

Energy Efficiency Comparison of Asynchronous and Synchronous Circuits Operating in the Sub-threshold Regime

Omer Can Akgun (*), Yusuf Leblebici

Address: EPFL-STI-IEL-LSM Bat. ELD, Station 11
Lausanne, CH-1015, Switzerland

Phone: +41-21-693-6955, Fax: +41-21-693-6959

E-mail: {omercan.akgun, yusuf.leblebici}@epfl.ch

Abstract—Digital circuits operating in the sub-threshold regime are able to perform minimum energy operation at a given delay. In the sub-threshold regime circuit delay, hence the leakage energy consumption depends on the supply voltage exponentially. By reducing the idle time of the circuit both the supply voltage that realizes minimum energy operation and the energy consumption can be reduced.

This paper presents an in-depth comparison of synchronous and asynchronous techniques in the sub-threshold operating regime for their energy efficiency. First, transistor level accurate high level sub-threshold energy consumption model is developed for both techniques. Afterwards, using the model energy consumption reduction due to the asynchronous operation is investigated analytically. Different architectural improvements such as pipelining and parallelism are considered. The model has also been applied to the benchmark circuits for comparing real world energy consumption values.

From our analysis and simulations we have found out that asynchronous operation in the sub-threshold regime significantly lowers the supply voltage value that realizes the minimum energy operation and operating the digital circuits at a lower supply voltage value result in lower energy operation. Asynchronous operation resulted in energy consumption savings of up to 51% on the ISCAS85 benchmark circuits synthesized in a digital CMOS $0.18\mu\text{m}$ process.

I. INTRODUCTION

Power density and power consumption of complex digital systems have become a major concern during the recent years, both due to thermal concerns and due to limited battery life-time in mobile applications. Any significant reduction in power dissipation can only be achieved by lowering the operating voltage of the circuits [1]. 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 sub-threshold (weak-inversion) regime.

In sub-threshold mode of operation, the supply voltage can be scaled aggressively and power dissipation can be decreased significantly. Sub-threshold operation of static CMOS logic was analyzed using the EKV model in Ref. [2]. In this analysis, it was shown that static CMOS logic can be operated with a supply voltage as low as 50 mV at ambient temperature. In Refs. [3] and [4] the occurrence of minimum energy point in the sub-threshold regime was shown. There are also several

successful implementations of digital circuits operating in the sub-threshold regime in the literature such as, an FFT processor that is operational down to 180 mV [5] and a sub-threshold SRAM which operates with a supply voltage of 160 mV [6].

Circuits operating at these extreme low supply voltages work at much lower speeds, as an example the FFT processor presented in Ref. [5] works with a maximum clock frequency of 10 kHz with a power supply of 350 mV. Even so, their extreme low power consumption results in excellent power delay product (PDP) values, making such circuits very interesting candidates for ultra-low power applications which do not have very high processing requirements.

In the sub-threshold regime the leakage energy of the circuit increases exponentially with decreasing supply voltage. At the same time the dynamic energy consumption of a circuit operating in the sub-threshold regime decreases quadratically with decreasing supply voltage. Therefore for circuits operating in the sub-threshold regime a supply voltage that minimizes the sum of the dynamic and leakage energy can be found. We will call this supply voltage value the *energy-minimum operating voltage* throughout the paper. In Section III we will show that the key to reduce the energy-minimum operating voltage is to lower the leakage energy of the circuit and that this is only realizable by employing asynchronous operation. We can employ asynchronous operation without changing the circuit structure.

One solution to reduce the leakage energy is to use asynchronous circuits which can be designed to work at average-case performance. During our analysis we assume that asynchronous circuits work at average-case performance. Recently in Ref. [7] asynchronous circuits were studied from a low power and energy efficient operation perspective. At least in theory, average case performance property would allow asynchronous sub-threshold circuits to work at an operating point of higher energy efficiency (lower energy-minimum operating voltage). There are many approaches to design asynchronous circuits, some of which have inherent completion detection capabilities. In this paper we focus on single-rail circuits, i.e., circuits that encode data using one wire per data bit. This is

the default scenario for synchronous design styles, and typical for low-power asynchronous design styles (Refs. [7], [8]). We assume that for single-rail CMOS circuits there is an external completion detection mechanism.

The remainder of this paper is organized as follows. Section II provides background information and motivation for asynchronous sub-threshold operation. In Section III we present an accurate analytical model for minimum energy operation of asynchronous circuits and synchronous circuits. In Section IV, we investigate the effects of architectural modifications on the energy consumption of the circuits, and finally conclusions are drawn in Section V.

II. MOTIVATION AND BACKGROUND

The total energy consumption of static CMOS digital circuits is given by the following well-known formula:

$$E_{total} = \underbrace{\alpha C_{load} V_{DD}^2}_{E_{dynamic}} + \underbrace{I_{leak} V_{DD} t_{leak}}_{E_{leakage}} + \underbrace{I_{peak} t_{sc} V_{DD}}_{E_{short-circuit}} \quad (1)$$

where $E_{dynamic}$ is the total dynamic energy consumed while charging the load capacitance C_{load} , with a switching probability of α . When the circuit is not switching there is some leakage energy $E_{leakage}$ that is consumed during the leakage time t_{leak} . In addition, when a switching occurs, for the duration of the switching time t_{sc} , when both nMOS and pMOS transistors are conducting, some short circuit energy $E_{short-circuit}$ will be consumed. In our energy consumption analysis we will neglect the contribution of the short circuit energy in the sub-threshold regime, as it is known to contribute only a small portion of the overall energy consumption [2].

From equation (1) it is immediately clear that the energy consumption of digital circuits can be reduced by lowering the supply voltage. It was first shown by Swanson as early as 1972 that CMOS digital operation can be realized with ultra-low supply voltages [9]. When the supply voltage is lowered aggressively, below the threshold voltage (V_T) of the MOS transistors, the digital circuit operates in the sub-threshold regime.

The current equations of MOS transistors operating in the sub-threshold regime are different than those working in the super-threshold regime. For an nMOS transistor, the drain current in the sub-threshold regime is given in Ref. [2] by

$$I_{DS} = I_S e^{\frac{V_{GS} - V_T}{n U_t}} \left(1 - e^{-\frac{V_{DS}}{U_t}} \right) \quad (2)$$

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. 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,

$$I_S = 2n\mu C_{ox} U_t^2 \frac{W}{L} \quad (3)$$

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

Due to the second term in (2), the drain current is 0 when $V_{DS} = 0$ but reaches its maximum value for a given V_{GS} and

saturates with V_{DS} values higher than a few U_t . As it is apparent from (2), the drain current of a MOS transistor operating in the sub-threshold regime shows exponential dependence on the gate-to-source, drain-to-source voltages, slope factor, and the operating temperature. This exponential dependence of the drain current on the node voltages causes near-exponential changes in the operating speed of the circuit as the supply voltage varies [2]. As the supply voltage is lowered in the sub-threshold regime, the circuit delay hence the leakage energy consumption increase exponentially and the switching energy decreases quadratically, resulting in an energy-minimum operating point to occur. This is in contrast to the super-threshold operation where an energy-minimum operating voltage does not occur. By operating asynchronously, both leakage and dynamic energy components can be reduced. The reduction in the leakage energy is due to the reduction of idle time of the circuit and the reduction in the dynamic energy is due to the moving of the energy-minimum operating voltage to a lower value as will be explained in Section III-A.

A. Effects of Process Variation on Sub-threshold Operation

Semiconductor fabrication like all the manufacturing processes is subject to both random and deterministic process variations. The effects of process variation has been studied extensively in the literature and they can be grouped into global and local variation [10].

Global variation affects all the devices on a single die equally and results in device characteristic variations between the dies. On the other hand local variation affects the devices on the same die and consists of both random and systematic components. In this part the effects of the random threshold voltage (V_T) variation on the circuit operation in the sub-threshold regime will be shown.

In probability and statistics, the log-normal distribution is defined as the probability distribution of any random variable whose logarithm is normally distributed. The log-normal distribution occurs frequently in sub-threshold circuit design due to the fact that the current depends exponentially on V_T and it is assumed that V_T is normally distributed in local process variation. A basic definition of the log-normal function is given in Appendix I.

Because the sub-threshold current depends on the V_T exponentially, the sub-threshold current and parameters that have a first order relationship with the sub-threshold current such as leakage energy and circuit delay show a log-normal distribution under the assumption of normal distribution of V_T mismatch.

The delay variation for a single inverter from a standard cell library in a 0.18 μm process is shown in Figure 1. The data for the delay variation is gathered from 1000 point Monte Carlo simulations. As it can be seen from the figure, the delay variation is minimal at the nominal supply voltage of the process (VDD=1.8V) and has a Gaussian distribution as expected. On the other hand in the sub-threshold operating regime the delay has a log-normal distribution with a long tail on the right (VDD=0.3V). This kind of distribution curve implies that the below average delays deviate slightly from the

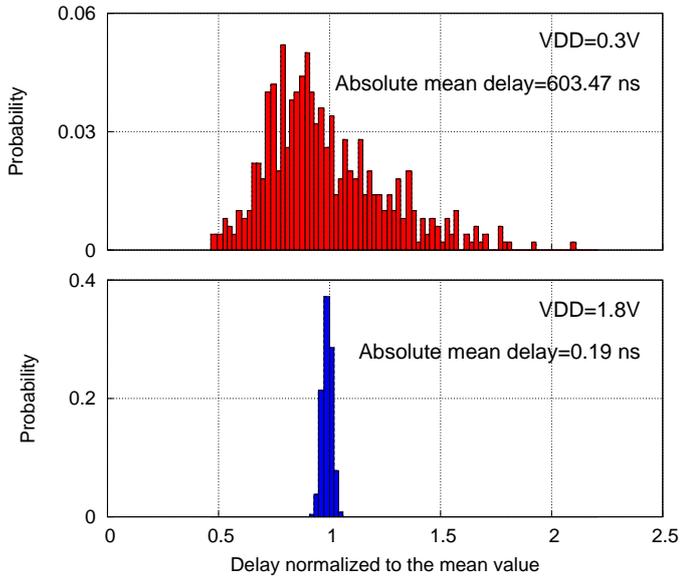


Fig. 1. Comparison of the delay times for a single inverter under V_T variation. The variation of the average delay is gathered from SPICE level Monte Carlo simulations.

expected value while above average delays can be as high as several times the mean.

For comparing the effects of supply voltage scaling on the circuit performance parameters, the coefficient of variation is a better metric. The coefficient of variation (CV) is defined as the measure of dispersion of a probability distribution and it is defined as the ratio of the standard deviation to the mean and is given by

$$c_v = \frac{\sigma}{\mu} \quad (4)$$

The coefficient of variation for various supply voltage values is shown in Figure 2. The sub-threshold operating regime can be easily distinguished on the figure for the supply voltage values less than or equal to 0.4V. Above 0.4V, the delay distribution is Gaussian and dispersion is much less than the sub-threshold supply voltages.

From the above analysis it can be concluded that for reliable operation in the sub-threshold regime, synchronous circuits must be over designed by a huge margin when compared to the superthreshold operation. On the other hand asynchronous circuits are insensitive to process variations. For an asynchronous system a job duration is irrelevant because the completion of operation is signaled to the following modules by the previous ones. So for better reliability and yield, asynchronous operation is a strong candidate for sub-threshold operation.

In Figure 3 the energy consumption overhead due to operating at a lower frequency for realizing higher yield (95%) for the ISCAS85 benchmark circuits is shown. Lowering the operation frequency for higher yield results in more leakage energy consumption and moving away from the energy-minimum operating point. From the SPICE level simulations we found out that the energy consumption penalty for targeting a higher yield can be up to 40.8% with respect to the typical operation case on the ISCAS85 benchmark circuits.

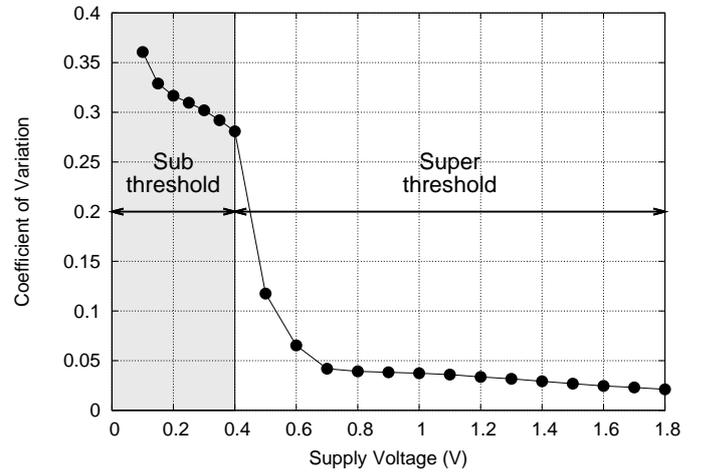


Fig. 2. Coefficient of variation for various supply voltage values. The sub-threshold and super-threshold regions are marked on the graph for $0.18\mu\text{m}$ digital CMOS process.

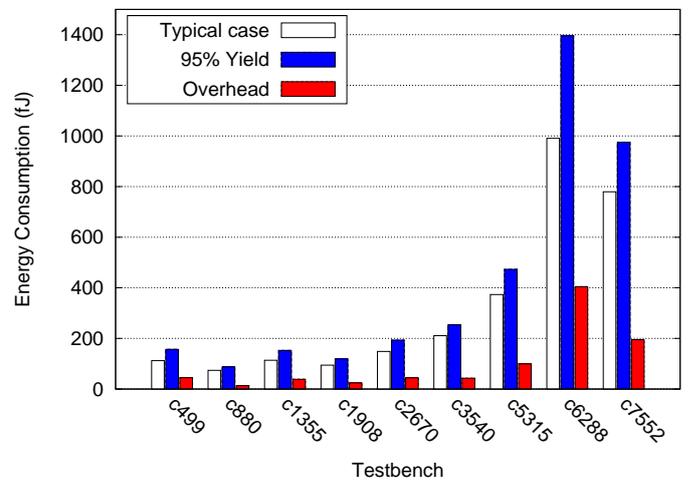


Fig. 3. Energy consumption overhead due to operating the sub-threshold circuits for correct operation yield with a 95% yield.

B. Asynchronous Operation

Asynchronous circuit design techniques employing completion detection are very attractive for digital circuits operating in the sub-threshold regime because of their better than worst case operation speeds, hence higher average throughput and lower leakage energy consumption. Asynchronous circuits rely on handshaking signals to communicate between functional units. Once an asynchronous block has been triggered (by the arrival of new data for example), it will start processing. At the end of the operation, the asynchronous sub-block will signal that it has completed its operation. After this time operation will continue depending on the specific asynchronous protocol that has been used.

Efficient asynchronous operation requires a reliable method to determine the time required to complete processing of information. There are many methods to realize asynchronous circuits, some of which have inherent completion detection capability such as methods based on n -out-of- m coding. In this paper, we will concentrate on a subset of asynchronous circuits that are based on asynchronous micro-pipelines first intro-

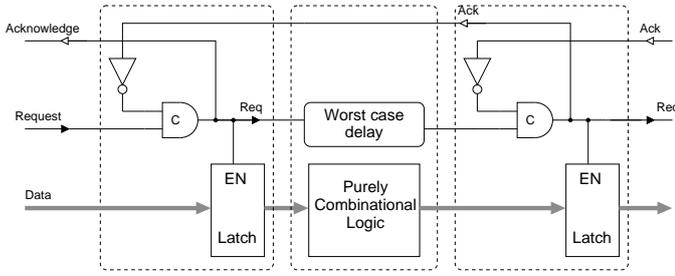


Fig. 4. 4-phase bundled data asynchronous pipeline. In this implementation control signals are delayed by a fixed amount equal to the critical path of the purely combinational block (after Ref. [8])

duced by Sutherland [11]. The asynchronous circuit model that we will use in the remainder of this paper is based on a single-rail (also known as bundled-data) four-phase handshake circuit shown in Figure 4 and is taken from Ref. [8, Figure 2.9].

In this type of asynchronous circuits, consecutive pipeline stages are separated using latches controlled by an asynchronous finite state machine (AFSM). The *req* line is used to signal that new data is available for processing. Once the pipeline stage is ready to process new data, the AFSM will acknowledge this request by using the *ack* line. This will enable the latch, and new data will become available for processing by the combinational circuit. The completion of this operation will generate a new *req* signal to the following stage. Implementations differ depending on the signaling scheme used between AFSMs. Without loss of generality we will use the four-phase signaling scheme in our examples.

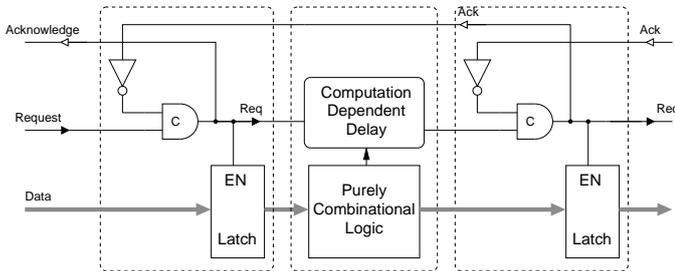


Fig. 5. 4-phase bundled data asynchronous pipeline with computation dependent delay. In this implementation control signals are delayed by a variable delay generated by the external completion detection circuit.

Traditional implementations of this circuit frequently use a *matched delay line* that has been engineered to have a delay that corresponds to the worst case delay through the combinational circuit as shown in Figure 4. There are obvious disadvantages of using a fixed delay element for both performance and energy consumption reasons, especially for coarse grained pipeline stages, where there is substantial variation in the operating speed depending on the input data switching probability. By operating such systems in a fixed delay fashion, unnecessary leakage energy consumption and throughput degradation occur. As it will be shown in Section III, operating asynchronous circuits with completion detection results in lower energy-minimum operating voltage hence lower energy consumption. In the remainder of this paper we will investigate the benefits of asynchronous operation

with accompanying completion detection circuitry as shown in Figure 5.

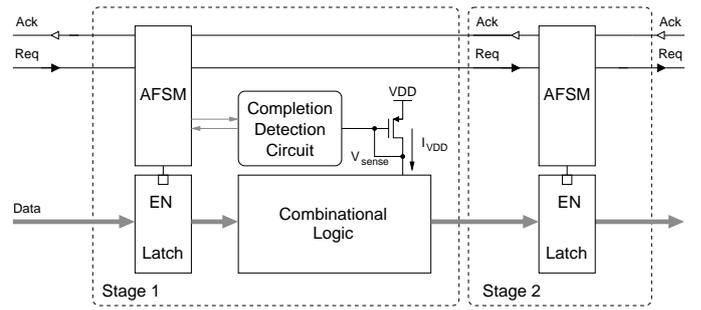


Fig. 6. General block diagram of the completion detection system. The system consist of an AFSM, completion detection circuitry and a sensor transistor (After Ref. [12]).

In Refs. [12] and [13], the authors presented a current-sensing completion detection system. The system can sense the currents in the pA to nA range, making the proposed technique applicable to sub-threshold circuits. Figure 6 shows the general block diagram of the current sensing based completion detection system. The completion detection system consist of AFSM, completion detection circuit, which consist of pulse generators and an AC-coupled amplifier, and a single pMOS transistor. The pMOS transistor is used for sensing the dynamic current consumption of the combinational logic block. The sensed signal is amplified inside the completion detection circuit and the amplified signal is used for setting the width of the generated pulse. AFSM is triggered by the generated pulse and then request and acknowledge signals are generated.

The waveforms in Figure 7 show the operation of the completion detection system for three consecutive cycles. In the shown example the completion detection circuit is used for generating the completion detection signals of a 16-bit adder. *DataInA* and *DataInB* are the inputs to the adder, *DataOut* is the adder output, *mmout* is the completion detection signal and *latch/trigger* signal is the latching signal controlling the latches at the output of the adder. In Figure 7 the *switching* phase of the first computation continues for a long time ($51.753 \mu s$). During the second computation, the input data does not change, and consequently there is no *switching* activity within the combinational block. In this case a fast completion detection signal is generated by the circuit ($12.334 \mu s$). The width of this fast signal is set by the internal design of the completion detection circuit. The last operation is much shorter than the first one and the sensed voltage signal V_{sense} settles more quickly to its final value, resulting in a shorter pulse ($31.709 \mu s$) when compared to the first generated pulse.

Presented completion detection circuit is shown as a proof of concept example of completion detection in the sub-threshold regime. For the given example the energy consumption of the completion detection circuitry is 41fJ per calculation at a supply voltage of 0.4V. This value is equal to the energy consumption of 40 inverter gates. Extra energy consumption due to the completion detection system defines

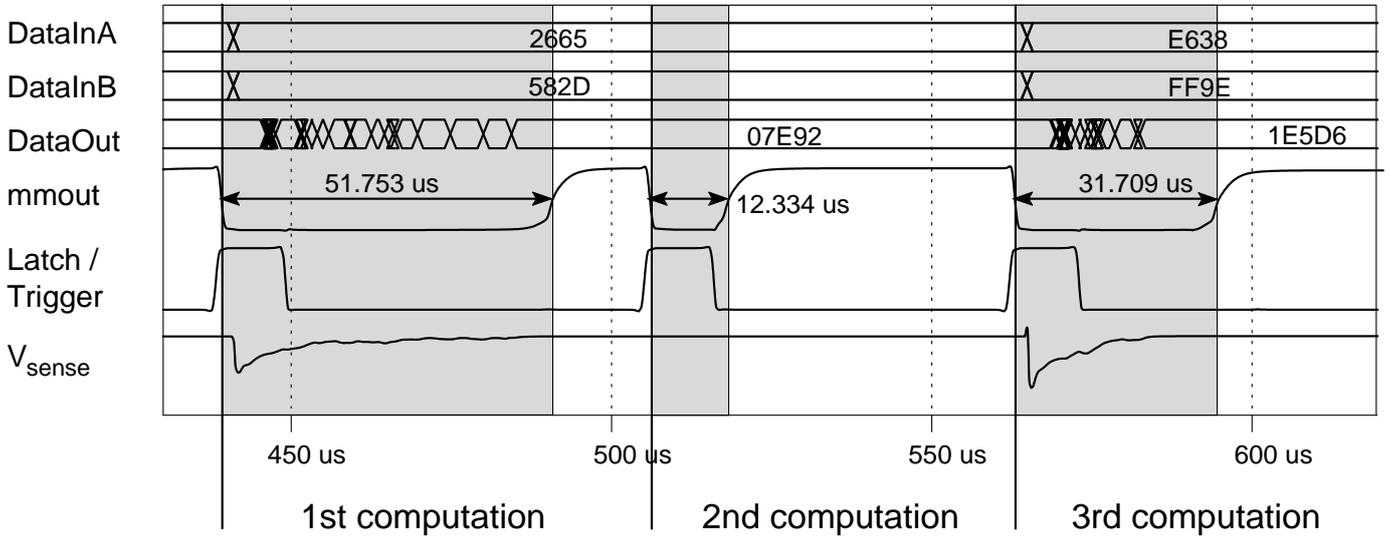


Fig. 7. Operation of the completion detection circuitry. Three consecutive computations are shown. The time differences between the computations are emphasized (After Ref. [12]).

the lower bound for the size of the circuits the shown completion detection solution can be applied. If the size of the circuit working with completion detection is too small, then the energy consumption overhead due to the extra circuit will overcome the benefits of asynchronous operation. In this paper we will work on isolated logic blocks to be able to fairly compare synchronous and asynchronous operations in terms of sub-threshold energy consumption. We will neither include the effects of energy consumption reduction techniques specific to synchronous or asynchronous operations nor the energy consumption overhead, such as control path and clock tree energy consumption, specific to the two mentioned domains.

Assumption-1: As soon as the asynchronous block finishes processing the current data, a new data input can be applied, i.e., there is no idle time between data inputs to the asynchronous block.

Assumption-2: The energy consumption and processing delays of the circuit per computation are randomly distributed. This assumption is guaranteed by applying a randomly distributed data set to the input of the circuit.

Assumption-3: Synchronous circuits work at their maximum speed, i.e., clocked at a speed equal to their critical path delay.

III. SUB-THRESHOLD ENERGY CONSUMPTION MODEL

In this section an energy model for asynchronous systems working in the sub-threshold regime is presented. In the asynchronous systems, the operation of the system is both dictated by the switching and delay properties of the system under investigation and the external request and acknowledge signals. During the development of the model for simplifying the mathematical operations and derivation of the model, following assumptions are made:

Assumptions 1 and 3 guarantee that synchronous and asynchronous operations can be compared fairly in terms of energy consumption. Together they guarantee that circuits run at their maximum speeds possible, hence consuming minimum leakage energy and working at the lowest energy-minimum operating voltage possible. Assumption-2 is used for simplifying the created energy model. As long as the energy consumption and processing delay of the circuit are randomly distributed with a mean, energy consumption model can be applied to any circuit that is operated with an arbitrary data set. While deriving the energy consumption model, assumptions made will be emphasized wherever required.

The energy consumption model presented in this section is comparable to other sub-threshold energy consumption models that have been published previously ([2]–[4]). All the models that were developed focused on synchronous sub-threshold operation. In Ref. [2] Vittoz investigated and proved the energy-minimum operation property of sub-threshold logic. In the model developed, an expression for the energy-minimum

operating voltage was not derived and energy-minimum operating point was shown by numerically inverting the duty factor for minimum energy numerically. In Ref. [3] occurrence of the energy-minimum operating voltage was shown again but the energy-minimum operating voltage equation was solved by curve fitting. In Ref. [4] sub-threshold energy-minimum operating voltage was solved analytically. In Calhoun's model average switched capacitance and average leakage current were specified as parameters and they were extracted from SPICE level simulation results. As will be shown in this section, we developed our model for asynchronous operation and later extended to include synchronous operation as well. Another benefit of the developed model is that it gives accurate results without requiring computation and time intensive SPICE simulations, making the model usable for all phases of the design.

A. Energy Model

The conceptual operation of an asynchronous block is shown in Figure 8a. In this example, the circuit is observed for an arbitrary time frame T , where four distinct sets of input data are processed. The time spans where the circuit is purely *leaking* (waiting for handshake completion) are denoted with T_{Li} and the time spans where the circuit is both *switching* (processing input data) and *leaking* are denoted with T_{Si} . In this diagram we assume that as soon as the *RequestOut* signal is lowered *RequestIn* goes high (Assumption-1) and the stage begins the *switching* phase; so all the purely leakage time spans (T_{Li} s) are fixed, equal, and represent the asynchronous communication overhead.

In the development of the asynchronous model, we assume a total of N operations take place during the arbitrarily long observation time frame T . Each switching time T_{Si} within the time frame can be different than each other depending on the switching properties of the circuit and the applied data to the circuit input.

Elaborating general energy consumption equation (1), the dynamic energy consumption during the i th time span can be expressed as

$$E_{dynamic_i} = e_i C_{tot} V_{DD}^2 \quad (5)$$

where e_i is a scaling parameter that defines the switching property of the circuit for a specific input data transition and C_{tot} is the maximum possible switched capacitance of the circuit. The switching energy scaling parameter e_i is in the range $[0, 1]$ and without loss of generality can be expressed as a single value in a random process e (Assumption-2). By defining the e as a random process, we can define a mean μ_e for it and using this mean the average dynamic energy for N calculations can be expressed as

$$E_{dynamic} = N \mu_e C_{tot} V_{DD}^2 \quad (6)$$

In (5) and (6), the total capacitance C_{tot} can be normalized in terms of the total inverter capacitance using a capacitance scaling factor $k_{cap-logic}$ as $C_{tot} = k_{cap-logic} C_{inv}$ where C_{inv} is the switched capacitance of an inverter.

Assuming even during the *switching* time, most of the cells in the circuit are leaking, the leakage energy consumption during the observation period T can be defined as

$$E_{leak} = k_{leak} I_0 V_{DD} T \quad (7)$$

where k_{leak} is the average leakage scaling factor of the circuit, I_0 is the leakage current of a single inverter. From (7) total average leakage current of the circuit can be calculated as $k_{leak} I_0$. In this equation, the average leakage parameter k_{leak} can be obtained from the synthesis results by summing the individual average leakage currents of the digital gates, where average leakage current is the mean of the leakage current for all the combinations of input vectors applied to the logic gate, and normalizing the result to the average leakage current of a single inverter.

Combining equations (1), (6) and (7), total energy consumption during the monitoring time frame T can be defined as

$$E_T = N \mu_e k_{cap-logic} C_{inv} V_{DD}^2 + k_{leak} I_0 V_{DD} T \quad (8)$$

From Figure 8a, the total time spent during switching is

$$T_S = \sum_{i=1}^S T_{s_i} \quad (9)$$

and based on the switching/timing statistics of the circuit we are analyzing any switching activity (T_{s_i}) in Figure 8a can be defined as

$$T_{s_i} = d_i k_{crit} T_{sw_inv} \quad (10)$$

where d_i is a scaling parameter that defines the delay properties of the circuit processing the current data, k_{crit} is a coefficient that defines the critical path delay of the circuit in terms of the inverter delay and T_{sw_inv} is the delay of an inverter. The scaling parameter d_i can take any value in the range $[0, 1]$. Like e , if the d is modeled as a random process with the mean μ_d (Assumption-2), the total time spent during switching can be calculated approximately as

$$T_S = N \mu_d k_{crit} T_{sw_inv} \quad (11)$$

During our observation frame T , N switchings and N handshakes take place, so T can be expressed as

$$\begin{aligned} T &= T_S + T_L \\ &= N \mu_d k_{crit} T_{sw_inv} + N k_{com_oh} k_{crit} T_{sw_inv} \end{aligned} \quad (12)$$

where k_{com_oh} is a parameter defining the overhead caused by the asynchronous communication in terms of the critical path delay of the purely combinational logic block. The delay of an inverter working in the sub-threshold regime is given in Ref. [2] as

$$T_{sw_inv} = \frac{C_{inv} V_{DD}}{I_0 e^{V_{DD}/(nU_t)}} \quad (13)$$

By introducing (13) into (12), we get the total observation time as

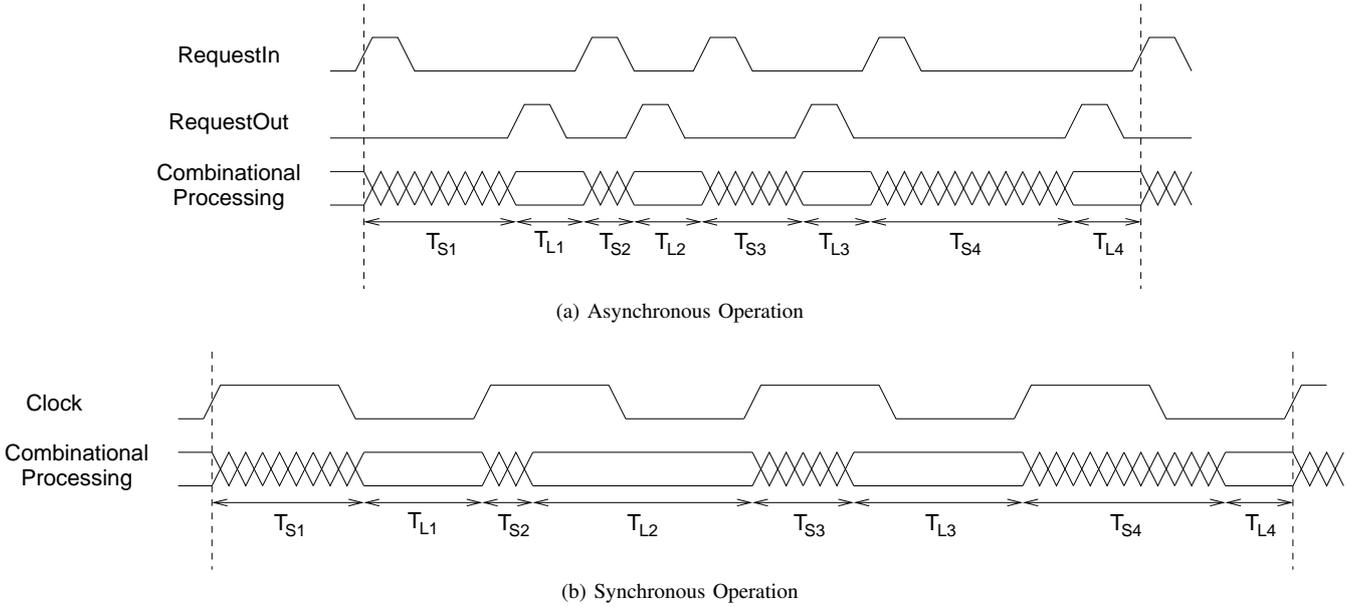


Fig. 8. Timing diagram showing the same logic block operating in synchronous and asynchronous modes.

$$T = Nk_{crit} \frac{C_{inv}V_{DD}}{I_0 e^{V_{DD}/(nU_t)}} (\mu_d + k_{com_{oh}}) \quad (14)$$

and by introducing (14) into (8), we get the final total energy consumption equation for N switchings as

$$E_T = NC_{inv}V_{DD}^2 \left[\mu_e k_{cap-logic} + k_{crit} k_{leak} (\mu_d + k_{com_{oh}}) e^{-V_{DD}/(nU_t)} \right] \quad (15)$$

By setting $N = 1$ in (15), average energy consumption per operation can be found. The optimal operating voltage for minimum energy operation can be found by taking the derivative of (15) with respect to V_{DD} , equating the result to 0, and solving for V_{DD} . The energy-minimum operating voltage is given in (16)

$$V_{opt-async} = 2nU_t - nU_t W_{-1} \left[-\frac{2e^2 k_{cap-logic} \mu_e}{k_{crit} k_{leak} (k_{com_{oh}} + \mu_d)} \right] \quad (16)$$

where W_{-1} is the -1 branch of the LambertW function. LambertW function is examined thoroughly in Ref. [14] and briefly explained in the Appendix II. All the k -parameters in (15) and (16) can be found from the synthesis results of the digital circuit, and the μ -parameters can be found after running switch level (synthesized Verilog) simulations. Hence the total simulation time for characterizing the sub-threshold performance of the circuit can be reduced greatly when compared to the SPICE-level simulation.

A similar modeling approach can be also taken for modeling a synchronous system operating as shown in Figure 8b. By assuming only one set of data is processed during a clock period ($N = 1$) and the clock period is equal to the critical

path of the logic circuit ($T = k_{crit} T_{sw,inv}$), the energy per operation can be derived from equations (14) and (15) as

$$E_T = C_{inv} V_{DD}^2 \left[\mu_e k_{cap-logic} + k_{crit} k_{leak} e^{-V_{DD}/(nU_t)} \right] \quad (17)$$

As in the asynchronous case, by taking the derivative of (17), equating the result to 0 and solving for V_{DD} , we get the optimum voltage that realizes the minimum energy operation as

$$V_{opt-sync} = 2nU_t - nU_t W_{-1} \left[-\frac{2e^2 k_{cap-logic} \mu_e}{k_{crit} k_{leak}} \right] \quad (18)$$

To be able to compare the energy consumption of the synchronous and asynchronous circuits at their energy-minimum operating points, the V_{DD} parameters in (15) and (17) are replaced by the optimum voltages $V_{opt-async}$ and $V_{opt-sync}$, respectively. The resulting energy equations depend only on the circuit implementation related k -parameters and the switching/delay properties of the circuit.

Figure 9 shows the energy profile of our randomly generated test circuit at a switching/delay mean of 0.1 for both asynchronous and synchronous operation. The k -parameters of the test circuit were chosen such that the circuit has energy consumption equivalent to 1000 inverter gates with a drive capability of 1, and the critical path was chosen to be 25 inverter delays. In the calculations made on the randomly generated circuits unless otherwise noted, the communication overhead parameter $k_{com_{oh}}$ is taken as 0.1. The optimum operating voltages of the same circuit for synchronous and asynchronous operations for the specified mean values occur at 210mV and 170mV, respectively.

When the energy-minimum operating voltages for the asynchronous and synchronous cases, (16) and (18) respectively,

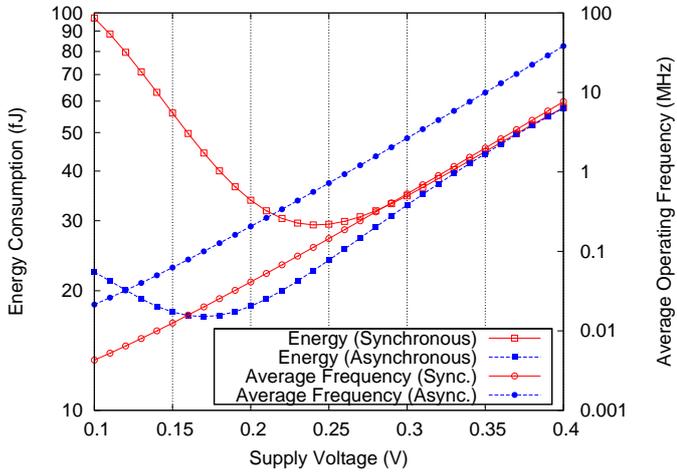


Fig. 9. Total energy consumption and average operation frequency for changing supply voltage values for a switching/delay mean value of 0.1. The energy-minimum operating points occur at different voltage values for synchronous and asynchronous cases.

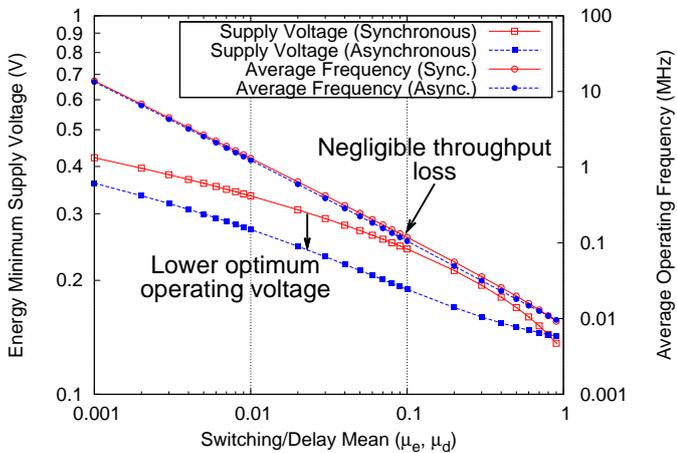


Fig. 10. Energy-minimum operating voltage and the respective throughput change for varying switching/delay mean values. The throughput loss due to the lower energy-minimum operating voltage is negligible.

are compared it is seen that the equations differ by the $1/(k_{com_{oh}} + \mu_d)$ factor in the argument of the LambertW function. Due to this difference, as long as $k_{com_{oh}} + \mu_d$ value is less than 1, the energy-minimum operating voltage for the asynchronous operation will always be lower than the synchronous operation, resulting in lower energy operation. $k_{com_{oh}} + \mu_d$ can be higher than 1 because of (1) high communication delay overhead and because of (2) majority of logic path delays being comparable to the critical path delay which increase μ_d to a value close to 1.

Advantages of operating in an asynchronous manner in the sub-threshold regime are two fold for energy consumption reduction. First the leakage energy is reduced by reducing the average time the circuit leaks. Due to the lower leakage energy consumption, the energy-minimum operating voltage value of the circuit decreases and the circuit can be operated at a lower energy-minimum operating voltage. This reduction in the supply voltage effectively reduces the switching energy consumption. This can be seen in the plot of energy-minimum

TABLE I
PARAMETER VALUES EXTRACTED FROM THE SYNTHESIS RESULTS FOR THE ISCAS85 BENCHMARK CIRCUITS.

Testbench	$k_{cap-logic}$	k_{leak}	k_{crit}	μ_d	μ_e
c499	559	437	57	0.31	0.7
c880	568	207	78	0.32	0.41
c1355	588	410	62	0.28	0.68
c1908	418	300	78	0.41	0.69
c2670	897	454	70	0.48	0.54
c3540	1228	470	120	0.58	0.47
c5315	2080	990	86	0.43	0.56
c6288	4310	1051	307	0.62	0.61
c7552	2498	1576	194	0.44	0.74

operating voltage values and their respective throughput values at those voltage values (Figure 10). In the figure, energy-minimum operating voltages and the operating frequencies for changing switching/delay properties of our randomly generated circuit are shown. While working under the conditions defined by Assumptions 1, 2 and 3, the energy-minimum operating voltage of the asynchronous operation is lower, thus reducing the switching energy. The throughput worsens due to lower operating voltage but it is negligible in asynchronous operation because of better than worst case computation delay.

B. Real-world Application Comparisons

To investigate the effectiveness of asynchronous operation on real world applications for sub-threshold energy consumption reduction, we extracted the model parameters from synthesized ISCAS85 benchmark circuits and their switch level simulations. Using the extracted parameters we calculated the energy-minimum operating voltages and respective energy consumption values. The values used in the calculations are given in Table I. $k_{cap-logic}$, k_{leak} and k_{crit} were extracted from the synthesis results and circuit delay and switching parameters μ_d and μ_e were extracted from switch level simulations. During the switch-level simulations randomly generated (Assumption-2) 10000 vectors were used.

The calculated energy-minimum operating voltage values for synchronous and asynchronous (under Assumption-1) operations are shown in Figure 11a. It is clearly shown in the plot that by employing asynchronous operation in the sub-threshold regime the energy-minimum operating voltage value can be reduced. For example for the test circuit *c1908*, working asynchronously results in the reduction of the energy-minimum operating voltage by 37.6%.

In Figure 11b the energy consumption of the test circuits at their respective energy-minimum operating voltages for synchronous and asynchronous cases are shown. Operating the circuits asynchronously in the sub-threshold regime result in considerable energy consumption savings as apparent from the figure. The reduction in the energy-minimum operating voltage and the total energy consumption are given in the second and third columns of Table II, respectively. By employing asynchronous operation in the sub-threshold regime, energy consumption reduction up to 51% has been observed.

While the lower-voltage operation might reduce the average throughput of the circuit as shown in Figure 12a, if ultimate

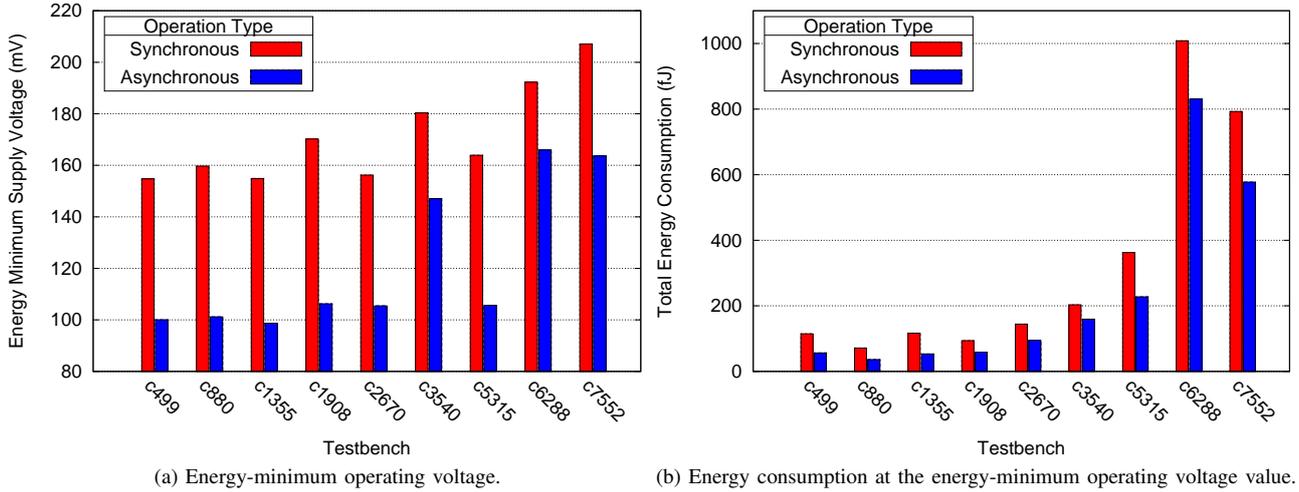


Fig. 11. Energy-minimum operating voltages and relative energy consumption of ISCAS85 benchmark circuits for both synchronous and asynchronous operation.

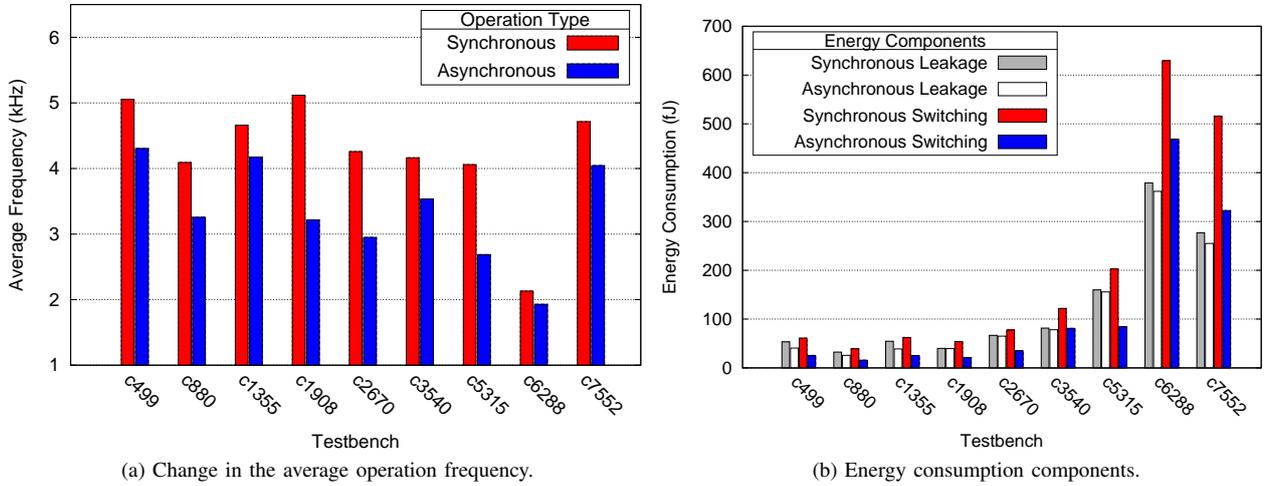


Fig. 12. Detailed average operation frequency change and energy consumption plots of the test circuits. The average operation frequency is lower due to lower energy-minimum operating voltage. Leakage and switching energy consumption elements in the energy plot are separated for both synchronous and asynchronous operation in the energy plot.

TABLE II

CHANGE IN THE ENERGY-MINIMUM OPERATING VOLTAGE, ENERGY CONSUMPTION, AVERAGE THROUGHPUT AND ENERGY DELAY PRODUCT FOR THE ISCAS85 BENCHMARK CIRCUITS.

Testbench	Supply voltage reduction (%)	Energy consumption reduction (%)	Average throughput loss (%)	EDP reduction (%)
c499	35.4	51.0	14.8	42.4
c880	36.7	49.6	20.4	36.6
c1355	36.3	54.6	10.5	49.3
c1908	37.6	38.0	37.1	1.4
c2670	32.5	34.2	30.8	5.0
c3540	18.5	21.7	15.1	7.7
c5315	35.6	37.3	33.8	5.2
c6288	13.7	17.6	9.6	8.9
c7552	21.	27.2	14.2	15.1

minimum energy operation is targeted, asynchronous operation in the sub-threshold domain is the only option that allows the reduction of the supply voltage. Due to the reduction of the *energy-minimum operating voltage*, both the leakage and the switching energy components of the circuits are reduced. The changes in the mentioned components are shown

in Figure 12b. For all the circuits examined, which have different structures and switching/delay properties, the energy consumption in asynchronous operation is lower than the synchronous operation. In all the test circuits reduction in the switching energy consumption due to lower *energy-minimum operating voltage* is more significant than the leakage energy

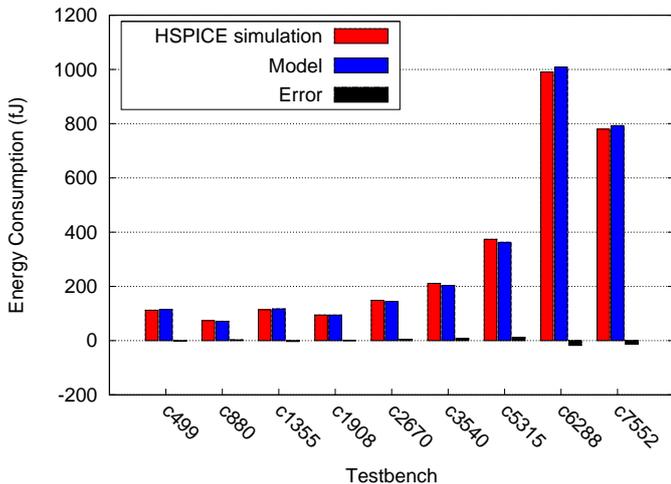


Fig. 13. Comparison of the energy consumption values gathered from the HSPICE simulation results and the model based on the synthesis results.

consumption reduction. For all the test circuits although the average throughput is lower for the asynchronous case (Figure 12a), leakage energy, which is inversely proportional to the average throughput of the circuit and linearly proportional to the supply voltage is always lower for the asynchronous case.

Although our aim is to realize minimum energy operation, we also checked the efficiency of the energy-delay trade-off in the sub-threshold regime between the synchronous and asynchronous operations. For checking the energy-delay trade-off efficiency we used the energy delay product (EDP) metric. The improvement in the EDP due to asynchronous operation is shown in the last column of Table II. Although the average throughput is lower for all the test circuits, because of greater reduction in energy consumption, the EDP of the asynchronous operation is always lower than the synchronous one. From this result it can be concluded that by operating asynchronously even in the worst case (*c1908*) the operation speed is efficiently traded for lower energy operation and for most of the cases significant reductions in energy consumption and EDP are realized.

C. Synthesis Results Based Model Accuracy

We ran simulations for analyzing the accuracy of using the proposed model employing the parameters gathered from the synthesis results and switch-level simulations. We gathered the energy consumption values for the test circuits from HSPICE transient simulations. During the simulations 1000 random vectors were applied to the test circuits working at their maximum allowed speed (speed set by the critical path of the circuit) at their synchronous energy-minimum operating voltages. Same test vectors were applied to the circuits during the switch-level simulations and their switching activity factors were calculated from the PrimeTime PX results. The comparison of the two methodologies in terms of accuracy are shown in Figure 13. From the comparison of the HSPICE simulation results and the developed model we found out that our methodology can predict the energy consumption values

TABLE III
ACCURACY COMPARISON FOR ENERGY CONSUMPTION AT THE ENERGY-MINIMUM VOLTAGE.

Testbench	Error (%)	Spice CPU Time	Model CPU Time
c499	-2.26	511.09	3.13
c880	3.49	351.35	4.16
c1355	-2.61	471.58	4.21
c1908	0.40	456.08	3.17
c2670	3.05	798.46	5.37
c3540	3.73	1845.05	5.36
c5315	2.98	2529.48	5.61
c6288	-1.79	10118.50	40.34
c7552	-1.70	4368.91	8.38

with less than 3.8% error for all the test circuits. The most important benefit of the developed model is the low run time when compared to SPICE level simulations. The error values and the CPU time spent during the simulations for all the test circuits are given in Table III for reference. From the table it can be seen that the developed model and the simulation methodology can be used for early design decisions in a more complex system with very high accuracy.

IV. ARCHITECTURAL IMPROVEMENTS

In this section architectural improvements such as parallelism and pipelining is investigated for their energy efficiency in the sub-threshold regime for both synchronous and asynchronous operations. During the application of the model developed in Section III to parallel and pipelined cases we focused on direct transformations from synchronous design to asynchronous design.

A. Parallelism

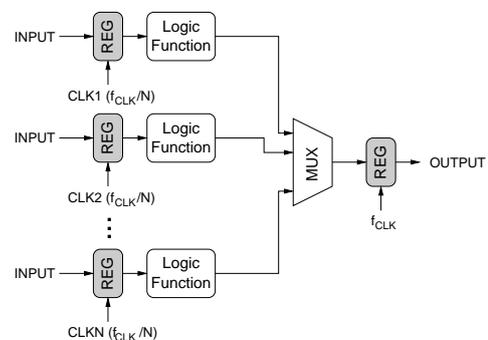


Fig. 14. Parallel system that consist of N copies of the same logic function running in parallel.

Parallelism (Figure 14) has been used in super-threshold logic circuits for either improving the throughput or for reducing the energy consumption by trading the throughput improvement for lower supply voltage, hence lower energy consumption [1]. In this section we study the effects of parallelism on the operation of synchronous and asynchronous circuits. Our study in parallel architectures focuses on so-called embarrassingly parallel operations that run N copies of the same logic function in parallel.

Assuming the parallelizing energy consumption overhead is negligible when compared to the energy consumption of

the main logic block, the asynchronous parallelized energy equation for one computation can be defined as

$$E_T = N\mu_e k_{cap-logic} C_{inv} V_{DD_{parM}}^2 + Mk_{leak} I_0 V_{DD_{parM}} T_{ref} \quad (19)$$

where M is the degree of parallelization (number of parallel blocks) and T_{ref} is the observation frame for the reference case. For trading the improvement in throughput for lower energy consumption, the supply voltage can be reduced such that the critical path delay is M times the reference delay of the circuit ($T_{parM} = MT_{ref}$). Replacing MT_{ref} by T_{parM} in (19) it can be easily seen that the resulting equation has the same form as (8) and the energy-minimum operating voltages are the same as (16) and (18) for the asynchronous and synchronous cases, respectively. This means regardless of the number of parallel copies of the same circuit, energy-minimum operating voltage is fixed and the same as the single copy case.

As a different case, the voltage can be lowered so that the throughput is fixed for M copies, resulting in energy sub-optimal operation. For this case the open form of timing relation in terms of the new and reference supply voltage is given by

$$k_{crit} \frac{C_{inv} V_{DD_{parM}}}{I_0 e^{V_{DD_{parM}}/(nU_t)}} = Mk_{crit} \frac{C_{inv} V_{DD}}{I_0 e^{V_{DD}/(nU_t)}} \quad (20)$$

Solving (20) for $V_{DD_{parM}}$ we can find the new supply voltage that realizes the same throughput for M parallel copies of the circuit as

$$V_{DD_{parM}} = -nU_t W_{-1} \left[-\frac{MV_{DD} e^{-V_{DD}/(nU_t)}}{nU_t} \right] \quad (21)$$

If the same procedure is followed for the calculation of the reduced voltage for synchronous operation as well, the same result as in (21) is found. To be able to see the effects of parallelization on the energy consumption of both synchronous and asynchronous operation, the switching/delay distribution mean is fixed to a value, and affects of changing the supply voltage is observed. During the calculations a communication overhead of two-tenth of the critical path of the main circuit for asynchronous operation is assumed. For the asynchronous and synchronous cases, the effects of parallelism are shown in Figure 15.

As it is clear from the plots, parallelism do not reduce the energy per operation for both the asynchronous and synchronous cases. In the ideal case without any parallelism overhead, the energy-minimum operating point moves but still results in the same minimum energy consumption. This comes from the fact that while reducing the voltage for trading the throughput for lower energy consumption, delay of the circuits increase exponentially and the parallelization and supply voltage reduction only result in the shifting of the energy-minimum operating point to higher voltage values. Moving of the optimum operating voltage to higher values comes from the fact that the leakage energy increases more than the switching

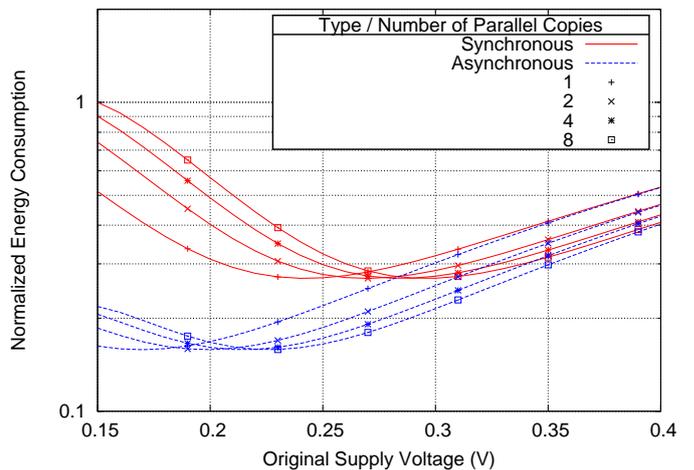


Fig. 15. Effects of parallelism on the energy consumption.

energy with parallelization and the optimum operating point for the system results in a relatively faster operating region while the switching and leakage energy consumptions are equal. Based on the presented results, the only benefit of parallelization in the sub-threshold regime is for increasing the throughput of the circuit. Once the energy-minimum operating voltage of the circuit is known, for increasing the throughput of the circuit, parallel copies of the circuit should be employed. This technique was employed by Sze et al. in Ref. [15] for realizing a UWB baseband processor operating with a supply voltage of 0.4V.

When compared to the parallelized case, the energy consumption reduction due to the asynchronous operation for different levels of parallelization is the same regardless of the number of parallel blocks for a fixed communication overhead making asynchronous sub-threshold operation more energy efficient for the required throughput. It should be emphasized that if asynchronous operation specific techniques, such as parallel (wagging) FIFOs [7], [16] are used during the implementation of parallel circuits, further energy consumption reduction in the sub-threshold operation is possible. We thank the anonymous reviewer for pointing out this detail.

B. Pipelining

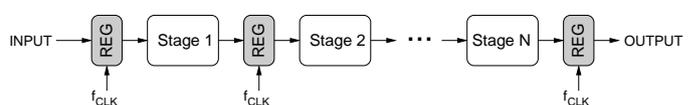


Fig. 16. Pipelined system with a pipelining level of N .

Another widely used architectural improvement for either improving the throughput or for reducing the energy consumption is pipelining (Figure 16). In this section effects of pipelining on the energy consumption of sub-threshold operation will be analyzed. As in the previous section we will focus on direct architectural transformations. During our simulations we employ the same memory elements, D type flip-flops in our case, for both synchronous and asynchronous

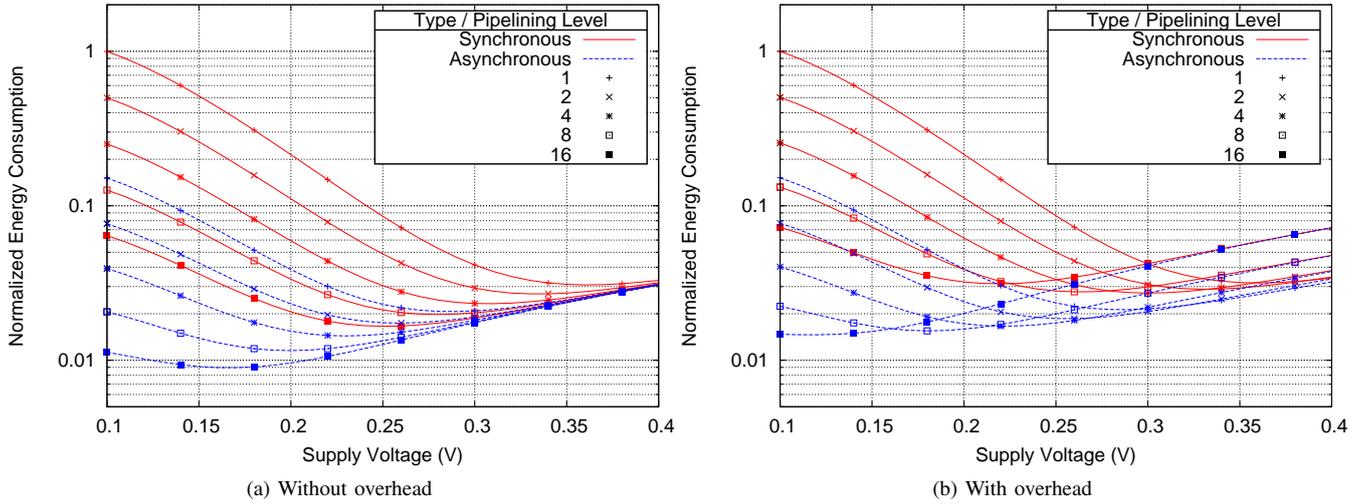


Fig. 17. Energy consumption calculations for different levels of pipelining.

operation. Latches, which are usually used as memory elements in asynchronous design increase the energy efficiency of asynchronous designs when compared to flip-flop based synchronous design. Although latches can also be used for synchronous design as well, their inclusion in the design is costly in terms of design and verification time and is not the common practice. To be able to investigate the improvement in energy consumption solely due to the average-case performance property of asynchronous operation and not to give asynchronous design an unfair advantage, we will assume that same type of memory elements are used for pipelining for both cases.

Unlike employing the parallel copies of the circuit, the hardware overhead due to the pipelining is not negligible and should be included in the energy consumption equations. The hardware overhead in pipelining is coming from the memory elements for dividing the circuit into equal-delay parts and their related routing capacitance. In our analysis we will only include the extra switched capacitance of the memory elements. Following the pipelining analysis methodology presented in Ref. [17] for super-threshold operation and adapting it for sub-threshold operation, the asynchronous energy consumption for a single computation with a pipelining level of N can be defined as

$$E_T = C_{inv}V^2 \left(\mu_e k_{cap-logic} + \mu_e B k_{cap-flop} N^\rho + \frac{k_{crit}}{N} (k_{leak} + B k_{leak-flop} N^\rho) (\mu_d + k_{comm_{oh}}) \right) \quad (22)$$

where B is the bit-width of the circuit, $k_{cap-flop}$ is the extra capacitance factor due to a single memory element, $k_{leak-flop}$ is the average leakage factor of a single memory element and ρ is the memory element growth factor that is super-linearly proportional to the pipelining depth N [17]. If the number of memory elements grow linearly with the number of pipelining stages, $\rho = 1$. But it has been shown in Ref. [18] that the number of memory elements grow super-linearly with

increasing pipelining, so in our analysis we will assume that $\rho = 1.2$. It should still be noted that ρ parameter depends on the circuit structure and is a design-specific parameter.

Applying the same methodology to the synchronous case, we get the synchronous pipelined energy consumption as

$$E_T = C_{inv}V^2 \left(\mu_e k_{cap-logic} + \mu_e B k_{cap-flop} N^\rho + \frac{k_{crit}}{N} (k_{leak} + B k_{leak-flop} N^\rho) \right) \quad (23)$$

In equations (22) and (23), $k_{cap-flop}$ and $k_{leak-flop}$ can be estimated from the standard cell library specifications, making the model usable just after the synthesis of the circuit. The calculation results for both synchronous and asynchronous pipelining cases with and without the overhead due to the memory elements are shown in Figure 17. The following deductions from the calculations and the figure can be made:

- Although without overhead significant energy consumption can be realized for synchronous operation, when the overhead is added in the calculations, the energy consumption reduction is marginal.
- Asynchronous operation can benefit from pipelining more than the synchronous operation and in the $0.18\mu\text{m}$ CMOS technology the limits of pipelining is set by the lowest working voltage of the memory elements (0.1V).

Calculations are also made for analyzing the energy consumption reduction of asynchronous operation with respect to the synchronous operation for different levels of pipelining for changing switching/delay probabilities. The results of these calculations are presented in Figure 18. The top curve in the plot shows a pipelining level of 16 and by employing pipelining at this level, an energy consumption reduction of up to 49% can be realized with respect to synchronous operation. In the figure for switching/timing coefficients greater than 0.9, asynchronous energy consumption is higher. Higher energy consumption is due to increased leakage energy consumption which is a result of the asynchronous communication overhead. An analysis for the optimum pipelining for different

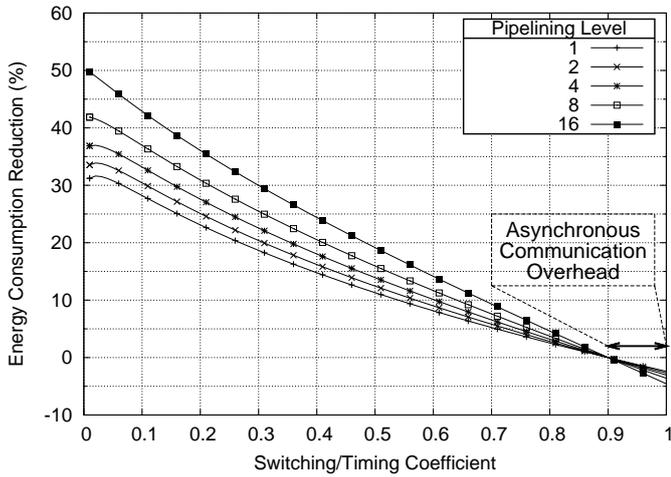


Fig. 18. Realizable energy gain with asynchronous operation for different pipelining levels. The region where synchronous operation results in lower energy consumption due to the asynchronous communication overhead is emphasized.

switching/delay probabilities of the circuit is also made. The results of the analysis are shown in Figure 19. For the calculations in Figure 18, same level of pipelining was applied to both synchronous and asynchronous operation at a specific switching/delay mean. On the other hand, for the results shown in Figure 19, minimum energy pipelining level and respective energy consumption for a specific switching/delay mean is calculated for both synchronous and asynchronous operations separately and the obtained results are compared. As it can be seen from the latter figure, asynchronous operation can realize an energy consumption reduction of up to 47% with respect to the synchronous operation with optimal pipelining.

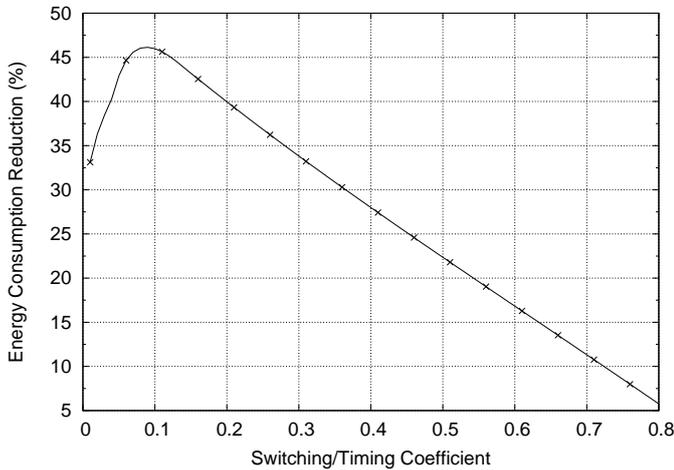


Fig. 19. Realizable energy gain for optimum pipelining at different switching/delay probabilities.

Similar numerical simulations were run for a relatively bigger system. The simulated system consist of 25000 NAND2 equivalent gates. The energy consumption and energy-delay product (EDP) of the system for both synchronous and asynchronous operation for changing pipelining levels are calculated. The results are shown in Figure 20. The optimum

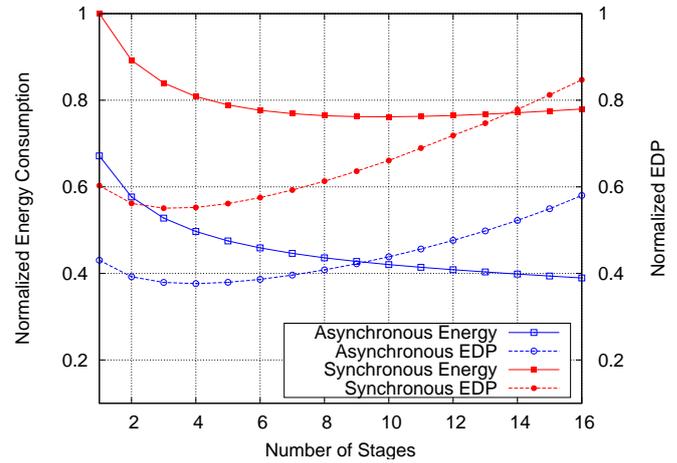


Fig. 20. Optimum number of stages for EDP/Energy minimization.

pipelining depth for synchronous and asynchronous minimum EDP is different, 3 and 4 levels, respectively. One important thing that should be noted is although there exists a minimum energy operation pipelining level for the synchronous case, for the asynchronous case no such limit exists and even lower minimum energy consumption occurs with increasing pipelining level. If synchronous and asynchronous operations are compared for their energy efficiency and EDP at their minimum EDP points, asynchronous operation consumes 41% less energy for a 32% lower EDP.

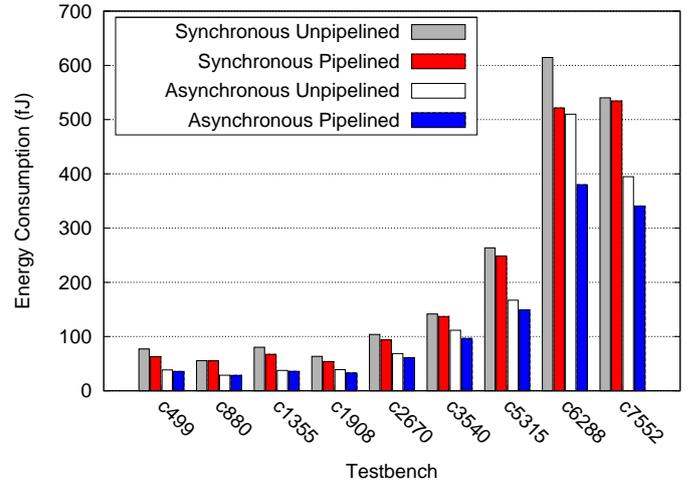


Fig. 21. Energy consumption reduction in ISCAS85 benchmark circuits due to pipelining. Both pipelined and unpipelined minimum energy consumption values for asynchronous and synchronous are shown.

We investigated the energy consumption reduction due to pipelining for real world applications on the benchmark test circuits. The energy consumption values for unpipelined and for pipelined cases are shown in Figure 21. The pipelined energy consumption values represent the consumption at the optimal pipelining level for minimum energy operation. In the figure the unpipelined case has flip-flops at the output pins, hence the energy consumption differences when compared to Figure 11b. The bit-width value (B) is chosen as the minimum of number of input pins, number of output pins or

the number of flip-flops that have a total capacitance of 25% of total switched capacitance of the purely combinational logic block. 25% value is chosen based on the complex industrial circuits and taken from Ref. [17]. The bit-width values used in the computations and the percentage energy consumption reduction are given in Figure IV. The percentage reduction values given in the table are relative to the unpipelined cases for synchronous and asynchronous operation separately. One thing that should be noted from Figure 21 is that for all the test circuits unpipelined asynchronous energy consumption is lower than the synchronous operation energy consumption for both pipelined and unpipelined cases. By applying pipelining on asynchronous circuits, even lower energy consumption can be realized.

V. CONCLUSIONS

In this paper an energy consumption model for sub-threshold operation is presented. The model is based on the circuit parameters that are obtainable from the standard cell synthesis results and switch level (synthesized Verilog) simulations. Same model is applicable to both synchronous and asynchronous operation with a slight modification. By using the proposed model, energy efficiency of the same circuit structure can be easily investigated in different manufacturing processes without the need for time and resource expensive SPICE level simulations.

Using the model the effects of architectural improvements on the sub-threshold circuit operation for synchronous and asynchronous cases have been investigated. It has been shown that the parallel processing in the sub-threshold domain does not reduce the energy per operation and should only be employed for increasing the throughput of the circuit. On the other hand, by employing pipelining, especially in the asynchronous operation, substantial energy reductions can be realized. It has been also shown that asynchronous operation in the simple case reduce the energy consumption by 43% on our randomly generated test circuit. Furthermore, by employing pipelining this reduction can be increased.

Same experiments are applied to ISCAS85 benchmark circuits for demonstrating achievable savings in energy consumption by operating asynchronously. It is found that for the basic case by operating the circuits asynchronously, energy consumption reductions of up to 51% and EDP reductions of up to 49.3% have been realized.

APPENDIX I

LOG-NORMAL DISTRIBUTION

The probability density function (PDF) of a log-normal distribution is characterized by the mean and the standard deviation of the variable's logarithm, μ and σ , respectively. The PDF of the log-normal distribution is given by

$$f(x, \mu, \sigma) = \frac{e^{-(\ln x - \mu)/(2\sigma^2)}}{x\sigma\sqrt{2\pi}} \quad (24)$$

and the mean and variance of the log-normal variable X are given by

$$E(X) = e^{\sigma^2/2 + \mu} \quad (25)$$

$$Var(X) = e^{\sigma^2 + 2\mu}(e^{\sigma^2} - 1) \quad (26)$$

APPENDIX II

LAMBERT W FUNCTION

In mathematics, the *Lambert W* function, $W = LambertW(x)$, gives the solution to the equation $We^W = x$. The plotted function is shown in Figure 22. W function has a branch point of order 2 at $x = -1/e$ and has two real solutions in the range $(-1/e, 0)$. W function assumes complex values for $x < -1/e$. In general, the branch satisfying $-1 \leq W(x)$ is denoted by $W_0(x)$ and the branch satisfying $W(x) \leq -1$ by $W_{-1}(x)$. The branch defined by $W_0(x)$ is called the *principal branch* of the W [14].

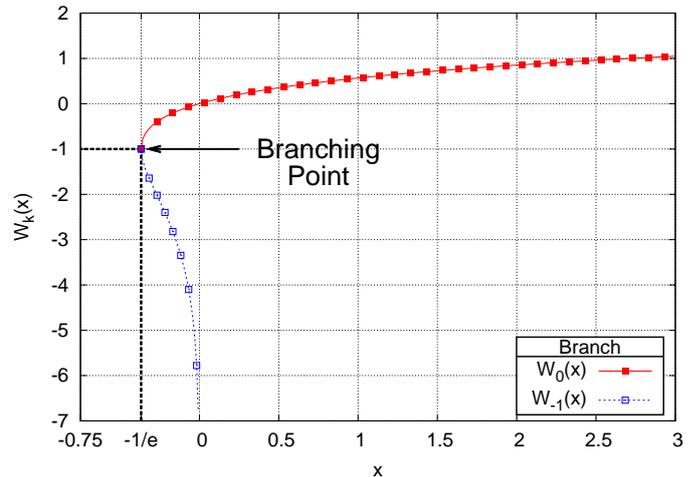


Fig. 22. Plot of *Lambert W* function showing the real valued branches.

REFERENCES

1. A. Chandrakasan, S. Sheng, and R. Brodersen, "Low-power CMOS digital design," *Solid-State Circuits, IEEE Journal of*, vol. 27, no. 4, pp. 473–484, 1992.
2. E. Vittoz, *Low-Power Electronics Design*. CRC Press LLC, 2004, ch. 16.
3. B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, "Theoretical and practical limits of dynamic voltage scaling," in *Proceedings of the 41st Annual Conference on Design Automation*. ACM New York, NY, USA, 2004, pp. 868–873.
4. B. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and sizing for minimum energy operation in subthreshold circuits," *Solid-State Circuits, IEEE Journal of*, vol. 40, no. 9, pp. 1778–1786, 2005.
5. 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.
6. J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mV robust schmitt trigger based subthreshold SRAM," *IEEE Journal of Solid-State Circuits*, vol. 42, no. 10, pp. 2303–2313, Oct. 2007.
7. P. Bearel and M. Roncken, "Low Power and Energy Efficient Asynchronous Design," *J. Low Power Electronics*, vol. 3, no. 3, pp. 234–253, 2007.
8. J. Sparso and S. Furber, Eds., *Principles of Asynchronous Circuit Design - A Systems Perspective*. Kluwer Academic Publishers, 2001.
9. R. Swanson and J. Meindl, "Ion-implanted complementary MOS transistors in low-voltage circuits," *Solid-State Circuits, IEEE Journal of*, vol. 7, no. 2, pp. 146–153, 1972.
10. K. A. Bowman, S. G. Duvall, and J. D. Meindl, "Impact of die-to-die and within-die parameter fluctuations on the maximum clock frequency distribution for gigascale integration," *IEEE Journal of Solid-State Circuits*, vol. 37, no. 2, pp. 183–190, Feb. 2002.
11. I. Sutherland, "Micropipelines," *Communications of the ACM*, vol. 32, no. 6, July 1989.
12. O. C. Akgun, F. K. Gurkaynak, and Y. Leblebici, "A current sensing completion detection method for asynchronous pipelines operating in the sub-threshold regime," *International Journal of Circuit Theory and Applications*, in press.

TABLE IV
BIT-WIDTH VALUES USED IN THE CALCULATIONS AND ENERGY CONSUMPTION REDUCTION DUE TO PIPELINING.

Testbench	Bit-width B	Synchronous Energy Reduction (%)	Asynchronous Energy Reduction (%)
c499	14	18.6	7.7
c880	14	0.8	1.3
c1355	15	16.2	4.9
c1908	10	15.0	15.8
c2670	22	9.8	11.0
c3540	22	3.3	13.6
c5315	52	5.9	10.8
c6288	32	15.1	25.5
c7552	62	1.1	13.7

13. O. C. Akgun, Y. Leblebici, and E. A. Vittoz, "Current sensing completion detection for subthreshold asynchronous circuits," in *Proceedings of the European Conference on Circuit Theory and Design 2007*, 2007, pp. 376–379.
14. R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, "On the LambertW function," *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
15. V. Sze and A. Chandrakasan, "A 0.4-V UWB baseband processor," *Proceedings of the 2007 international symposium on Low power electronics and design*, pp. 262–267, 2007.
16. E. Brunvand, "Low latency self-timed flow-through FIFOs," in *Proceedings of the 16th Conference on Advanced Research in VLSI (ARVLSI'95)*. IEEE Computer Society Washington, DC, USA, 1995.
17. N. Kim, T. Kgil, K. Bowman, V. De, and T. Mudge, "Total power-optimal pipelining and parallel processing under process variations in nanometer technology," in *Computer-Aided Design, 2005. ICCAD-2005. IEEE/ACM International Conference on*, 2005, pp. 535–540.
18. V. Srinivasan, D. Brooks, M. Gschwind, P. Bose, V. Zyuban, P. Strenski, P. Emma, I. Center, and N. Yorktown Heights, "Optimizing pipelines for power and performance," *Microarchitecture, 2002.(MICRO-35). Proceedings. 35th Annual IEEE/ACM International Symposium on*, pp. 333–344, 2002.