AN ENERGY EFFICIENT SUB-THRESHOLD BASEBAND PROCESSOR ARCHITECTURE FOR PULSED ULTRA-WIDEBAND COMMUNICATIONS

V. Sze, R. Blázquez, M. Bhardwaj, A. Chandrakasan
Massachusetts Institute of Technology
sze@mit.edu, rbf@mit.edu, manishb@mit.edu, anantha@mtl.mit.edu

ABSTRACT

This paper describes how parallelism in the digital baseband processor can reduce the energy required to receive ultra-wideband (UWB) packets. The supply voltage of the digital baseband is lowered so that the correlator operates near its minimum energy point resulting in a 68% energy reduction across the entire baseband. This optimum supply voltage occurs below the threshold voltage, placing the circuit in the sub-threshold region. The correlator and the rest of the baseband must be parallelized to maintain throughput at this reduced voltage. While sub-threshold operation is traditionally used for low energy, low frequency applications such as wrist-watches, this paper examines how sub-threshold operation can be applied to low energy, high performance applications. The correlators are further parallelized for a 31x reduction in the synchronization time, which along with duty-cycling, lowers the energy per packet by 43% for a 500 byte packet. Simulation results for a 100Mbps UWB baseband processor are described.

1. INTRODUCTION

The FCC has authorized UWB wireless communications in the 3.1GHz to 10.6GHz band with a minimum bandwidth of 500MHz and a maximum equivalent isotropic radiated power spectral density of -41.3dBm/MHz [1]. IEEE working group 802.15.3a is developing a high data rate standard for wireless personal area networks using UWB.

Applications of UWB include battery-operated devices such as mobile phones, handheld devices and sensor nodes. Consequently, there is a strong demand for an energy efficient UWB system. This paper will describe how operating the digital baseband in the sub-threshold region and increasing the degree of parallelism can translate into energy savings across the entire UWB receiver.

2. UWB SYSTEM ARCHITECTURE

The UWB packets are built from a sequence of binary phase-shift keying pulses with a 500MHz bandwidth. The transmitter generates approximate Gaussian pulses and upconverts the packet to one of 14 channels in the 3.1GHz to 10.6GHz band. Each packet, shown in Figure 1, is divided into two sections: preamble and payload. The preamble contains multiple repetitions of a N_c=31 bit Gold code sent at a pulse repetition frequency (PRF) of 25MHz, or T_{pre}=40ns. The payload contains the actual data and is sent at a PRF of 100MHz, or T_{pay}=10ns, for a 100Mbps data rate with no channel coding.

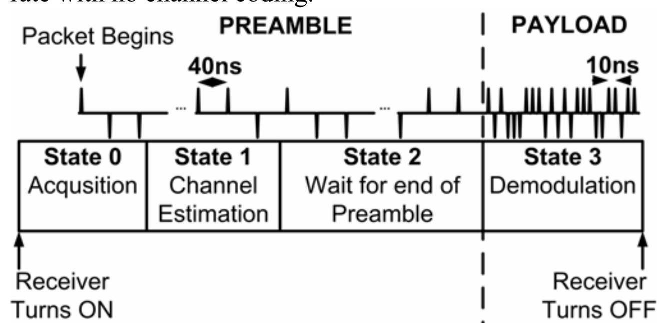

Fig. 1. UWB Packet Format

The receiver, shown in Figure 2, uses a direct conversion architecture in the front-end and the in-phase and quadrature components are sampled at 500MSPS by two 5-bit ADCs. For real-time demodulation of the UWB packet, the digital baseband must perform the signal processing with a throughput of 500MSPS. Synchronization is performed entirely in the digital domain. Only the automatic gain control (AGC) is fed back to the analog domain so that the digital baseband can scale to lower geometries. The baseband was simulated using the digital logic cell library of a 90-nm process.

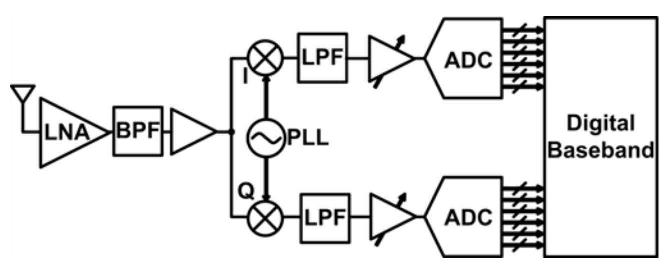

Fig. 2. UWB Receiver

3. DIGITAL BASEBAND PROCESSOR

The digital baseband performs packet detection, acquisition, delay correction and channel estimation using the preamble, followed by demodulation of the payload. Additional repetitions in the preamble are required for AGC, but will not be included in the discussion. Figure 1 outlines the baseband processor's four states of operation with respect to the packet.

In State 0, the acquisition phase, the baseband detects the presence of a packet and provides an initial estimate of its delay. This is accomplished by performing a correlation of the input with an unknown delay against a 31-bit Gold code. Each correlation takes place over $T_{code}=N_c\times T_{pre}=$ 1240ns. The delay must be resolved up to 2ns accuracy; therefore, there are a total of 620 possible delays and corresponding correlations: 20 to match the pulse position, and 31 to match the Gold code. Until acquisition is achieved, the baseband remains in State 0 and performs these correlations. When a correlation exceeds a predefined threshold, acquisition is declared (i.e. lock is detected) and the baseband retimes the input so that it is aligned before moving on to State 1. If all 620 delays are checked and the baseband does not detect lock, the UWB receiver turns off.

In State 1, the channel estimation phase, the baseband must acquire channel estimates from the output of the correlators. This must be done before demodulation in order to compensate for the detrimental effects in the UWB channel [2]. The channel estimates are used to construct a five tap FIR matched filter that takes both the pulse shape and channel impulse response into account.

In State 2, the detection of payload phase, the baseband waits for the end of the preamble which is indicated by an inverted replication of the Gold code. During State 1 and 2, the baseband continuously performs correlations to check that the baseband remains locked. If a threshold is not met, the packet is assumed to be lost or to have been a false packet lock, the baseband and the rest of the UWB receiver turns off. In addition, the baseband performs delay correction with the use of a delay locked loop which is part of the retiming block.

Finally, in State 3, the demodulation phase, each pulse of the payload is filtered by the matched filter derived from the channel estimates and then passed through a decoder that resolves the bit.

A block diagram of the baseband is shown in Figure 3. This paper exploits two forms of parallelism. N defines the degree of parallelism required to operate the digital baseband in sub-threshold. M is defined as the number of Gold Code correlations performed simultaneously. Each sub-bank, composed of N correlators, checks for one Gold Code delay. The trade-offs involved in the specification of M and N will be discussed in the following sections. Other papers have discussed the use of parallelism to reduce *power* consumption for a baseband that uses both autocorrelation and cross-correlation [3]; however, the metric here is to reduce the *energy* consumption for a baseband that uses only cross-correlation.

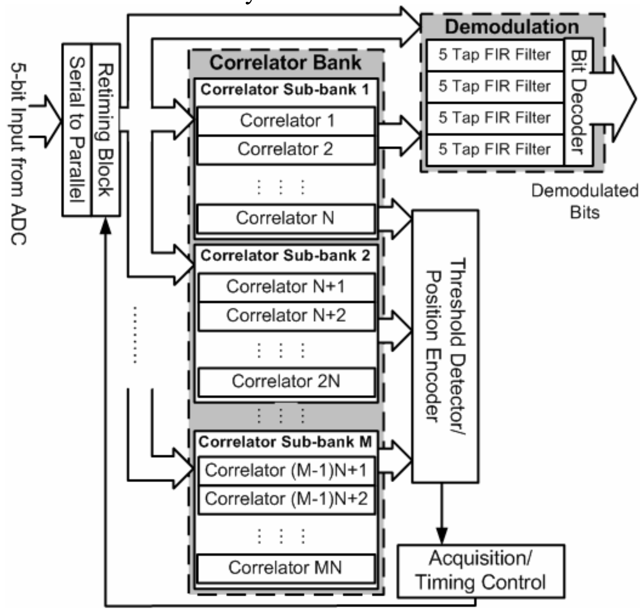

Fig. 3. UWB Parallelized Digital Baseband

4. SUB-THRESHOLD OPERATION (IMPACT OF N)

As previously mentioned, since the input from the ADC arrives at a rate of 500MSPS, a serial baseband must run at a frequency of 500MHz if the input is to be processed in real time. In order that the critical paths, through the correlator and through the matched filter, meet the timing constraint, the digital circuitry must run at its maximum supply voltage. However, running at the maximum voltage is not energy-efficient. It is important to reduce the energy of the correlator since it consumes the largest portion of energy in the baseband during synchronization. The energy per operation can be reduced by lowering the supply voltage (V_{dd}) [4]. At maximum V_{dd}, the transistors in the circuit operate in the active region. If V_{dd} is lowered below the threshold voltage (V_{th}) of the device, the circuit is said to be operating in the sub-threshold region. Lowering V_{dd} increases the latency per operation (T_{period}) linearly in the active region, and exponentially in the sub-threshold region. This increases the leakage energy as it is linearly related to T_{period}. There is a minimum operating energy point since the dynamic energy and the leakage energy scale in an opposite manner with V_{dd} [5]. Spectre simulations of the correlator in the 90-nm process show that operating at the minimum energy point of 0.3V rather than at the maximum V_{dd} of 1V reduces the energy per operation of the correlator by 89% (Figure 4).

At the minimum energy point, the baseband processing must be parallelized to maintain a throughput of 500MSPS. For ease of design, it is desirable that the PRF of the preamble be a multiple of the clock frequency of the baseband. Since this is not possible at 0.3V, the baseband operates slightly above the minimum energy point at 0.4V with a frequency of 25MHz which requires N=20 correlators to form a sub-bank of correlators. Lowering the supply voltage from 1V to 0.4V (sub-threshold) results in an overall energy savings of 83% for the correlators and 68% for the entire baseband. The energy savings for the entire baseband is less since buffers are inserted in some paths to compensate for the increased transition time at 0.4V.

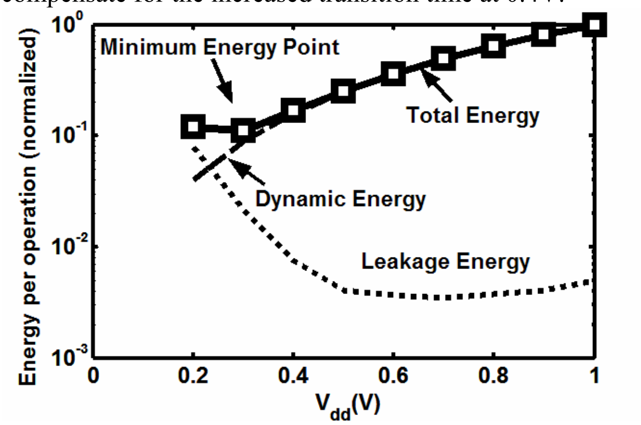

Fig. 4. Simulated energy plot for the correlator

5. REDUCED ACQUISITION TIME (IMPACT OF M)

Increasing M increases the number of code shifts that are simultaneously checked. Although this increases the power consumed by the baseband during acquisition, the time spent in acquisition decreases proportionally. In this section a model is developed to show that the baseband energy remains approximately the same for any M, while the energy spent by the rest of the receiver scales inversely with M. This results in an overall reduction in energy per packet.

5.1. Modeling Energy per Packet.

The average time and amount of energy the baseband spends in each state must be determined. The total time the baseband spends in State 0 and 2 is set by the number of times the Gold code is repeated in the preamble, R(M), which can be reduced by increasing parallelism M.

$$R(M) = \left\lceil \frac{N_c}{M} \right\rceil \tag{1}$$

While the time spent in State 1 and State 3 is fixed, the distribution of time between State 0 and 2 is dictated by when the baseband detects lock. The maximum time the baseband will remain in State 0 is R(M)×T_{code}. In this case, no time is spent in State 2.

Let D be the number of code shifts between the Gold code in the preamble and the Gold code in the baseband. Assume that D is uniformly distributed over [0, N_c-1]. Let I(D,M) be the number of code durations (T_{code}) required to achieve acquisition in State 0. Let P_d be the probability that the baseband detects lock when the input and the Gold code are aligned, and P_{fa-M} be the probability that the baseband detects lock in one or more of M delays which are not aligned to the code. For small P_{fa} (=P_{fa-1}), $P_{fa-M} \approx M \times P_{fa}$.

Assuming the detector is ideal (P_d=1, P_{fa-M}=0),

$$I(D,M) = \left\lceil \frac{D}{M} \right\rceil \tag{2}$$

The maximum number of code durations, T_{code}, required to achieve acquisition in State 0 is R(M). As the baseband performs different operations during each state, the energy per T_{code} varies per state. E_0 and E_2 are the energies consumed over T_{code}, in State 0 and State 2, while E_1 and E_3 are the energies required to perform channel estimation and demodulation respectively. It is important to note that E_0, to a first order, scales linearly with M. After acquisition, M-1 of the correlator sub-banks can be turned off through the use of clock gating and power gating so that E_1, E_2 and E_3 are not dependent on M to a first order.

The energy per packet is computed as follows,

$$\begin{aligned} Energy(D,M) &= \sum_{X=1}^{R(M)} \Pr(X,D,M) \times Energy(X,D,M) \\ &= \alpha E_0 + \beta E_1 + \gamma E_2 + \varepsilon E_3 \end{aligned} \tag{3}$$

Pr(X,D,M) is the probability that the baseband will stay in acquisition for X units of T_{code} given a packet with delay D. Energy(X,D,M) is the energy consumed by the baseband.

For all X≠I(D,M),

$$\Pr(X,D,M) = P_{fa-M}\left(1 - P_{fa-M}\right)^{X-1} \tag{4}$$

$$Energy(X,D,M) = \left(XE_0 + E_1\right) \tag{5}$$

When X=I(D,M),

$$\begin{aligned} &\Pr(X,D,M) \times Energy(X,D,M) \\ &= P_d\left(1 - P_{fa-M}\right)^{X-1}\left\{XE_0 + E_1 + \left(R(M) - X\right)E_2 + E_3\right\} \\ &\quad + (1 - P_d)\left(1 - P_{fa-M}\right)^{X-1} R(M) E_0 \end{aligned} \tag{6}$$

This paper assumes P_d=0.9 and $P_{fa}=10^{-5}$, which were derived from the 802.15.3a proposal. The average energy required by the baseband to process a packet for a given degree of parallelism M is computed by taking the expected value of the energy per packet over all possible delays D, conditioned on M. If P_{fa} is small, this average baseband energy does not change significantly since, to a first order, the same number of operations occur for any M. In addition, for a small P_{fa}, the required preamble time and hence energy spent during acquisition by the rest of the receiver scale inversely with M.

5.2. Impact of M on Energy per Packet

The energy per packet can be broken down into the preamble energy and the payload energy. The payload energy is fixed by the number of bits transmitted per packet. However, the length of the preamble, and consequently the preamble energy, can be reduced based on the configuration of the baseband. A previous version of the UWB baseband checked one combination of the Gold code at a time [6]. In order to check all shifted combinations of the 31-bit Gold code, the baseband must perform at least 31 correlations. The preamble must last for $N_c \times T_{pre} \times R(M=1) = 34.880\mu s$.

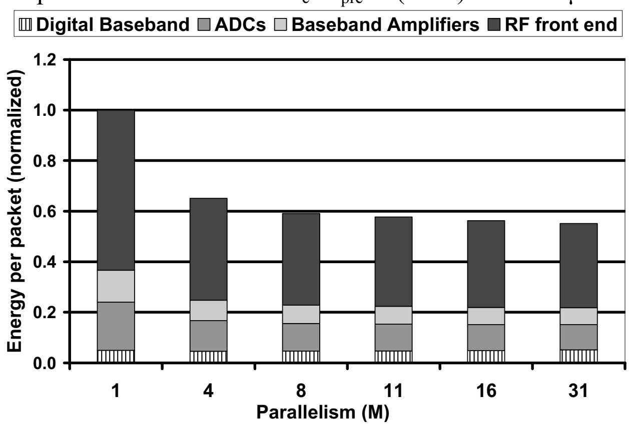

Fig. 5. Average packet energy consumption of the receiver subsystems for various degrees of parallelism.

By using multiple sub-banks of correlators that operate in parallel, the number of Gold code shifts that can be checked in one cycle is increased, which reduces the number of repetitions required in the preamble. In a fully parallelized baseband, with 31 sub-banks of correlators, all 31 shifted possibilities of the Gold code are checked simultaneously, and the Gold code only has to be repeated once in the preamble for acquisition. This results in a 31x reduction in the preamble length. As previously stated, for varying degrees of M, the energy spent by the baseband on the acquisition is almost the same with a slight increase due to increased interconnect capacitance that results from parallelism. The actual energy savings result from the other circuitry in the UWB receiver. Reduction in acquisition time implies that the entire receiver needs to be on for a much shorter period of time. The RF front end, ADCs and the baseband amplifiers can be turned off once the packet has been demodulated. The measured power of these blocks is approximately 79% of the receiver power [7], [8]; shutting them off earlier translates into significant energy savings. Figure 5 shows the reduction in energy per packet, with payload size of 500 bytes, for various degrees of parallelism. It can be concluded that faster synchronization, combined with duty-cycling, reduces the energy required to receive a UWB packet. As increasing M only affects preamble energy, the impact of using parallelism to reduce energy per packet varies with payload size (Figure 6).

It is important to note that reduction in preamble energy should not be made at the expense of the payload energy. Techniques such as clock gating and power gating are used to ensure that the power consumption of the baseband during demodulation does not increase with parallelization. During State 1, 2 and 3, either most or all correlators are turned off and power gated to reduce leakage, and clock gating should be used to reduce the impact of the increased interconnect capacitance due to parallelism.

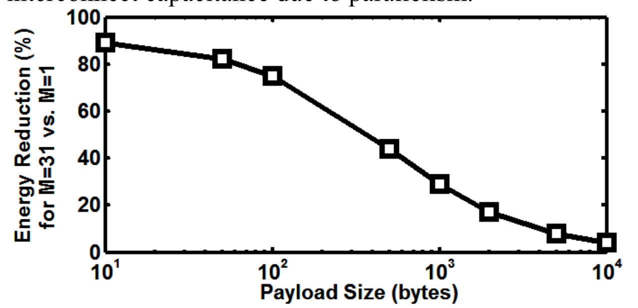

Fig 6. Energy reduction for various payload sizes

6. CONCLUSION

This paper discusses how parallelism allows for voltage scaling and reduced acquisition time, which reduces the energy required to receive a UWB packet. Voltage scaling to sub-threshold allows the correlator sub-banks to operate near the minimum energy point, resulting in an energy per operation reduction in the correlators of 83% and energy reduction of 68% across the entire baseband. The reduced acquisition time through further parallelization of the correlator sub-banks by 31 led to a 43% reduction in energy per packet for a 500 byte packet. The analysis in this paper can be mapped to other high performance communication applications using sub-threshold operation and parallelism.

ACKNOWLEDGMENTS

This research is sponsored by DARPA and an NSERC Fellowship. We thank STMicroelectronics for providing process and cell library access.

REFERENCES

[1] FCC, "Ultra-wideband First Report and Order", February 2002.
[2] J. Foerster, "Channel Modeling Sub-Committee Report Final," IEEE P802.15 Working Group for Wireless Wireless Personal Area Networks (WPANs). February 2002.
[3] M. Verhelst, W. Dehaene, "System Design of an Ultra-low Power, Low Data Rate, Pulse UWB Receiver in the 0-960MHz Band," *IEEE International Conference on Communications*, 2005.
[4] A. Wang, A. Chandrakasan, "A 180mV FFT Processor Using Subthreshold Circuit Techniques," *ISSCC*, February 2004.
[5] B. H. Calhoun, A. P. Chandrakasan, "Characterizing and Modeling Minimum Energy Operation for Subthreshold Circuits," *ISLPED*, August 2004.
[6] R. Blázquez, A. Chandrakasan, "Architectures For Energy-Aware Impulse UWB Communications," *ICASSP*, 2005.
[7] B. P. Ginsburg, A. Chandrakasan, "Dual Scalable 500MS/s, 5b Time-Interleaved SAR ADCs for UWB Applications,"*CICC*, 2005.
[8] F. S. Lee , A. P. Chandrakasan, "A BiCMOS Ultra-Wideband 3.1-10.6GHz Front-End," *CICC*, 2005.