshold mode of operation, the supply voltage can be scaled aggressively and power dissipation can be decreased significantly. There are successful implementations of digital circuits working in the subthreshold region [1]–[3]. Subthreshold operation of static CMOS logic has been analytically analyzed using the EKV model in [4]. According to the analysis, to benefit the most from the subthreshold operation, the logic circuits should be run at their maximum operating frequency with an activity factor $\alpha$ as close to 1 as possible.

Maximizing $\alpha$ is non-trivial in synchronous circuits. Because of the fixed time slot for computation, data-dependent processing times cannot be exploited in synchronous operation. Also in a synchronous system, power is consumed with each clock transition regardless of the data or state change. Although clock gating has been proposed as a solution to this problem [5], it cannot provide a comprehensive solution to the problem of data-dependent computation times. On the other hand, asynchronous circuits do not consume clocking power and have the potential of operating with tunable voltage supplies at an $\alpha$ factor of 1 at the optimum supply voltage. In addition, asynchronous operation also enables the absorption of time-domain (delay) variations that inevitably become more prominent in the subthreshold regime due to the device parameter and temperature variations. Therefore, implementation of efficient asynchronous circuits operating in subthreshold regime would be a subject of high interest.

In this work, we demonstrate the design of a novel completion detection system (CDS) for self-timed circuits based on current sensing. The proposed CDS can operate in the subthreshold regime and sense currents in the $pA$ - $nA$ range. The proof of concept is shown by using the proposed system to generate the completion detection signals for an 8-bit ripple carry adder. It has been observed that the improvement in the delay of the overall circuit operation is significant, justifying the use of the CDS.

The remainder of the paper is organized as follows. In Section II the basic current equations for MOS transistors operating in the weak inversion regime is given. In Section III the design of the CDS emphasizing the trade-offs in the design of the AC-coupled amplifier is presented. Section IV shows the application and improvements due to the CDS in terms of the delay of the pipeline stage.

## II. MOS Weak Inversion Operation

The MOS digital circuits operate in subthreshold regime when the supply voltage is lower than the threshold voltage ($V_T$) of the transistors. The drain current of an n-channel MOS transistor operating in this regime is given by [4]

$$I_{DS} = I_S \exp\frac{V_{GS} - V_T}{nU_T}\left(1 - \exp\frac{-V_{DS}}{U_T}\right) \qquad (1)$$

where $n$ is a process dependent term called slope factor and is typically in the range of $1.3-1.5$ for modern CMOS processes. The value of $n$ depends on the depletion region characteristics of the transistor. $V_{GS}$ and $V_{DS}$ are the gate to source and

drain to source voltages, respectively. The parameter $I_S$ is the specific current which is given by,

$$I_S = 2n\mu C_{ox} U_T^2 \frac{W}{L} \tag{2}$$

where $\mu$ is the mobility of carriers, $C_{ox}$ is the gate oxide capacitance per unit area, $U_T$ is the thermal voltage whose value is $26mV$ at 300K and $\frac{W}{L}$ is the aspect ratio of the transistor.

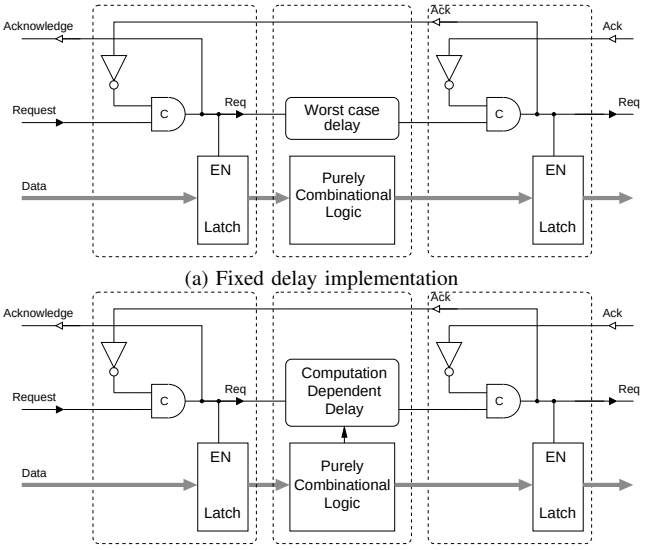
Due to the second term in (1), the drain current is 0 when $V_{DS} = 0$ but reaches its maximum value and saturates with $V_{DS}$ values higher than a few $U_T$. As it is apparent from (1), the drain current of a MOS transistor in subthreshold region shows exponential dependence on the gate-to-source, drain-to-source voltages, slope factor and the operating temperature. To alleviate the problems resulting from the exponential dependence on the supply voltage, process parameters and the operating temperature, self-timed circuits emerge as strong candidates for subthreshold operation.

## III. Completion Detection Circuitry

Asynchronous circuits are fundamentally different from their synchronous counterparts. Although the signaling convention is binary, the data propagation through the stages take place by employing handshaking signals, not by using a common clock. A basic 4-phase, bundled data asynchronous pipeline is shown in Fig. 1. In Fig. 1(a) the most common implementation (fixed delay) is shown. In this implementation the *Request* signal is delayed by an amount equal to the worst case delay of the purely combinational logic stage. This implementation is a similar approach to the synchronous operation, where unnecessary delay, which is fixed regardless of the logic computation time, is introduced. Another version of the 4-phase bundled data pipeline is shown in Fig. 1(b), where the delay should mimic the computation completion time of the logic block as much as possible.

In the literature there are examples of completion detection using current sensing methods [7], [8]. The techniques in the mentioned papers use either bipolar transistors, which are not available in standard CMOS processes, or high valued resistors. Although these techniques can be used for detecting current values in the $\mu A$ - $mA$ range, they cannot be used for detecting the currents for subthreshold operation which are in the $pA$ - $nA$ range because of the high $\beta$ requirements of the bipolar transistors and difficulty in implementing very high value resistors.

The proposed completion detection system consists of a single low threshold voltage ($V_T$) MOS transistor, an AC-coupled amplifier and a monostable multivibrator (Fig. 2). The current signal is sensed by the diode connected low $V_T$ PMOS transistor, converted to a voltage signal and compressed in the log domain. The conceptual sensed waveform that corresponds to the combinational block processing a new input set is shown in Fig. 3. Two operation regions can be discerned in the figure. One is the time when the actual computation is taking place and the other is the settling of the supply node of the




(a) Fixed delay implementation

(b) Computation based delay implementation

Fig. 1. 4-phase bundled data pipeline (After [6]).

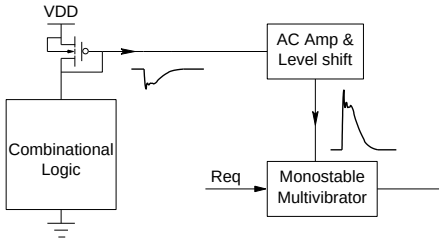



Fig. 2. Block diagram of the proposed completion detection system.

combinational logic block. The settling time depends on the capacitance of the supply node of the logic block and the resistance of the diode connected PMOS transistor. In order to avoid excessive computation delay when compared to the standard implementation, the resistance of the diode connected transistor should be kept as small as possible.

After sensing the current signal and converting to a log-compressed voltage signal, amplification of this signal is necessary before feeding it into the monostable multivibrator. The basic AC-coupled amplifier used for amplification is shown in Fig. 4. Diode connected transistors $mp1$ and $mn1$ bias the transistors $mp2$ and $mn2$, which are acting as an

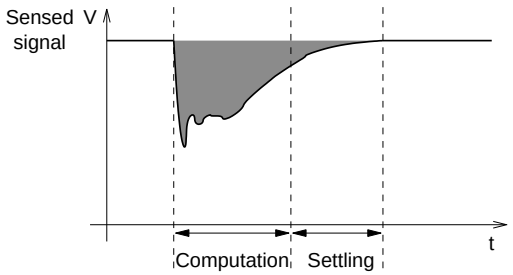



Fig. 3. Computation signature detected as the temporary drop of the supply voltage at the drain node of the current sensor device.

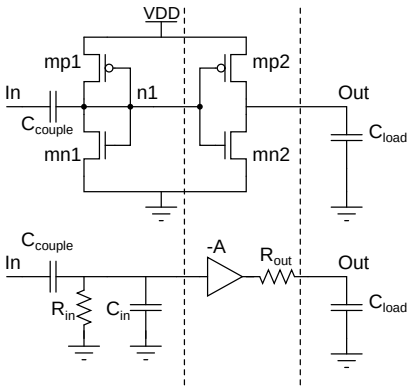



Fig. 4. Schematic and the small signal model of the AC-coupled amplifier used to amplify the detected signal.

amplifier, at the maximum gain point for a given size and DC level. In the small signal model the $C_{in}$ is the total capacitance at the node $n1$, $R_{in}$ is the resistance at the same node, $-A$ is the gain of the inverter, $R_{out}$ is the output resistance of the inverter and $C_{load}$ is the total capacitance consisting of the parasitic capacitance of the output inverter (amplifier) and the load capacitance.

$$\frac{V_{out}(s)}{V_{in}(s)} = \underbrace{\frac{sC_{couple}R_{in}}{1+sR_{in}(C_{in}+C_{couple})}}_{\text{HP Response}} \cdot \underbrace{\frac{-A}{1+sC_{load}R_{out}}}_{\text{LP Response}} \quad (3)$$

The transfer function of the amplifier is given by (3). As it can be seen from the transfer function, the AC-coupled amplifier shows a band-pass response. The high-pass characteristic is due to the coupling capacitor and the input stage, and the low-pass response with gain is the characteristic of the output stage. The important parameters while designing the amplifier and optimizing the whole CDS are the coupling capacitance, parasitic capacitance at node $n1$, the load capacitance and the gain of the amplifier. By changing the coupling capacitance the first pole location and the amplitude of the signal at the input of the output amplifier can be modified. The frequency domain response of the amplifier for multiple $C_{couple}$ values is shown in Fig. 5. The main constraint limiting the coupling capacitance selection value is the available chip area. The load capacitance, affecting the low-pass response of the amplifier is as important as the coupling capacitance and it should be kept as small as possible not to degrade the frequency response of the amplifier. Gain and bandwidth of the AC-coupled amplifier for multiple values of $C_{couple}$ and $C_{load}$ are given in Table I. It can be seen from the data in the table that with increasing load capacitance value, both the maximum gain and the bandwidth of the amplifier degrades, therefore when designing the CDS, the capacitive load of the AC-coupled amplifier should be kept minimum.

While designing the system there is a trade-off between the gain of the AC-coupled amplifier and the delay caused by the sensor transistor. If greater delay caused by the sensor transistor (larger spikes in the supply node of the combinational logic

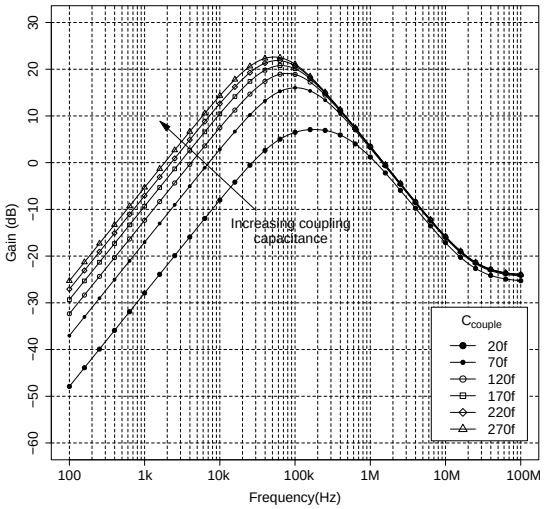



Fig. 5. Frequency domain response of the AC-coupled amplifier for multiple coupling capacitor values.

TABLE I

MAXIMUM GAIN(DB) AND BANDWIDTH DATA FOR DIFFERENT VALUES OF $C_{couple}$ AND $C_{load}$

| | | $C_{load}$ (fF) | | | | | |
|---|---|---|---|---|---|---|---|
| | | 5 | | 10 | | 15 | |
| | | Gain | BW | Gain | BW | Gain | BW |
| $C_{couple}$ (fF) | 20 | 7.1 | 568k | 3.07 | 528k | 0.321 | 511k |
| | 70 | 16 | 228k | 12.7 | 196k | 10.3 | 183k |
| | 120 | 19.1 | 163k | 16.2 | 133k | 14.1 | 120k |
| | 170 | 20.8 | 136k | 18.3 | 106k | 16.3 | 93.7k |
| | 220 | 21.9 | 120k | 19.6 | 91k | 17.9 | 78.9k |
| | 270 | 22.6 | 111k | 20.6 | 81.6k | 19 | 69.6k |

block) can be tolerated, the gain, thus the power consumption of the AC-coupled amplifier can be reduced.

The monostable multivibrator [9] is used for converting the sensed signal to a pulse, whose width is proportional to the time spent during computation (shown in Fig. 6). In this monostable multivibrator implementation, a PMOS transistor acts as a variable resistor, whose resistance is modulated by the amplified current signal. By modulating the instantaneous resistance of the PMOS transistor, the $RC$ time constant of the multivibrator is modulated as well, resulting in a pulse whose width is proportional to the area under the current curve of the combinational block. The pulse width of the multivibrator is given by equation (4)

$$T = C(R+R_{on})\ln\left[\frac{R}{R+R_{on}}\frac{V_{DD}}{V_{DD}-V_{th}}\right] \quad (4)$$

where $R$ is the average resistance of the PMOS transistor during pulse generation, $R_{on}$ is the resistance of the NOR gate and $V_{th}$ is switching threshold of the inverter. Assuming $R_{on} << R$ and $V_{th} = V_{DD}/2$, equation (4) can be simplified as $T = 0.69CR$. This means the pulse generated by the monostable multivibrator will be compressed by a factor of 0.69 compared to the actual computation time in the ideal case.

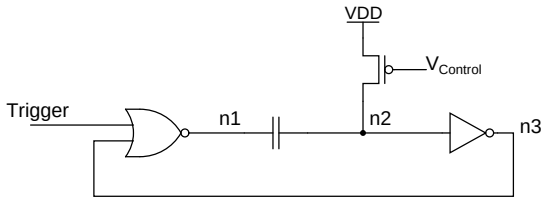



Fig. 6. Basic monostable multivibrator with variable resistor.

In the standard implementation of the multivibrator, node $n3$ is used as the output. For some cases, the resistance of the PMOS transistor might be too low, so a very fast switching occurs at node $n2$ and no switching occurs at node $n3$, resulting in no signal generation. To prevent this situation, signal at node $n1$ was used as the output. The signal at $n1$ goes low when the trigger signal goes high and if there is no change in the signal at $n3$, it goes high as the trigger signal goes low. This operation guarantees a completion detection signal is generated for each trigger input. The minimum width of the completion detection pulse is thus set by the trigger pulse width, setting the minimum delay of the *Request* signal of the preceding stage.

## IV. Simulation Results

To prove the effectiveness of the proposed completion detection method, the concept was applied to an 8-bit ripple carry adder in a standard $0.18\mu m$ process. The simulations on the adder were run for 1000 random input vectors, and the pulse width of the completion detection signal, representing the delay of the *Request* signal, were measured. The results of the simulations for $V_{DD}$=0.4V operation are shown in Fig. 7. In the figure, the response of the CDS is also plotted as a continuous curve on the y-axis and the maximum computation time for the sample data set is marked with a horizontal line. The shaded area under the curve below the marker line represents the gain in the operation speed of the asynchronous pipeline with regards to the worst case computation delay in the applied dataset. By changing the $R$ and $C$ values of the multivibrator and matching the slowest response to the worst case delay of the combinational logic block, substantial gain in the speed of any logic pipeline can be realized. A dashed line with a slope of 1 is also drawn in the figure as a reference. As long as all the data points are above this curve, the delayed *Request* signal will not arrive at the next stage before the computation is completed . The pulse width (delay) generated by the completion detection circuitry and the computation time has a correlation coefficient of 0.72, showing a clear correlation between the computation time and the generated delay. The data points in Fig. 7 have a slope of 0.5 which clearly shows the compression of the computation time by the monostable multivibrator. It can be seen that the mean delay is improved by about 17% compared to the maximum (worst-case) computation time. The corner simulations were also run and the completion detection circuit functioned without any loss of performance for all the corner cases except the slow-slow corner.

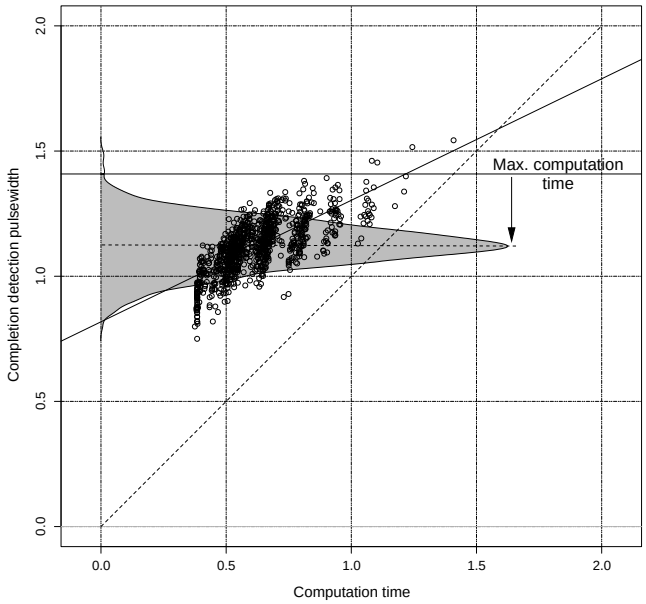



Fig. 7. Comparison of the actual computation time of the 8-bit adder and the monostable multivibrator response for randomly generated input vectors.

## V. Conclusions

In this work design of a novel completion detection system based on current sensing has been demonstrated. The proposed system can be integrated into the standard CMOS design flow very easily. The current sensing completion detection system can operate with subthreshold and superthreshold circuits without any modification. The proposed CDS creates pulses correlated to the actual computation time. It can be concluded from the simulation results that the operating speed of any asynchronous pipeline can be increased substantially by introducing computation based delay signal generation.

## References

[1] A. Wang and A. Chandrakasan, "A 180-mV subthreshold FFT processor using a minimum energy design methodology," *IEEE Journal of Solid-State Circuits*, vol. 40, no. 1, pp. 310–319, 2005.

[2] J. Kao, M. Miyazaki, and A. Chandrakasan, "A 175-mV multiply-accumulate unit using an adaptive supply voltage and body bias architecture," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 11, pp. 1545–1554, 2002.

[3] C.-I. Kim, H. Soeleman, and K. Roy, "Ultra-low-power DLMS adaptive filter for hearing aid applications," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 11, no. 6, pp. 1058–1067, 2003.

[4] E. Vittoz, *Low-Power Electronics Design*. CRC Press LLC, 2004, ch. 16.

[5] V. Tiwari, D. Singh, S. Rajgopal, G. Mehta, R. Patel, and F. Baez, "Reducing power in high-performance microprocessors," in *Design Automation Conference, 1998. Proceedings*, June 1998, pp. 732–737.

[6] J. Sparso and S. Furber, Eds., *Principles of Asynchronous Circuit Design - A Systems Perspective*. Kluwer Academic Publishers, 2001.

[7] M. E. Dean, D. L. Dill, and M. Horowitz, "Self-timed logic using current-sensing completion detection (CSCD)," in *International Conference on Computer Design*, 1991, pp. 187–191.

[8] H. Lampinen and O. Vainio, "Dynamically biased current sensor for current-sensing completion detection," in *Circuits and Systems, 2001. ISCAS 2001. The 2001 IEEE International Symposium on*, vol. 4, May 2001, pp. 394–397.

[9] A. S. Sedra and K. C. Smith, *Microelectronic Circuits*, 4th ed. Oxford University Press, 1998.