

# Serial Sub-threshold Circuits for Ultra-Low-Power Systems

Sudhanshu Khanna and Benton H. Calhoun

ECE Dept., University of Virginia, 351 McCormick Road, PO Box 400743, Charlottesville, VA  
<sudhanshu, bcalhoun>@email.virginia.edu

## ABSTRACT

This paper explores the use of serial circuits for ultra-low-power sub-threshold systems. A serial system leads to a smaller design and higher utilization, yielding 40% active energy, 15x active power, and 32x leakage power benefits. Further, we show that using a serial system in the sub-threshold regime decreases both active energy and leakage power even *at the same speed* as a parallel system. This is in sharp contrast to strong inversion, where larger bit widths give lower energy and power for the same delay. We identify the unique properties of sub-threshold operation that creates these differences.

## Categories and Subject Descriptors

B.6.1 [Logic Design]: Design Styles – *combinational logic, parallel circuits, sequential circuits.*

## General Terms

Performance, Design

## Keywords

Ultra low power, serial systems, bit width, sub-threshold, leakage.

## 1. INTRODUCTION

This paper compares serial and parallel structures for energy- and power-constrained applications that use sub-threshold operation and shows that unique properties of sub-threshold circuits and the applications they support make serial design more beneficial, whereas in strong inversion, parallel circuits are generally better. Many emerging applications require miniaturized electronics to operate for long lifetimes. These competing requirements place tight energy or power constraints on the circuits. Example applications include wireless microsensors, RFIDs, implantable electronics, micro vehicles, and body area sensor networks. Sub-threshold digital circuits that operate with a supply voltage,  $V_{DD}$ , less than the transistor threshold voltage,  $V_T$ , significantly reduce active and leakage energy, so they provide a promising approach to severely energy- or power-limited applications.

The low  $V_{DD}$  in sub-threshold circuits decreases the on-current and substantially slows circuit speed. However, most ultra low power (ULP)<sup>1</sup> applications have reduced performance needs that align with the lower operating frequencies that sub-threshold circuits can achieve. Furthermore, many ULP devices have a low duty cycle and spend large fractions of their lifetimes in sleep

---

<sup>1</sup> Without loss of generality, we use ULP to refer to systems with strict power or energy constraints.

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

ISLPED'09, August 19–21, 2009, San Francisco, California, USA.  
Copyright 2009 ACM 978-1-60558-684-7/09/08...\$10.00.

mode. This characteristic makes leakage power during the sleep mode extremely influential on the total system energy budget. Techniques to reduce leakage power, in both active and sleep modes, can improve the lifetime of ULP systems. Many sub-threshold system designs port standard architectures from strong inversion into the lower voltage domain, and the bit width of sub-threshold systems remains an unexplored knob.

In this paper, we examine the tradeoffs between serial and parallel component and system design. A serial implementation uses a smaller bit width, but must use more cycles to perform the same operation (e.g. a 32b addition) as a parallel system. The serial system uses less area, and thus it has less switched capacitance and leakage power. This fact is well known for strong inversion systems, but the delay overhead of serial systems often limits their use. In sub-threshold, different equations govern the energy-performance trade off, and we find that serial systems become preferred over a broad range of application requirements. Drawing from our analysis and observations, this paper makes the following key contributions:

- Quantifies major leakage savings from serialization
- Shows that active energy/operation of a serial system over multiple cycles is less than for a single cycle parallel system
- Shows that a sub-threshold serial implementation provides active energy and leakage power savings *at the same speed* as a parallel version, along with area savings
- Clarifies overhead tradeoffs of mixed serial/parallel systems

This paper is organized in the following manner. Section 2 briefly reviews sub-threshold circuit research and motivates the need to examine serialization in sub-threshold systems. Section 3 compares systems of varying bit width for minimum energy operation considering variable sleep time and performance constraints. Section 4 describes full systems and discusses tradeoffs that arise from the overhead of using serial or parallel components. Section 5 concludes the paper.

## 2. MOTIVATION FOR BIT WIDTH OPTIMIZATION

ULP systems require extremely energy efficient circuits to meet their stringent size and lifetime requirements. Advances in sub-threshold circuit techniques have made sub-threshold operation the prominent approach for ULP applications, but we argue in this paper that an opportunity exists to reconsider the bit width (e.g. degree of serialization) in sub-threshold systems to further reduce system energy and power consumption.

### 2.1.1 Energy vs. Power Constraints

A variety of emerging applications require long lifetimes and small form factors, which limit the amount of energy storage (e.g. battery size) and impose strict energy constraints. Sub-threshold circuits meet these application needs by minimizing energy per operation [1]-[3]. Chip demonstrations of successful logic [1][2], memory [4]-[7], DSPs [8][9], and processors [10]-[12] have

conclusively shown the viability of sub-threshold circuits and their ability to reduce energy consumption significantly.

ULP systems fall broadly into two categories: energy-constrained and power-constrained. Energy-constrained systems operate on a finite supply of energy (battery). Each operation draws some energy from the battery (active energy), so the key goal is to minimize energy per operation (E/op), which comprises both dynamic energy and active leakage energy.

Power-constrained applications operate from an effectively infinite energy source that supplies a limited amount of power. Examples include inductive power coupling (as in RFID) and power scavenging. Reducing the power consumption of a power-constrained circuit allows for more functionality, smaller coils, or longer range. Sub-threshold circuits minimize power by using the lowest possible  $V_{DD}$ .

Both energy- and power-constrained systems would benefit from less capacitance and leakage power. Lower leakage reduces active and idle energy and extends lifetime for energy constrained applications. Less active energy spent over a longer time reduces active power for power-constrained devices. Serial implementations potentially address both of these needs.

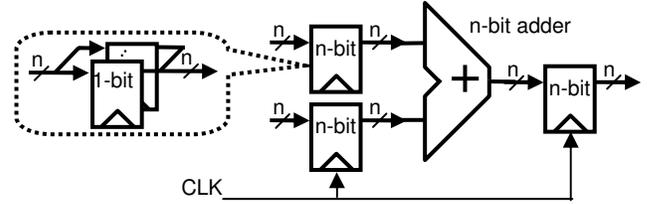
### 2.1.2 Leakage in Active and Sleep Modes

Many ULP applications deal with low duty cycle operation that leaves circuits in sleep mode for large fractions of the lifetime. For example, physiological metrics may require sampling every 0.25s (heart rate), every few minutes (blood pressure, temperature), or every few hours (temperature, glucose level). Structural health sensors may sample every day or every few days. We previously described the impact of active leakage on energy constrained systems, but sleep mode leakage can dominate the energy consumption of low duty cycle systems. A number of existing circuit techniques for reducing idle mode leakage can apply equally well in the sub-threshold regime. These include stacking, power gating, and reverse body bias. These techniques reduce sleep mode leakage current to a small fraction of the active mode leakage. Consequently, these techniques are limited in effectiveness by the amount of active mode leakage itself. A small fraction of the active mode leakage, integrated over a large sleep time, still poses a huge roadblock to extended system operation.

Since the active mode leakage of a digital circuit is roughly proportional to the number of leaking paths, serializing the implementation will lower active leakage. Since the active leakage is the starting point from which circuit level leakage reduction techniques begin to reduce standby leakage, this has the additional benefit of making those techniques more effective in standby by lowering the initial leakage amount. Less area in serial circuits also reduces switched capacitance. For these reasons, serial implementations deserve investigation for reducing both active and standby leakage in ULP systems. In this paper, we show the value of serialization for reducing both active mode and sleep mode leakage, resulting in substantial energy savings.

### 2.1.3 Serial vs. Parallel (or Bit Width)

So far, we have observed that serialization will save area and reduce both active and leakage power. This lowers leakage energy in the sleep mode; we explore the impact on active leakage and dynamic energy in Section 3.2, which is complicated by the execution in a serial design over multiple cycles. The observation



**Figure 1: An n-bit adder in a larger n-bit digital system.**

that serial designs reduce leakage power has appeared in many previous works that focus on strong inversion, for example [13].

The study of low power strong inversion DSPs in [13] shows that leakage power dominates more for longer sleep times (lower frequency), where serial implementations become the lowest power solution. In strong inversion, serial designs have significantly larger delay than parallel, which limits their use. Reference [13] also shows how distributed arithmetic offers efficient serial circuits and discusses how system redesign can specifically target serial topologies. In this paper, we investigate how serial and parallel architectures compare for sub-threshold circuits, where different equations govern the energy-delay tradeoff and where low frequency and long sleep time are the rule.

## 3. SYSTEM LEVEL BIT WIDTH

Shrinking circuit area reduces energy by lowering both switched capacitance and leakage power. This section compares systems with bit widths of 1, 16, and 32, using adders as a representation of a digital system. A bit width of 1 corresponds to a fully serial addition system. Such a system takes multiple cycles to complete the same operation (e.g. a 32b add) that a parallel system completes in a single cycle. Serial systems surely lower leakage power, but we must examine how they compare in sub-threshold in terms of delay and active mode energy.

### 3.1 Minimizing Energy

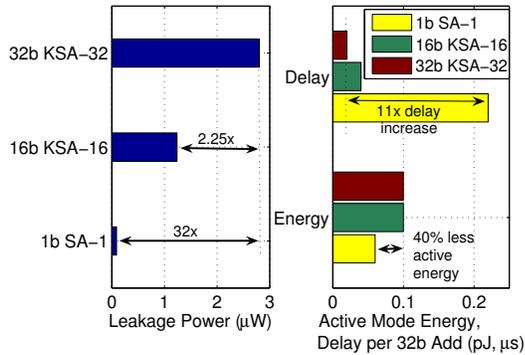
Sensor applications present design constraints that are skewed towards energy conservation, justifying sub-threshold operation. However, more savings can be achieved by recognizing that in this application space, sleep mode energy is dominant. Small system level bit width may introduce concerns for active mode operation, but offers significant benefits during the sleep mode. In this section we report the benefits of serial implementations for reducing energy in sub-threshold, starting with a scenario that excludes sleep mode and then accounting for sleep mode energy.

#### 3.1.1 Minimizing Active Energy

We use addition systems to represent digital systems of varying bit widths ( $n$ ): 1, 16 and 32. The sum of a serial addition depends on the current two operands and only one previously stored bit, the carry. The compared addition systems are as follows:

- 1b Serial Addition using a full adder and carry flop (1b SA-1)
- 16b addition using 16b Kogge-Stone Adder (KSA) and carry flop (16b KSA-16)
- 32b addition using a 32b Kogge-Stone Adder (32b KSA-32)

Figure 1 shows a generic n-bit addition system with an adder, two input registers, and an output register. We define our quality metric (E/op) as the amount of energy per 32b add, including energy in the adder, registers, and clock system. An n-bit addition system takes  $32/n$  clock cycles to complete this operation. Note



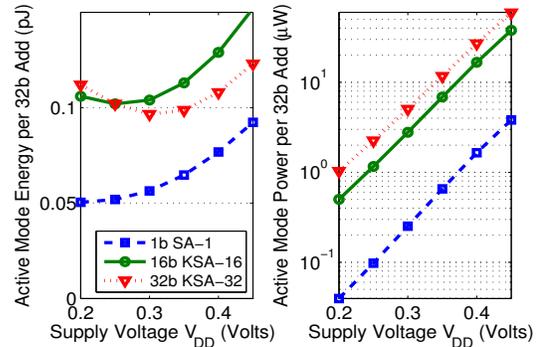
**Figure 2: (a) Active mode leakage power, (b) energy (including dynamic energy), and delay per 32b add for different addition systems. (22nm CMOS [15],  $V_{DD}=300\text{mV}$ ).**

that this corresponds to active mode energy, consisting of active mode leakage energy and dynamic energy.

Figure 2 compares the simulation results for leakage power, active energy, and delay per 32b add at 300mV, which closely approximates the optimal  $V_{DD}$  point for many sub-threshold systems [1]-[12]. Figure 2(a) shows that 1b and 16b systems demonstrate 32x and 2.25x leakage power reduction over a 32b system, respectively. Further, Figure 2(b) shows that the 1b SA-1 consumes 40% less active mode energy than the 32b KSA-32. In other words, total energy consumed by a 1b system over 32 cycles is less than that consumed by a 32b system in just one cycle. The reason for this is that complexity in terms of transistor count increases super linearly with bit width. The 1b SA-1 has higher resource utilization (each node in a 1b SA-1 is on the critical path) as compared to 32b KSA-32. Figure 2(b) also shows that the cost of this lower active energy and leakage power is longer delay, although additional delay may not be problematic for many ULP applications. The lesson is that multi-cycle operation does not increase active mode energy. Instead, small bit widths lower active mode energy. The results shown are for 22nm predictive technology models (PTMs) [15], but at 90nm, the trends remain the same. In summary, smaller bit width results in near linear leakage power savings and modest active mode energy reduction. The number of cycles per 32b add increases as bit width lowers. However, the critical path gets smaller with bit width, and thus the delay per 32b add does not change as fast as the bit width.

Figure 3(a) reports the active mode energy as a function of  $V_{DD}$  for the same 3 adder systems. Again, we see that 32b and 16b systems have similar active mode energy, while the 1b system has lower active mode energy across the  $V_{DD}$  range. The 32b Ripple Carry adder (not shown in graph) also showed much higher active mode energy than the 1b SA-1. In power constrained applications, serial systems prove quite beneficial because they spread the active energy over multiple cycles, lowering active power. Figure 3(b) shows how active power scales with  $V_{DD}$  for 1b, 16b and 32b addition systems. Across the  $V_{DD}$  range, the benefit of using a 1b system over a 32b system exceeds 15x in terms of active mode power. This enables us to put more processing power on a microsensor, enabling further potential power savings by pre-processing data to reduce power hungry wireless communication.

In summary, serial sub-threshold systems reduce active mode energy, lower active mode power by over 15x and cut leakage by up to 32x in our example circuits. In the next section, we will show how this translates to large sleep mode energy benefits.



**Figure 3: Active mode (a) energy per 32b add and (b) power for 1-, 16- and 32-bit addition systems as a function of  $V_{DD}$ .**

### 3.1.2 Accounting for Sleep Time

The previous section showed the active mode energy and power benefits of using serial systems. However, the biggest benefit of serial systems occurs during sleep mode. Sensor applications usually spend substantial amounts of time in sleep mode. To model the impact of sleep mode energy, we let the addition system sleep for a given duration after each 32b add from 0 to 1 second. Note that, while 1 second sufficiently demonstrates the benefit of using a serial system, many ULP sensing systems sleep for minutes, hours, or even days between operating periods. During sleep mode, energy consumption results primarily from leakage. As mentioned in the introduction, sleep mode techniques like power gating can help reduce the sleep mode leakage to a small fraction of the active mode leakage. In our simulations, we assume that the system uses a leakage reduction mechanism that cuts leakage power to 10% of its active mode value.

The sum of active and sleep mode energy changes with  $V_{DD}$  and has an optimal point that minimizes this sum. As sleep time increases, this optimal  $V_{DD}$  point shifts to lower values. Table 1 shows the total energy consumed as sleep time varies, with each case calculated at its optimal  $V_{DD}$  value. A 1b system gives an energy benefit of 50% at zero sleep. For longer sleep times, the ratio of total energy approaches the ratio of leakage power, which means the serial case uses 32X less energy than the 32b system.

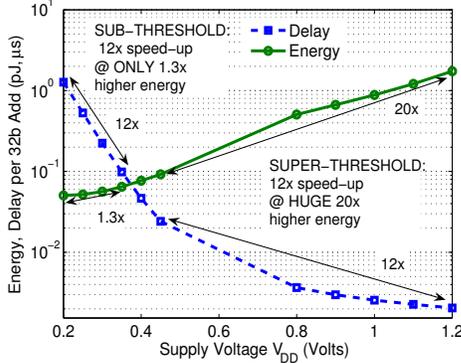
**Table 1: Total energy (pJ) accounting for sleep time after active operation.  $V_{DD}$  set to minimize energy for each entry.**

Topology	Zero Sleep	10μs	1ms	1s
1b SA-1	0.05	0.07	2.72	2723
16b KSA-16	0.10	0.48	38.10	38096
32b KSA-32	0.10	0.96	85.85	85852

### 3.2 Energy-Delay Analysis in Sub-threshold

The previous section showed that serial systems provide substantial energy and power reduction at a cost of increased delay, which results from multi-cycle operation to complete the same work. While more delay may not be a problem for some ULP systems, other systems may face a firm delay constraint. This section analyzes energy and delay in sub-threshold and finds the surprising result that sub-threshold serial systems are not only more energy efficient, but also as fast as parallel systems.

Figure 4 shows how active mode energy and delay of 1b SA-1 scales with  $V_{DD}$  from 0.2V to 1.2V. While the delay scales



**Figure 4: Total active energy (including dynamic energy), and delay per 32b add for 1b SA-1 as a function of  $V_{DD}$ .**

exponentially in sub-threshold, it scales roughly linearly in strong inversion. Conversely, energy has a steeper slope in strong inversion (quadratic), but a relatively shallow slope in sub-threshold near the minimum energy  $V_{DD}$ . This means that in sub-threshold the sensitivity of delay to  $V_{DD}$  far exceeds the sensitivity of energy to  $V_{DD}$ . In other words, we can increase  $V_{DD}$  only slightly to achieve a dramatic speedup with very little energy cost. Further, the shallow energy- $V_{DD}$  curve results in only a slight increase in energy consumed. Thus, as we confirm in Section 3.3, by increasing its  $V_{DD}$  slightly, we can potentially achieve energy savings with a 1b SA-1 over the 32b KSA-32 *at the same delay*.

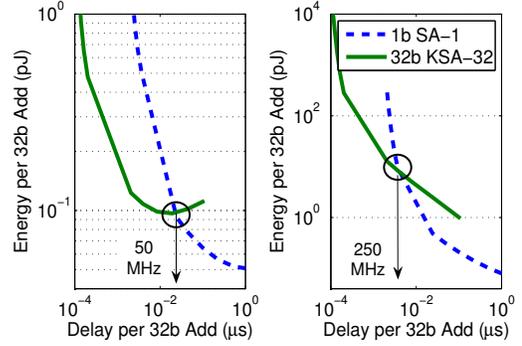
In contrast, in strong inversion, the opposite energy-delay sensitivities would require a huge increase in  $V_{DD}$  to make a serial system as fast as a parallel system, which would be very costly in energy. This explains why strong inversion designs use larger bit-widths. To probe this point further, Figure 5 shows the pareto-optimal energy-delay curves for 1b SA-1 and 32b KSA-32. For large delay values, a 1-bit system has lower energy at the same delay, and at small delay values, the situation is the opposite. The cross-over shows that below a certain  $V_{DD}$  (i.e. in sub-threshold, or above a certain delay requirement), a small bit-width gives less total energy *at the same delay*. When sleep time is included in (b), the cross-over moves to shorter delays, making our proposal of reconsidering the system bit width compelling for a wider range of applications.

### 3.3 Minimum Energy with a Delay Constraint

Figure 2 shows that smaller bit widths improve leakage power and active mode energy but increase delay by using multiple cycles to do the same work. The last section showed that in sub-threshold we can buy a speed-up at very little energy cost. In this section, we use this concept to show how serial systems retain their power and energy benefits at the same speed as parallel systems.

Increasing  $V_{DD}$  decreases delay and helps a simple topology (e.g. serial) be as fast as a topology built for high speed (e.g. Kogge Stone adder). This basic but powerful concept allows us to make a small bit width system as fast as a larger bit width system.

The interesting question is whether this  $V_{DD}$  increase tips the energy and power of a 1b system over that of the 32b system. The answer to this question differs in sub-threshold and super-threshold. Table 2 compares the energy consumed by different addition systems given a constant frequency constraint of 10MHz for the 32b add. This means that the 32b add must complete in 0.1 $\mu$ s, after which the sensor enters sleep mode. Sleep time



**Figure 5: Pareto-optimal E-D curves across sub-threshold and super-threshold (a) active mode energy (b) total energy with 10 $\mu$ s of sleep time.**

**Table 2: E / 32b add (pJ) including sleep E, for same delay.  $P_{lkg}$  ( $\mu$ W) (22nm CMOS [15],  $V_{DD}$  set to achieve 10MHz)**

Topology	Zero Sleep	10 $\mu$ s	1s	$P_{lkg}$	$V_{DD}$
1b SA-1	0.06	0.21	14919	0.15	0.35
16b KSA-16	0.10	0.80	70552	0.70	0.25
32b KSA-32	0.10	0.96	85852	0.86	0.20

changes from zero to 1 second. As sleep time increases, the contribution of sleep mode leakage energy increases.

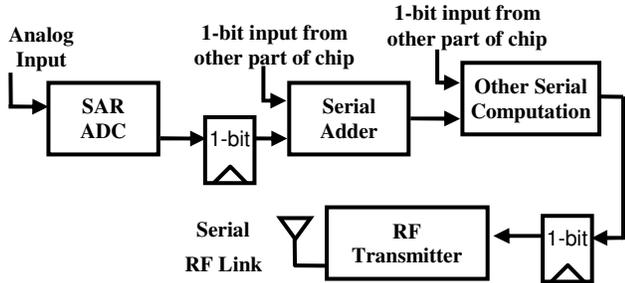
Counter to intuition, Table 2 shows that even at a raised  $V_{DD}$ , a 1b system consumes less active mode energy than a 32b system (the zero sleep time case). The benefit in leakage power does decrease from 32x to 5.7x, but these savings occur for the same operating speed. Thus, even for a fixed performance constraint, both active energy and leakage power are less for the small bit width system in sub-threshold. Our conclusions to this point are:

- At a fixed  $V_{DD}$ , a 1b serial adder has 40% lower active mode energy (including dynamic energy), 15x lower active mode power and 32x lower leakage power than a 32b version.
- By increasing  $V_{DD}$ , serial systems can be made as fast as parallel systems, while retaining the above energy and power benefits albeit with lower margins.
- The above optimal behavior of lower energy and power *at the same speed* can only be achieved in the sub-threshold regime, meaning that designers must reconsider the system level bit width while porting digital systems from strong inversion to sub-threshold.

The reason for this divergence lies in the basic transistor equations for sub- and super-threshold, which means that similar behavior with respect to bit width would be expected for any digital system. Further, we can generalize this observation to say that a small change in voltage in sub-threshold allows a “simple” topology to catch up in speed to a “high speed” topology. This implies that sub-threshold designs should use simple topologies.

## 4. SERIAL SYSTEM IMPLEMENTATION

Our analysis of adder systems shows convincing benefits in sub-threshold for using fully serial systems. This section envisions completely serial systems in Section 4.1 and examines limitations that might prevent complete serialization. Section 4.2 shows the



**Figure 6: A simple, fully serial digital system with a serial ADC (e.g. SAR topology), serial processing, and RF transmission.**

impact of different sources of overhead for mixed serial-parallel systems.

#### 4.1 Complete Serial Systems

This section presents a vision for a fully serial sensing node. Interestingly, existing research in ULP analog to digital converters (ADCs) favors successive approximation register (SAR) designs that inherently produce 1 bit at a time in steady state. A SAR ADC generates the MSB first, but most serial processing requires the LSB first. By buffering one word of data initially, the SAR can provide serial data with the LSB first as the next word resolves its MSBs. Thus, the SAR ADC acts as a serial source of sensed data to the rest of the system. Many protocols for offloading data from a chip use serial links, including I<sup>2</sup>C, UARTs, SPI, SLIMbus, etc. Likewise, most radios serialize data prior to sending it over the RF link.

Since serializing the input data and the output data for a system represents common practice, a fully serial system needs only to fill in the processing components with serial computation units. These may include a serial adder or various serial distributed arithmetic blocks (e.g. [13]). Figure 6 shows a block diagram of a representative serial system that leverages existing serial input and output streams. Data enters the 1b SA-1 as a stream of 1b operands. Every 16 cycles, for example, a 16b data word moves through the system, undergoes processing, and is communicated off-chip using the wireless link.

Clearly, this serial system will achieve the substantial area, active energy, and leakage power reductions that we observed in the previous section. Moreover, for sub-threshold operation, we can theoretically tune  $V_{DD}$  to provide the same speed as a parallel system but with area and leakage improvements that dramatically reduce standby power. One additional advantage of the serial system comes from its energy scalable properties. As fidelity requirements in the system vary, the required precision of data can change. In some cases, a smaller number of bits can adequately represent the incoming data. A SAR ADC can scale its energy with precision by stopping its conversion after fewer cycles. Parallel systems can use smaller precision data with energy savings, but only after including additional circuitry to mask off portions of the circuits (e.g. [8]). In the serial system, the same sort of graceful energy scaling can occur with much less overhead by simply changing the number of cycles used for each operation to match the precision of the incoming data.

While fully serial systems will in some cases provide the optimal ULP systems, many applications use operations that make full serialization difficult. For example, operations like multiplication require the storage of intermediate results as parallel numbers. Storing this information in parallel and then reconverting it to

serial (e.g. with a shift register) will increase active energy and leakage power. This overhead may prevent serialization in some cases. We can use a very simple analysis to evaluate the impact of such overhead in a general case. The following equation states the total energy required for a system to perform  $N$  operations followed by  $T_{SLP}$  amount of sleep time:

$$E_{TOTAL} = NE_{OP} + T_{SLP}P_{STANDBY} \quad (1)$$

To compare two systems, we can equate their total energy and solve for  $T_{SLP}$ :

$$T_{SLP} = N(E_{OP1} - E_{OP2}) / (P_{STANDBY2} - P_{STANDBY1}) \quad (2)$$

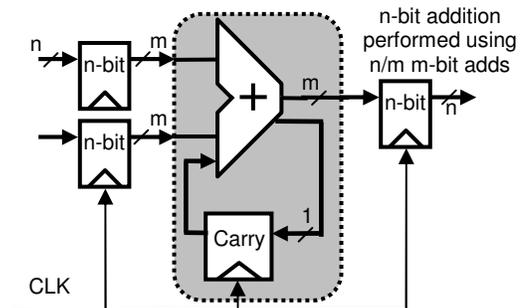
The *break even time* provided by (2) shows us how much sleep time is necessary to make the total energy consumed by the two systems the same. For a given amount of required work and a given active duty cycle, this equation tells us which system to use. If the break even time for a new system exceeds the sleep time that the duty cycle imposes, its overhead does not warrant use. If, on the other hand, a system change reduces leakage power enough to offset any active energy overhead in less than the sleep time, then that change saves overall power. Clearly, ULP systems with long sleep times will favor designs that reduce standby power, even at a cost of additional active energy per operation.

#### 4.2 Serial Components in Parallel Systems

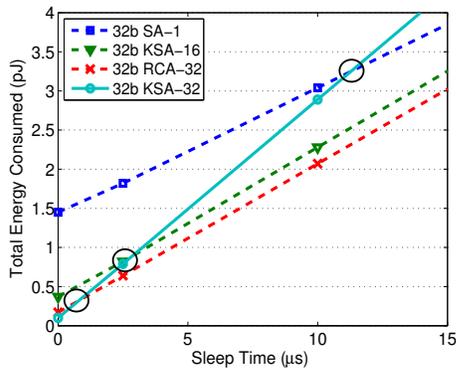
In some cases, application constraints will require a system to have parallel inputs and outputs, or an initial design may be fully parallel. If so, we can identify blocks to serialize even if the rest of the system cannot change. In other words, the block would still have a parallel interface, but would do its internal processing in a serial fashion. This allows us to leverage the area and leakage power benefits of a serial implementation without the need for system wide changes. To evaluate serialization of components, consider an  $n$ -bit addition system with a serial adder embedded in an  $n$ -bit system. The input and output register size equals the system bit width ( $n$ ), but the adder bit width ( $m$ ) can be smaller, as Figure 7 shows. For example, a 16b Kogge-Stone adder runs for 2 cycles to do a 32b add by saving the first carry. Note that the register size must be  $n$  because the  $n$ -bit addition system needs to be part of a larger  $n$ -bit synchronous digital system. In this section we compare addition systems with a common 32b parallel interface ( $n$ ), but varying adder block bit-width ( $m$ ):

- 32b addition using a 32b Kogge-Stone Adder: 32b KSA-32
- 32b addition using a 32b Ripple Carry Adder: 32b RCA-32
- 32b Serial Addition using a full adder and carry flop: 32b SA-1
- 32b Serial Addition with 16b KSA and carry flop: 32b KSA-16

The first part of the name refers to the system bit-width ( $n$ ) and



**Figure 7: A generic  $n$ ,  $m$ -bit addition system in a larger  $n$ -bit digital system.**



**Figure 8: Energy per op as a function of sleep time. 32b SA -1 starts off with lowest energy, but because of lower leakage current, 32b SA has lowest energy above a certain sleep time. (22nm CMOS [15],  $V_{DD}=300mV$ ).**

the second part refers to the adder block bit-width ( $m$ ). The number of cycles taken by a  $(n, m)$  addition system is  $n/m$ . We again account for energy in the adder block, registers, and clock for the entire 32b add and include sleep time after every add. Figure 8 shows how  $E/op$  varies with sleep time for the 4 adders. Notice that this figure essentially plots (1) for each adder system.

The energy at time zero equals the active mode energy, and as sleep time increases,  $E/op$  increases due to sleep mode leakage. The active energy of the serial adder exceeds the energy of the parallel adders by more than 14x, and this overhead occurs almost completely due to the 32b shift register, which is clocked 32 times per 32b add. However, the lower leakage power of the 32b SA-1 results in a break even time of 12  $\mu s$ , which is a fairly short sleep time for ULP applications. The break even time will be longer in older technologies with less leakage relative to the active energy. Similar crossover points exist for 32b RCA-32 and 32b KSA-16 at shorter sleep times, indicating that less serialization will be optimal for a range of intermediate duty cycles. The 32b RCA-32 and 32b KSA-32 break even point shows how merely changing the adder topology can produce significant energy savings. The RCA in this case uses a simpler topology that reduces leakage power. As we described in Section 3.2, tuning the sub-threshold voltage for the RCA can allow it to equal the speed of the more sophisticated KSA topology while still reducing power.

Performing the same analysis with a multiplier leads to larger break even times. A multiplier is less amenable to serialization because each output bit depends upon all previous input bits. However, we can still serialize the partial product addition. The resulting implementation will increase the active energy by roughly 32x (for fully serial additions), and (2) tells us that we must wait longer during sleep mode to compensate for this overhead. However, for ULP applications like structural monitors with very long sleep times, almost any active energy overhead becomes tolerable if it leads to a reduction in leakage power. Thus we see that even component level serialization conserves energy and power and extends the life-times of sub-threshold ULP applications with long sleep time.

## 5. CONCLUSION

Serial circuits reduce active energy and leakage power at the cost of delay, which suits them for ULP devices. For ULP applications

with large amounts of sleep time, the lower leakage power provided by serialization can dramatically extend device lifetime. Serial circuits achieve low leakage power by using the minimum possible circuits needed to get the job done. At the same voltage (300mV), a serial adder system achieves 40% active energy, 15x active power and 32x leakage power reduction over a 32b system. Moreover, the exponential delay dependence on  $V_{DD}$  in sub-threshold allows us to increase  $V_{DD}$  slightly to equate the delay of a serial system and a parallel system, and the serial system still saves active energy and leakage power. This indicates that serial systems are preferable whenever possible for sub-threshold operation, unlike in strong inversion. For a serial system, we recognize that SAR ADCs and RF communication are inherently serial techniques that can directly drive serial computation circuits. Even when some parts of the system must retain wider bit widths, we show that converting components to serial implementations provides leakage power reduction that compensates for active energy overhead for modest sleep times. In conclusion, serialization provides a strong knob for reducing energy and power in energy- or power-constrained ULP systems.

## 6. ACKNOWLEDGMENTS

This work was funded in part by NSF award number 0831426.

## 7. REFERENCES

- [1] A. Wang, B. H. Calhoun, and A. Chandrakasan, Sub-threshold Design for Ultra-Low-Power Systems, Springer, 2006.
- [2] B.H. Calhoun, A. Wang, and A. Chandrakasan, "Modeling and Sizing for Minimum Energy Operation in Sub-threshold Circuits," JSSC, Vol. 40, No. 9, pp. 1778-1786, September 2005.
- [3] B. Zhai, D. Blaauw, D. Sylvester, and K. Flautner, "Theoretical and practical limits of dynamic voltage scaling," DAC, pp.868-873, 2004.
- [4] B.H. Calhoun and A. Chandrakasan, "A 256kb Sub-threshold SRAM in 65nm CMOS," ISSCC, pp. 628-629, February 2006.
- [5] N. Verma and A. Chandrakasan, "A 256 kb 65 nm 8T Subthreshold SRAM Employing Sense-Amplifier Redundancy," JSSC, Vol. 43, No. 1, pp. 141-149, January 2008.
- [6] T. H. Kim, J. Liu, J. Keane, and C. H. Kim, "A 0.2 V, 480 kb Subthreshold SRAM With 1 k Cells Per Bitline for Ultra-Low-Voltage Computing," JSSC, Vol. 43, No. 2, pp. 518-529, Feb 2008.
- [7] J. P. Kulkarni, K. Kim, and K. Roy, "A 160 mV Robust Schmitt Trigger Based Subthreshold SRAM," JSSC, Vol. 42, No. 10, pp. 2303-2313, October 2007.
- [8] A. Wang and A. Chandrakasan, "A 180mV FFT Processor Using Subthreshold Circuit Techniques," ISSCC, pp. 292-293, Feb 2005.
- [9] Y. Pu, et al., "An Ultra-Low-Energy/Frame Multi-Standard JPEG Co-Processor in 65nm CMOS with Sub/Near-Threshold Power Supply," ISSCC, pp. 146-147, February 2009.
- [10] M. Seok, et al., "The Phoenix Processor: A 30pW platform for sensor applications," Symp VLSI Circ, pp.188-189, 2008.
- [11] J.Kwong, et al., "A 65 nm Sub-V<sub>t</sub> Microcontroller With Integrated SRAM and Switched Capacitor DC-DC Converter," JSSC, Vol. 44, No. 1, pp. 115-126, January 2009.
- [12] S. Jocke, et al., "A 2.6- $\mu W$  Sub-threshold Mixed-signal ECG SoC," Symposium on VLSI Circuits, 2009.
- [13] R. Amirtharajah, et al., "DSPs for Energy Harvesting Sensors: Applications and Architectures," IEEE Pervasive Computing Magazine, Vol. 4, No. 3, pp. 72-79, 2005.
- [14] L. Nazhandali, et al, "Energy optimization of subthreshold-voltage sensor network processors," ISCA, pp. 197-207, 2005.
- [15] W. Zhao and Y. Cao, "New generation of predictive technology modeling for sub-45nm early design exploration," IEEE Trans.on Electron Device, vol. 53, no. 11, pp. 2816-2823, Nov. 2006.