

# Design and Implementation of a Sub-threshold BFSK Transmitter

Suganth Paul<sup>#</sup>  
suganth.paul@intel.com

Rajesh Garg<sup>‡</sup>  
rajeshgarg@tamu.edu

Sunil P Khatri<sup>‡</sup>  
sunilkhatri@tamu.edu

Sheila Vaidya\*  
vaidya1@llnl.gov

<sup>#</sup> Intel Corporation, Austin, TX 78746

<sup>‡</sup> Department of ECE, Texas A&M University, College Station TX 77843

\* Lawrence Livermore National Laboratories, Livermore, CA 94550

**Abstract**—Power Consumption in VLSI circuits is currently a major issue in the semiconductor industry. Power is a first order design constraint in many applications. However, a growing class of applications need extreme low power but do not need high speed. Sub-threshold circuit design can be used for these applications. Unfortunately, sub-threshold circuits exhibit an exponential sensitivity to process, voltage and temperature (PVT) variations. In this paper we implement and test a robust sub-threshold design flow which uses circuit level PVT compensation to stabilize circuit performance. We design and fabricate a sub-threshold BFSK transmitter chip. The transmitter is specified to transmit baseband signals up to a data rate of 32kbps. Experiments using the fabricated die, verify the functionality of the design show that the sub-threshold circuit consumes  $19.4\times$  lower power than the traditional standard cell based implementation on the same die.

**Keywords:** Sub-threshold Design, Low Power, BFSK, Adaptive Body Biasing

## I. INTRODUCTION

Power consumption is a dominant issue in contemporary circuit design. Due to their extreme low power consumption, sub-threshold design approaches are appealing for a widening class of applications which demand low power consumption and can tolerate larger circuit delays. Examples include sensor networks, wearable computers, certain portable electronic devices, etc. Here speed is a secondary design goal, whereas low power consumption is a primary design requirement. Sub-threshold circuit design is done by setting  $V_{DD} \leq V_T$  in the circuit.

Results from [1] indicate that a sub-threshold design approach will yield a  $100\text{-}500\times$  reduction in power, compared to traditional designs, with a  $10\text{-}25\times$  delay penalty. This analysis was performed using the Berkeley Predictive Technology Model [2] for the  $0.1\mu\text{m}$  and  $0.07\mu\text{m}$  processes. Table I shows the delay, power and power-delay product (P-D-P) for a 21-stage ring oscillator, implemented using both the traditional and sub-threshold design approaches. Note that Table I also indicates that the P-D-P of the sub-threshold design is  $10\text{-}25\times$  better than a traditional design.

However, sub-threshold circuits are exponentially sensitive to variations in supply, temperature and processing (PVT) factors. This is evident from the sub-threshold leakage current

of a MOSFET given by the equation:

$$I_{ds} = \frac{W}{L} I_{D0} e^{\left(\frac{V_{gs} - V_T - V_{off}}{n v_t}\right)} \left[1 - e^{-\frac{V_{ds}}{v_t}}\right] \quad (1)$$

In Equation 1,  $V_T$  is the device *threshold voltage*. It depends on process dependent factors like gate and insulator materials, thickness of insulator and channel doping density. It also depends on operational factors like  $V_{sb}$  (body effect) and temperature ( $V_T$  is inversely proportional to device junction temperature).  $W$  and  $L$  are the device width and length. Also,  $I_{D0}$  is a constant while  $v_t = \frac{kT}{q}$ . Here  $k$  is the Boltzmann's constant, and  $v_t = 26\text{mV}$  at room temperature.  $n$  is the sub-threshold swing parameter (a constant). Finally,  $V_{off}$  is a constant.

In order to center a sub-threshold design we need to stabilize the circuit delay by compensating for PVT variations. We have proposed a technique in [1] that uses self-adjusting body bias, to *phase lock* the circuit delay to a *beat clock*. The circuit is implemented using a network of interconnected, medium-sized Programmable Logic Arrays (PLAs). This phase locking is done for a group of spatially localized PLAs. Spatially localized PLAs are clustered, and each cluster of PLAs shares a common *Nbulk* node. This *Nbulk* node is driven by a bulk bias adjustment circuit (one per PLA cluster), whose task is to synchronize the delay of a representative PLA in the cluster, to a globally distributed *beat clock*. Since all PLAs in any cluster are identical, this approach compensates circuit delay. In this design we use a single cluster of 33 PLAs.

In this project we implement a BFSK (Binary Frequency Shift Keying) transmitter using sub-threshold circuits. The transmitter is capable of modulating message signals up to a data rate of 32kHz. We fabricate the design on a  $10\text{mm}^2$  die in a TSMC  $0.25\mu\text{m}$  triple well CMOS process. Test results indicate that the sub-threshold circuit is able to operate using  $19.4\times$  lower power than a traditional standard cell based implementation of the same function on the same die.

The **key contributions** of the paper are

- 1) A dynamic delay compensation scheme for sub-threshold circuits to combat PVT variations is implemented and tested in silicon,
- 2) An EDA flow is developed and validated for network of PLA (NPLA) based circuit design.

TABLE I  
COMPARISON OF TRADITIONAL AND SUB-THRESHOLD CIRCUITS

Process	Traditional Ckt			Sub-threshold Ckt ( $V_b = 0V$ )			Sub-threshold Ckt ( $V_b = VDD$ )		
	Dly (ps)	Pwr (W)	P-D-P (J)	Delay $\uparrow$	Power $\downarrow$	P-D-P $\downarrow$	Delay $\uparrow$	Power $\downarrow$	P-D-P $\downarrow$
bsim70	14.157	4.08e-05	5.82e-07	17.01 $\times$	308.82 $\times$	18.50 $\times$	9.93 $\times$	141.10 $\times$	14.43 $\times$
bsim100	17.118	6.39e-05	1.08e-06	24.60 $\times$	497.54 $\times$	20.08 $\times$	12.00 $\times$	100.96 $\times$	8.20 $\times$

- 3) We show a 19.4 $\times$  lower power as compared to traditional circuits (both implemented using a 0.25 $\mu$ m process. From simulations we get around 100 $\times$ -500 $\times$  lower power while using a 0.1 $\mu$ m and 0.07 $\mu$ m process.

Section II describes our approach to design and implement the BFSK transmitter using sub-threshold circuits. In Section III we present test data from the fabricated die. We conclude in Section IV.

## II. OUR APPROACH

In this section we describe the functional operation of the BFSK transmitter and provide details about its design and implementation.

### A. Functional Description

A typical BFSK transmitter generates one of two frequency tones at the output, and shifts the frequency of the output tone to any of two pre-determined values depending on the value of the binary input message signal which can be a logical HIGH or LOW. The input to the transmitter is assumed to be digitized and supplied to the transmitter at a rate of  $R_B$  bits/s. We design the system to be able to modulate signals up to 32k bits/s, which is good enough for voice transmission. The frequencies of the two tones that will be produced by the BFSK transmitter are given by  $f_1$  and  $f_2$ .  $\phi_1$  and  $\phi_2$  are phase offsets that the two tones could have. Depending on the value of the binary input, one of the tones is multiplexed to the output. We use non-coherent BFSK modulation, where  $\phi_1 \neq \phi_2$ . The two frequency tones are produced using digital circuits implemented in the form of a Numerically Controlled Oscillator (NCO).

The BFSK transmitter architecture shown in Figure 1 consists of a digital BFSK modulation circuit, along with a Digital to Analog Converter (DAC) and amplifier. The BFSK modulation circuit is made up of the Phase Accumulator, NCO and a Binary to Thermometer Code Converter. The analog components in the design include a DAC, an amplifier and an antenna for wireless transmission. This is shown in Figure 1. The BFSK modulator is implemented as a digital circuit, using a network of Programmable Logic Arrays (NPLAs). We next give a brief introduction to PLAs and how they are used in a network to do computations. We will also discuss in detail each of the digital and analog components that make up the design of the system.

### B. Description of Design: NPLA Based Dynamic Compensation

In this section we describe the NPLA style of circuit design. We explain the dynamic compensation scheme and state the

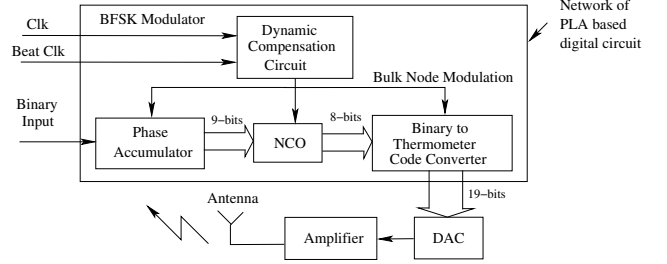


Fig. 1. System Architecture

advantages of using a NPLA based style over a standard cell based design.

1) *PLA Operation*: PLAs are the basic circuit modules used in this design. The PLAs operate in the sub-threshold region of conduction. Consider a PLA consisting of  $n$  input variables  $x_1, x_2, \dots, x_n$ , and  $m$  output variables  $y_1, y_2, \dots, y_m$ . Let  $k$  be the number of rows in the PLA. A *literal*  $l_i$  is defined as an input variable or its complement.

Suppose we want to implement a function  $f$  represented as a sum of cubes  $f = c_1 + c_2 + \dots + c_k$ , where each cube  $c_i = l_i^1 \cdot l_i^2 \cdot \dots \cdot l_i^{r_i}$ . We consider PLAs which are of the *NOR-NOR* form. This means that we actually implement  $f$  as

$$\bar{f} = \sum_{i=1}^k (c_i) = \sum_{i=1}^k (\bar{c}_i) = \sum_{i=1}^k (\bar{l}_i^1 + \bar{l}_i^2 + \dots + \bar{l}_i^{r_i}) \quad (2)$$

The PLA output  $\bar{f}$  is a logical NOR of a series of expressions, each corresponding to the NOR of the complement of the literals present in the cubes of  $f$ . Figure 2 illustrates the schematic of the PLAs used in this design.

The PLAs are dynamic logic blocks. They enter their precharge state when the  $CLK$  signal is low. During this time, the outputs are precharged. A special output line (which is inverted to produce the signal *completion* shown in Figure 2) also gets precharged. The *completion* signal is also the last output signal to switch, since it is maximally loaded, in comparison to other outputs. The *completion* signal switching low signals the completion of the precharge operation of the PLA. In the precharged state, all the wordlines and the output lines of the PLA are precharged. Now, when the  $CLK$  signal switches high, the PLA enters the *evaluation* phase. In evaluation, if any of the vertical bitlines are high, the wordline that it is connected to, gets pulled low. One of the inputs *and* its complement is connected to the dummy wordline, so that the dummy wordline switches low during *every* evaluate phase and effectively acts as a timing reference for the PLA. By

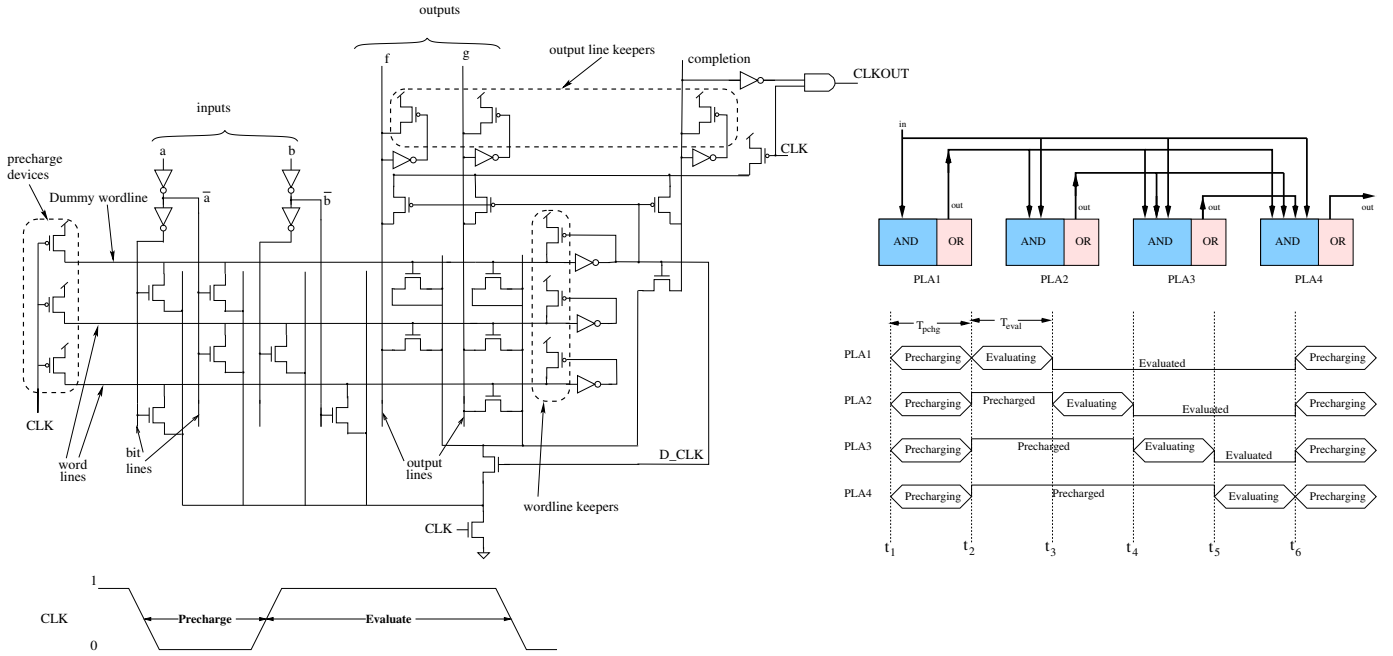


Fig. 2. Schematic View of PLA and Timing of NPLA

design, the dummy wordline is the last wordline to switch low. When the dummy wordline switches low, it makes the signal D\_CLK switch high, as a result of which the GND gating transistor driven by D\_CLK turns on. The output lines to which wordlines that have switched low are connected, will switch low. The *completion* line, which is connected to the complement of the dummy wordline is the last signal to switch high. This signals the completion of the evaluation operation. The completion signal of the PLA switches in each cycle.

2) *Network of PLA Operation*: A network of PLAs, NPLA is nothing but a multilevel network of PLAs. Each of the digital components that make up the digital BFSK modulator in Figure 1, i.e. the NCO and the Binary to Thermometer Code Converter are made of NPLAs. Each of these blocks are implemented as combinational circuits and the outputs of each of these blocks are registered using negative edge triggered flip-flops clocked by CLK. The flip-flops are negative edge triggered as the outputs of the flip-flops need to be stable when the CLK signal is HIGH when the PLAs are evaluating. The timing diagram of NPLAs in a single combinational circuit is shown in Figure 2. Notice from this figure that all the PLAs in a network precharge at the same time and start evaluating one after another in a cascading fashion. Hence an evaluation period has to be provided that is sufficient for *all* the PLAs to evaluate. A PLA at logical depth  $i$  in the network is clocked by the logical AND of all the CLKOUT signals of PLAs at logical depth  $i - 1$ , except for the first PLA in the chain which is clocked by the CLK signal. The CLKOUT signal of each PLA is the logical AND of its *completion* signal and the CLK signal. The maximum throughput that can be achieved depends on the delay of the slowest combinational block (i.e.

the maximal logic depth of all combinational blocks). When implemented as a network of PLAs, the throughput of the circuit can be approximately written as:

$$Throughput = \frac{1}{T_{pchg} + N * T_{eval}} \quad (3)$$

Here  $N$  is the number of levels of PLAs needed in the multilevel network of PLAs.

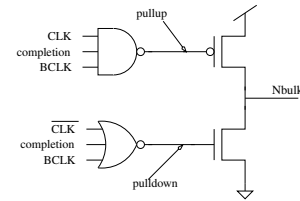


Fig. 3. Phase Detector and Charge Pump Circuit

3) *Dynamic Compensation Circuit*: The dynamic delay compensation circuit is shown in Figure 3. The task of this circuit is to synchronize the delay of a representative PLA in the cluster, to a globally distributed *beat clock*. The beat clock is an external signal, derived from the system clock. For a high speed of operation, the duty cycle of *beat clock* needs to be increased, and all PLAs in the design speed up to synchronize to *beat clock*. Conversely, reducing the duty cycle of *beat clock* slows down the PLAs to synchronize to *beat clock* again. In this way, we can implement a synchronous design using sub-threshold PLAs, in a manner that is insensitive to inter and intra-die processing, temperature and voltage variations.

This phase locking is done for a group of spatially localized Programmable Logic Arrays (PLAs). These PLAs are placed

such that they are part of a single cluster of PLAs sharing a common  $N_{bulk}$  node. This  $N_{bulk}$  node is driven by the charge pump shown in Figure 3. The self-adjusting body bias scheme controls the substrate voltage of the PLAs in a closed-loop fashion, by ensuring that the delay of a representative PLA in the cluster is phase locked to the *beat clock* signal. The phase detector and charge pump circuits used for the design are shown in Figure 3. Note that since PLAs have NMOS devices in their AND and OR planes, we only compensate the bulk node of the NMOS devices.

The NAND gate in this figure detects the case when the completion signal is too slow, and generates low-going pulses in such a condition. These pulses are used to turn on the PMOS device of Figure 3, and increase the  $N_{bulk}$  bias voltage, resulting in a speed-up in the PLAs. If the *completion* signal of the reference PLA has not occurred by the time  $BCLK$  rises, a downward pulse is generated on the *pullup* signal, which forces charge into the  $N_{bulk}$  node, resulting in faster generation of *completion*. At this time, *pulldown*, the signal which is used to bleed off charge from  $N_{bulk}$ , is low. The NOR gate in Figure 3 generates high-going pulses to turn on the NMOS transistor when the PLA delay leads *beat clock*. These pulses drive the NMOS device in Figure 3, bleeding charge out of  $N_{bulk}$  and thereby slowing the PLA down.

Figure 4 shows the effectiveness of the dynamic body biasing scheme. Here the envelope of delays under process variations is studied for a standard cell based 9 stage ring oscillator and a delay compensated PLA circuit. The nominal delays for both circuits are matched at room temperature. The temperature is varied from 0-100 degree C and at each temperature the parameters  $V_T$ ,  $V_{DD}$  and  $L_{min}$  are varied with  $3\sigma=15\%$ . The maximum delay encountered is shown in Figure 4 for both circuits. We see that the PLA delay is almost constant while the uncompensated delay varies by  $7\times$ . These simulations are performed for the TSMC  $0.25\mu$  process used to fabricate our chip.

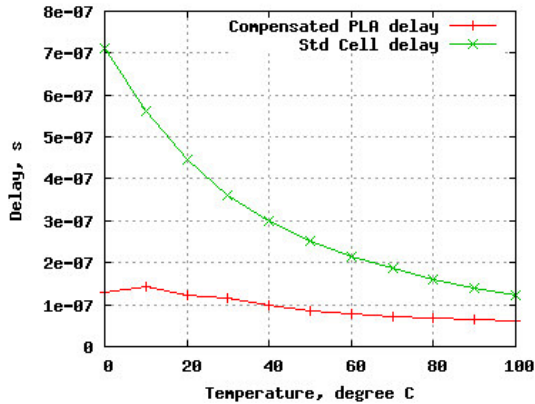


Fig. 4. Delay Compensation with  $V_T$ ,  $V_{DD}$  and  $L_{min}$  and Temperature Variations

4) *Advantage of NPLAs*: In order to dynamically compensate the delay of the design, we need to find the critical delay path in the circuit and use it as the representative signal to phase lock with the *beat clock* signal. In the case of standard

cell based circuit design this is a non-trivial task. PLAs are used here mainly because a PLA can be designed such that the delay of *all* outputs is constant, regardless of the input vector applied. Hence, the task of finding the critical delay path (which needs to be found in standard cell based bulk bias control approaches such as [3]) is avoided. If all the PLAs in the design have the same size (as in our case), the circuit delay to be monitored can be the delay of *any* of the PLAs (used as a representative block).

Also, design methodologies using a network of medium sized PLAs was shown [4] to be a viable way to perform digital design, resulting in improved delay for a design. In a standard cell based flow, there is an intervening technology mapping step, which often negates the benefits of technology-independent logic optimization. A network of PLAs on the other hand, allows us to carry forward the benefits of technology-independent multi-level logic synthesis.

### C. EDA Tools and Methodology

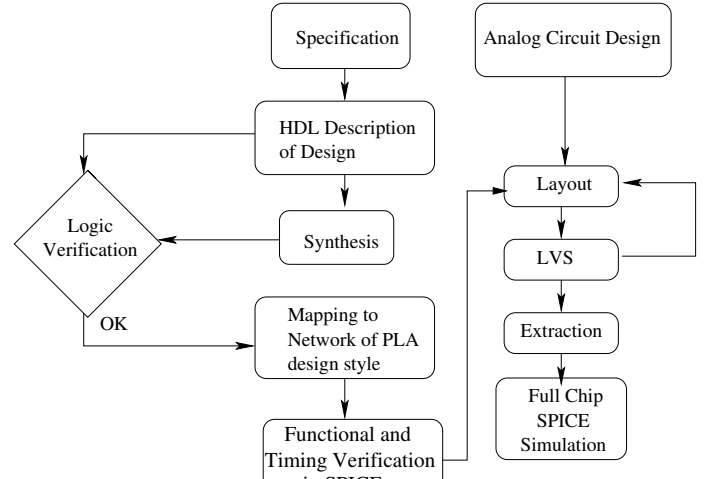


Fig. 5. Design Flow

The steps of the design flow to be used are shown in Figure 5, and briefly described here.

- 1) First the design specification (obtained by user requirements such as frequency of data being transmitted, available bandwidth, distance of transmission etc) are determined.
- 2) Next, the HDL code to implement the specification is developed. We used VHDL for this step.
- 3) This code was synthesized next, resulting in an RTL description of the design. We used the Xilinx XST synthesis tool for this. This step was done so as to obtain a technology independent logic netlist from the HDL code. This technology independent netlist was next translated to the *blif* format and the remaining synthesis steps were done in SIS [5].
- 4) The synthesized gate level netlist was verified against the HDL, by running functional test vectors. This was done using the Modelsim tool.

- 5) Next the design is decomposed to a network of PLA based netlist. We first translated the Xilinx XST synthesis netlist to *blif*. Then, we decompose the netlist into a NPLA. We used the synthesis code from [6] for this purpose. The synthesis tool used was SIS. The size of each of the PLAs (the number of inputs, outputs, and cubes) to be used in the design is determined at this point based on the number of PLAs required for the design (area) and the speed of operation of the PLAs (latency and throughput). At the end of this step, the NPLA netlist is translated into a SPICE netlist.
- 6) A functional and timing verification is done on the SPICE level schematic. This simulation is done across all process corners. This validated and tested the design of the circuit to some extent.
- 7) The analog circuits are designed separately. The analog circuits were simulated and laid out. We used the Cadence Schematic environment for simulation and the Virtuoso environment for layout of the analog blocks.
- 8) Using the netlist of PLAs which results from the mapping to NPLA step, the layout of each PLA is drawn using the TSMC 0.25 $\mu$ m process. Additionally, the layout of IO pads, ESD cells, IO drivers and power rails are drawn using TSMC layout guidelines.
- 9) Layout Versus Schematic (LVS) verification was performed next to ensure that there were no layout errors. We used the Assura LVS checker for this step.
- 10) Finally, the design parasitics were extracted, and the entire design is simulated in SPICE as a final sign-off.

#### D. Implementation Considerations

1) *Digital NPLA portion: PLA Design and Size:* All the PLAs in our design are of the precharged NOR NOR type, and have a fixed number of inputs (8), outputs (6) and cubes (12). This was found to be a good size for the design based on logic synthesis results in terms of number of PLAs needed and throughput achievable. We tried several medium sized PLAs (5-15 inputs, 3-8 outputs and 10-20 rows), and over several test examples, found 8 inputs, 6 outputs and 12 cubes to be optimal. The digital portion of the transmitter was synthesized into a total of 33 same size PLAs.

Also, PLA folding is used to allow a PLA to implement more complex logic without increasing its area. Folding is done by running two unconnected *bit-lines* corresponding to two different inputs on the same track, in the AND plane of a PLA. One of the *bit-lines* start from the top of the PLA and the other one starts from the bottom. In this way, more cubes can be fitted into the PLA in compact way.

Initial simulations using HSPICE [7], showed that precharge and evaluate time for the 8 input, 6 output, 12 cube NOR NOR PLA were,  $T_{pchg} = 45ns$  and  $T_{eval} = 35ns$ , for the TSMC 0.25 $\mu$ m process.

The maximum number of levels needed for the slowest combinational block for this design is 19. This gives us an estimate of the throughput from Equation 3 as approximately 1.4MHz, if we use  $T_{pchg} = 45ns$  and  $T_{eval} = 35ns$ . Now the

two tones produced have to be less than half the throughput according to the Nyquist Sampling theorem. We choose the frequency of the second tone as

$$f_2 = \frac{f_{clk} \times 117}{512} \quad (4)$$

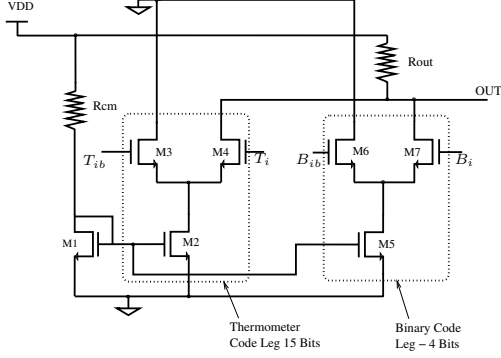
We choose the first tone to have a frequency three times less than that of  $f_2$ . If we assume a clock frequency of 1MHz, we get  $f_1 = 115kHz$  and  $f_2 = 350kHz$ .

Now we need to choose a reference PLA out of the chain of PLAs in the network. The *completion* signal of this reference PLA will be used as the reference circuit delay for the delay compensation circuit. Since there are many levels of PLAs in the synthesized network of PLAs, it is best to choose a PLA which completes its evaluation at approximately half the time it takes the entire network of PLAs to complete its evaluation. This is because the completion signal of the reference PLA would transition to a LOW value during the middle of the evaluation time span of the *CLK* signal. This gives the *BCLK* signal sufficient head room on both sides of the completion signal, to be able to generate equally long *pullup* or *pulldown* signals. In our case, we use a PLA at logical depth 10 (out of a maximum of 19) as the reference PLA.

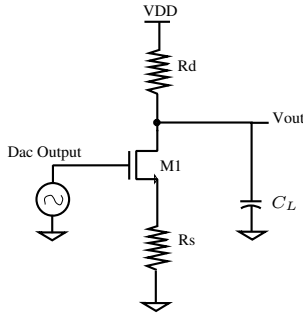
#### 2) Analog Portion: Digital to Analog Converter (DAC):

The circuit diagram of the 8-bit DAC is shown in Figure 6 (a). The DAC has a reference current mirror, M1 biased by resistor  $R_{cm}$ . It also has as many current mirrors reflecting the reference as the number of input bits. The input to the DAC is a 19bit digital signal. The top 15 bits are generated by thermometer encoding on the 4 MSBs of the DAC input. The 4 LSBs are binary encoded. Hence the DAC will have 19 current mirror legs. Figure 6 (a) shows two of the current mirror legs of the DAC (one leg shows the binary coded circuit, and the other is the thermometer coded circuit). The inputs  $T_i$  and  $T_{ib}$  are the  $i^{th}$  thermometer encoded bit and its complement. The inputs  $B_i$  and  $B_{ib}$  are the  $i^{th}$  binary encoded bit and its complement. The DAC works by switching the current mirrors ON depending on the value of the input bits and measuring the voltage across the  $R_{out}$  resistor due to this current. The input bits control the NMOS transistors, M3, M4, M6 and M7. For any of these legs, if the input bit is LOW, then the NMOS on the left i.e. M3 or M6 turns ON and prevents the current mirror leg from conducting current. If the input bit is HIGH, then the NMOS on the right turns ON and allows the leg to mirror the current in the reference transistor M1. The difference between the current mirrors for the thermometer code and the binary code is in the size difference between M2 and M5. The W/L of M5 used in the current mirrors for the binary encoded bits are 1.3,2.6,5.2,10.4 from LSB to MSB. The W/L ratio doubles for every next MSB. The transistors corresponding to M2 have a W/L of 20.8 for all the current mirror legs for the 15 thermometer encoded bits. This allows the DAC to modulate the voltage at OUT based on the weighted current flowing through  $R_{out}$  and through different current mirror legs. The resistors  $R_{cm}$  and  $R_{out}$  of the DAC are designed to be surface

mounted resistors outside the chip. This will allow us to tune these resistors in real time to enhance the output signal. Two external pins in the pin-out of the chip is reserved for these two resistors. The choice of a thermometer encoding for the 4 MSBs of the DAC is made to improve differential non-linearity (DNL) and minimize glitch energy.



(a) 8-bit Digital to Analog Converter



(b) Amplifier

Fig. 6. Analog Circuit Schematic

### Common Source Amplifier:

A common source amplifier is needed at the output of the DAC to amplify the signal and drive the antenna. The common source configuration is shown in Figure 6 (b). The common source amplifier is an inverting amplifier. In this configuration, note that there are no bias resistors biasing the gate of the transistor M1. The gate of M1 is connected to the output of the DAC. The gate is thus biased by the DC component of the sinusoidal voltage from the output of the DAC. The amplifier is powered by a very low  $VDD$  of 0.7V. Under this condition, other amplifiers such as the source follower or common drain amplifier do not function correctly.

**Pin Out and I/O Pad Cells:** A standard cell based implementation of the BFSK transmitter is also performed for comparison purposes. The two circuit realizations operate at different  $VDD$  values. In order to isolate these two implementations, we need one extra voltage domain for the standard cell implementation. This will be a 2.5V domain which is the nominal operating voltage for the TSMC 0.25 $\mu$ m

process. For the targeted process, we have specified the sub-threshold design to work at a  $VDD$  of 0.6V. The inputs to the sub-threshold digital modulator circuit cannot be on the same voltage domain. This is because designing I/O drivers at such a low voltage resulted in extremely large devices. Hence we use another voltage domain (higher than 0.6V) so that the inputs to the sub-threshold circuit are driven at this higher voltage. We have chosen the  $VDD$  of this domain to be 1V. One of the built in testability features of this chip is that the outputs of the sub-threshold digital modulator circuit, if needed, can be sent directly off-chip to an external DAC and antenna. We however found that there was no off-the-shelf DAC that had an input voltage rating of less than 2V. Hence the outputs of the sub-threshold circuit needed to be driven to a voltage value of at least 2V. Hence another voltage domain with a  $VDD$  of 2V was used.

We thus have four separate  $VDD$  domains on the chip. All these domains have a common  $GND$  to make the power distribution easier. The following special conditions need to be addressed when we have signals that cross two different voltage domains.

- 1) A higher voltage signal cannot drive a pass gate of a lower voltage domain. In this case we buffer the signal with a buffer operating on the  $VDD$  of the lower voltage domain before driving the pass gate.
- 2) A higher voltage signal can drive the gate of a transistor in a lower voltage domain.
- 3) To buffer a signal from lower voltage to higher voltage domain, we use custom designed level shifters.

Due to the different voltage domains, we required 5 different types of custom made I/O Cells. The total number of pins used were 72. Out of these pins, there were 18  $VDD$  pins and 18  $GND$  pins. The  $VDD$  or  $GND$  pins were used to shield sensitive analog I/O pins and clock signals, while designing the IO frame pinout.

### E. Test Strategies

**1) Testability and Redundancy:** Various testability features were built into the design. The use of these features is to test each component of the chip individually to verify functionality. They also serve as a backup against failure of one of the components. The following are the testability features that are incorporated in the design.

- 1) A standalone PLA is included in the design along with the other PLA components which make up the digital modulator circuit. The PLA is designed in such a way that the two outputs (which are brought out as pins on the packaged die) of the PLA toggle continuously when the clock waveform is applied. The result of this test verifies the baseline functionality of the PLAs which are the basic building blocks in the design.
- 2) The  $Nbulk$  node of the PLAs is connected to an external pin. This is to enable us to verify the functionality of the dynamic delay compensation circuit.
- 3) The 8bit output of the NCO block is directly sent to 8 I/O pads on the chip. These pads are bi-directional. This



means that these pads on the chip can either be used to get the digital 8bit sine-wave value from the output of the NCO, or can be used as an 8bit input to the binary to thermometer code converter. This feature is important since it takes into account the scenario in which only one of the digital modulator or the DAC is functionally correct. In this scenario, these bi-directional pins may be used to excite the correctly functioning blocks in the design.

- 4) The output of the DAC can be measured using an oscilloscope, at the pin which connects the external DAC drive resistor  $R_{out}$  to the chip. This allows the DAC to be tuned and tested individually based on its output waveform. This also gives us the option of directly using the DAC with an external amplifier and antenna, which we implemented.
- 5) The output of the common source amplifier also can be scoped externally using the pin connected to the  $R_D$  resistor. This signal may also modulate an off-chip antenna, instead of the on-chip antenna.
- 6) The output of the amplifier is connected to the antenna through a pass gate that is controlled by a signal called *Anton*. This signal is used to disconnect the on-chip coil antenna by turning off the pass gate if needed. This would ensure that the capacitance of the on-chip antenna is not switched when we use the external antenna.

Note that many of the above features were included as a fallback option, in case significant portions of the design failed. In reality our silicon was 100% functional and operates at speed.

### III. IMPLEMENTATION RESULTS AND COMPARISONS

#### A. Physical Description of the Design

Each of the PLAs have the same number of inputs, outputs and cubes. The logic implemented by the PLAs however is different. The transistors connected to the bitlines, wordlines and output lines need to be changed for each of the PLAs depending on the function implemented. The footprint of each PLA however remains the same. The layout of the DAC and the amplifier are also done. The transistor lengths used for these analog components are three times the minimum length. This increases the variation tolerance of these components. The standard cell layout is done using the SEDSM tool [8]. This layout is merged with the rest of the components to get the entire die layout. The die photo is shown in Figure 7.

The size of the die is  $10mm^2$ . The design has a total of 20882 transistors. The chip is fabricated using a TSMC  $0.25\mu m$  technology and is packaged in an 80 pin QFN package.

#### B. Test Results

In this section, we present results from the fabricated die. The range of operation of the circuit is tested. The functionality of the dynamic body bias delay compensation circuit is also verified. The sub-threshold implementation is compared with a standard cell based implementation of the BFSK circuit, which was also implemented on the same die. We have also

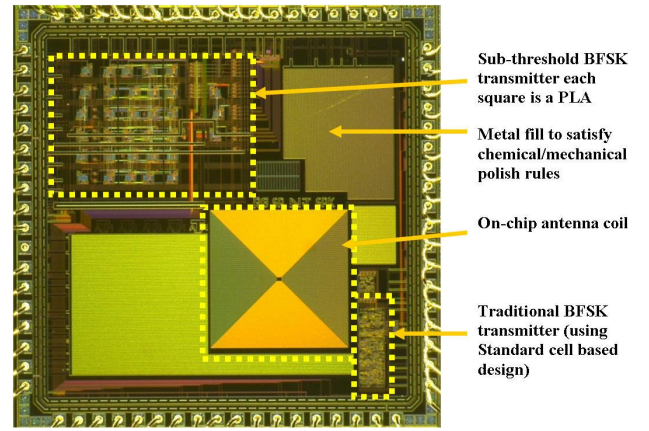
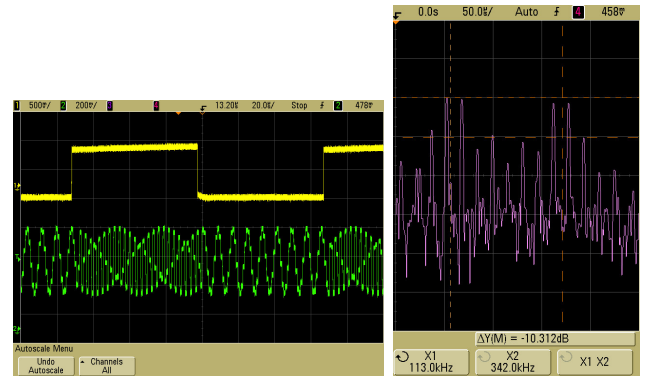


Fig. 7. Die Photo



(a) BFSK Modulation

(b) FFT of Modulated Signal

Fig. 8. BFSK Functionality

implemented an FPGA based receiver design (not discussed in detail due to lack of space) which was found to correctly demodulate the signal from our chip.

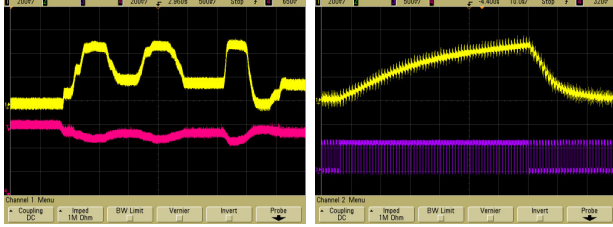
1) *Functional Verification:* The  $VDD$  domains 1 and 4, which correspond to the sub-threshold BFSK inputs, and DAC and amplifier outputs are powered ON. The *reset* signal is held LOW. The DAC and Amplifier are biased using resistances determined during the circuit design phase. The output of the DAC for an input signal that makes a LOW to HIGH transition is shown in Figure 8 (a). Note that the DAC output clearly shows two tones depending on the value of the input. The Fast Fourier Transform (FFT) of the output of the DAC is shown in Figure 8 (b).

Here the input bit-stream is continually alternating between a logical “zero” and a logical “one” at a frequency of 31.25kHz. The clock frequency,  $f_{clk}$  of the sub-threshold circuit is set at 1MHz. From the FFT we see the two transmitted tones at 113kHz and 342kHz respectively.

Notice that the secondary unwanted peak between the two tones is around -11dB below the fundamental tones as shown in Figure 8 (b). Also through our FPGA based receiver we found that a signal with a spectrum that has the secondary

unwanted peak at -10dB was demodulated correctly at the receiver side. The receiver architecture used was a standard receiver for demodulating non-coherent BFSK signals [9].

2) *Dynamic Compensation Circuit*: The dynamic compensation circuit stabilizes circuit delay by modulating the  $N_{bulk}$  node of the NMOS transistors in the design as explained in Section II-B.3. Figure 9 (a) shows an oscilloscope plot of the  $N_{bulk}$  node voltage (light signal) and  $V_{DD}$  (dark signal) of the sub-threshold circuit.



(a) Bulk Node Voltage Modulation with  $V_{DD}$

(b) Bulk Node Voltage Modulation with  $BeatClock$

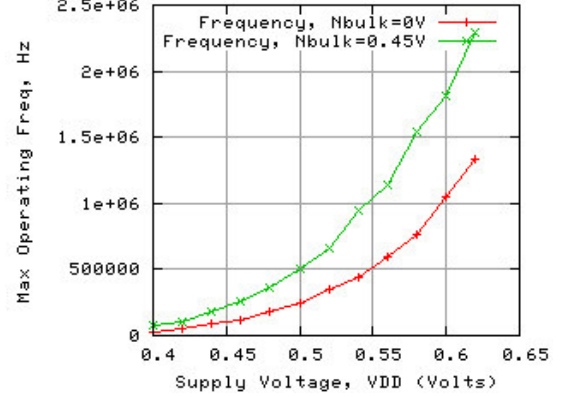
Fig. 9. Dynamic Delay Compensation by Modulating  $N_{Bulk}$  node

Here the external beat clock has been fixed to a particular delay. Notice that when the supply voltage (which is the bottom signal in the plot) fluctuates from its nominal value, the bulk node voltage (which is the top signal in the plot) is immediately modulated in the opposite direction to compensate the circuit delay with respect to power supply variation. Thus the reference circuit delay is kept in phase with the external reference signal.

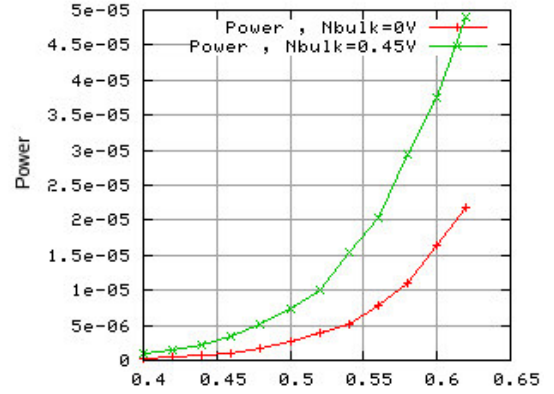
Figure 9 (b) plots the bulk node voltage in the top half and the external beat clock signal in the bottom half. Here the phase of the *beat clock* signal is changed alternatively. Initially it is set to run the PLAs as fast as possible. Due to this the charge pump forward biases the  $N_{bulk}$  node and the circuit speeds up. When the *beat clock* signal phase changes to slow down the PLAs, the  $N_{bulk}$  node is driven low and the circuit slows down. The  $N_{bulk}$  node is clearly modulated up and down when the phase of the *beat clock* signal changes verifying the operation of the dynamic body bias circuit with respect to the external reference signal.

3) *Operating Ranges*: The supply voltage for the digital BFSK modulator circuit was varied from 0.4V to 0.62V. Note that  $V_T$  for the TSMC 0.25 $\mu$ m process was about 0.54V. The maximum frequency of operation at these voltages was determined by observing the output of the source amplifier. When the frequency is too high, the sine wave at the output of the amplifier gets distorted and the output cannot be demodulated. The maximum operating frequencies over a set of supply voltages is plotted in Figure 10 (a).

This figure shows two curves which correspond to a bulk node voltage value of 0V and 0.45V respectively. This plot shows the range of frequencies over which the dynamic compensation circuit can track the reference beat clock. Notice that the maximum speed of operation increases quadratically as the supply voltage increases.



(a) Maximum Operating Frequency



(b) Power Consumed at Maximum Operating Frequency

Fig. 10. Chip Operating Ranges

The power consumed by the circuit at these operating voltages and frequencies is shown in Figure 10 (b). The power consumed is plotted for the maximum and minimum voltage value that the  $N_{bulk}$  node can have. The power consumed is the product of the average current flowing through the digital BFSK modulator voltage source. Note that a different voltage source is used for the DAC and the amplifier.

4) *Comparison with Standard Cells*: The power consumed by the sub-threshold BFSK modulator was compared with the power consumed by the standard cell BFSK implementation. This is shown in Table II. From this table we see that the **power consumed by the Standard Cell based circuit implementation is 19.4 $\times$  more**. The standard cell based design is specified to operate at a supply voltage of 2.5V. Note that the standard cell based design is capable of operating at higher speeds. The standard cell design does not have any compensation scheme that compensates circuit delay for PVT variations (which are higher when operating near the sub-threshold region) hence it would not function correctly in sub-threshold under varying operating conditions.



TABLE II  
SUB-THRESHOLD VS STANDARD CELL POWER CONSUMPTION

Design Style	VDD	Clock Frequency	Average Current	Power Dissipation
Sub-threshold	0.6	1.05MHz	44.7 $\mu$ A	26.8 $\mu$ W
Standard Cell	2.5	1.05MHz	208.0 $\mu$ A	520.0 $\mu$ W

#### IV. CONCLUSIONS

Power Consumption a key issue in the semiconductor industry today. Sub-threshold circuit design is an ultra-low power technique that can be used in applications where speed is not a primary concern. However sub-threshold designs are hard to center, as they exhibit an exponential sensitivity to process, voltage and temperature (PVT) variations. In this paper we implemented and tested a robust sub-threshold design flow which uses circuit level PVT compensation to stabilize circuit performance. We designed and fabricated a sub-threshold BFSK transmitter chip. The transmitter was tested to transmit baseband signals up to a data rate of 32kbps. Experiments using the fabricated die, verify the functionality of the design show that the sub-threshold circuit consumes **19.4** $\times$  lower power than the traditional standard cell based implementation on the same die. There are several future directions that this paper can take. With a small modification the BFSK transmitter can be used as a Software Defined Radio (SDR) in which the transmitted frequencies are flexibly programmed during runtime. Also from simulations we expect a 100-500 $\times$  reduction in power when a smaller process such as 0.1 $\mu$ m and 0.07 $\mu$ m is used. Also a natural extension of this project is to realize the entire transceiver system on the same die.

#### REFERENCES

- [1] N. Jayakumar and S. Khatri, "A Variation-tolerant Sub-threshold Design Approach," in *Proceedings, Design Automation Conference*, pp. 716–719, June 2005.
- [2] Y. Cao, T. Sato, D. Sylvester, M. Orshansky, and C. Hu, "New Paradigm of Predictive MOSFET and Interconnect Modeling for Early Circuit Design," in *Proc. of IEEE Custom Integrated Circuit Conference*, pp. 201–204, Jun 2000. <http://www-device.eecs.berkeley.edu/~ptm>.
- [3] J. Tschanz, J. Kao, S. Narendra, R. Nair, D. Antoniadis, A. Chandrakasan, and V. De, "Adaptive Body Bias for Reducing Impacts of Die-to-Die and Within-die Parameter Variations on Microprocessor Frequency and Leakage," *IEEE Journal of Solid-State Circuits*, vol. 37, pp. 1396–1402, Nov 2002.
- [4] S. Khatri, R. Brayton, and A. Sangiovanni-Vincentelli, "Cross-talk Immune VLSI Design Using a Network of PLAs Embedded in a Regular Layout Fabric," in *IEEE/ACM International Conference on Computer Aided Design*, pp. 412–418, Nov 2000.
- [5] E. M. Sentovich, K. J. Singh, L. Lavagno, C. Moon, R. Murgai, A. Saldanha, H. Savoj, P. R. Stephan, R. K. Brayton, and A. L. Sangiovanni-Vincentelli, "SIS: A System for Sequential Circuit Synthesis," Tech. Rep. UCB/ERL M92/41, Univ. of California, Berkeley, CA 94720, May 1992.
- [6] S. P. Khatri, *Cross-talk Noise Immune VLSI Design Using Regular Layout Fabrics*. PhD thesis, EECS Department, University of California, Berkeley, CA, Dec 1999.
- [7] "HSPICE." [www.synopsys.com/products/mixedsignal/hspice/hspice.html](http://www.synopsys.com/products/mixedsignal/hspice/hspice.html), May 2007.
- [8] Cadence Design Systems Inc., 555 River Oaks Parkway, San Jose, CA 95134, *Envisia Silicon Ensemble Place-and-route Reference*, Nov 1999.
- [9] F. Xiong, *Digital Modulation Techniques, Second Edition (Artech House Telecommunications Library)*. Norwood, MA: Artech House, Inc., 2006.