LOW POWER SCANNER FOR HIGH-DENSITY ELECTRODE ARRAY NEURAL RECORDING

A Thesis work submitted to the faculty of
San Francisco State University
In Partial Fulfillment of
the Requirements for
the Degree

Master in Sciences
In
Engineering: Embedded Electrical and Computer Systems

By
ELI LYONS
San Francisco, California
September, 2011

CERTIFICATION OF APPROVAL

I certify that I have read *Low Power Scanner for High-Density Electrode Array Neural Recording* by Eli Lyons, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree: Mast of Sciences in Engineering at San Francisco State University.

_____ _____

Dr. Hamid Mahmoodi Date

Professor of Electrical and Computer Engineering

_____ _____

Dr. Hao Jiang Date

Assistant Professor of Electrical and Computer Engineering

ABSTRACT:

LOW POWER SCANNER FOR HIGH-DENSITY
ELECTRODE ARRAY NEURAL RECORDING

Eli Lyons

San Francisco, California

2011

There is an increasing amount of interest in measuring signals from electrogenic cells, such as neurons or cardiomyocytes. A current trend is to increase the density of electrodes for higher resolution readings. Reading from all locations in a large electrode array, for example an array with 10,000 electrodes, leads to significant power consumption in interfacing ICs. Excessive power consumption not only reduces battery life in implantable applications, but also raises the temperature and damages biological cells. The primary goal of this research is to design a low power integrated system that can be used in vivo for scanning the electrode arrays. The design presented provides three main modes of operation implemented in a low power fashion. A model created in Python provides input vectors and output comparison for the verification process of the Auto Calibration mode. Clock Gating is applied to the design with a reduction of power in the gate level simulation of more than 70% for all three modes of operation. Further power savings are observed with the application of Power Gating.

I certify that the Abstract is a correct representation of the content of this thesis.

_____ _____

Chair, Thesis Committee Date

DEDICATION

To the kids in afterschool detention with their arms folded and head down, keep your heads up.

ACKNOWLEDGEMENTS

I would like to thank my advisor and mentor, Dr. Hamid Mahmoodi. I would not have gotten this far without his encouragement and support. I would also like to thank my aunts and uncles from the sometimes neurotic, but always courageous Lyons family, and my friends for their invaluable support as well. Finally, I would like to thank my grandfather George W. Lyons. Before he passed he recorded a message encouraging me to keep studying, I remember this often.

TABLE OF CONTENTS

CONTENTS **PAGE**

LIST OF FIGURES

INTRODUCTION

There is an increasing amount of interest in measuring signals from electrogenic cells, such as neurons or cardiomyocytes. There is currently a large research effort into methods of reading from and stimulating neurons in the brain. Not and such systems are referred to as Brain Machine Interfaces or

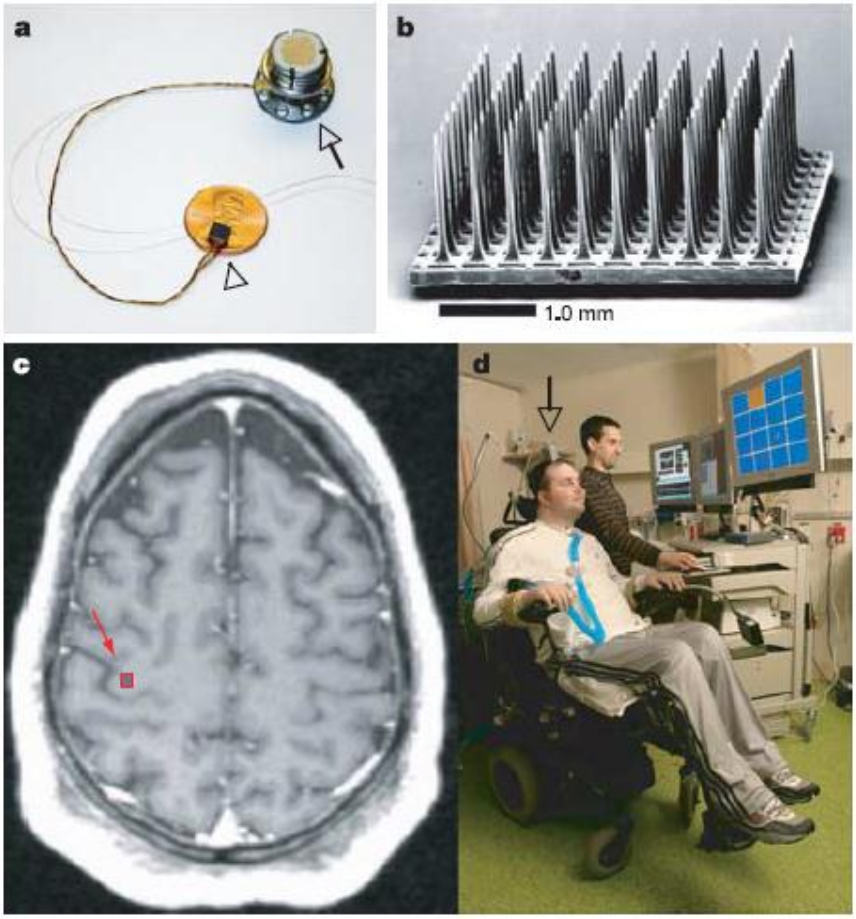


Figure 1: A paralyzed participant using a Brain Machine Interface (BMI) system to control the position of an orange square on a screen

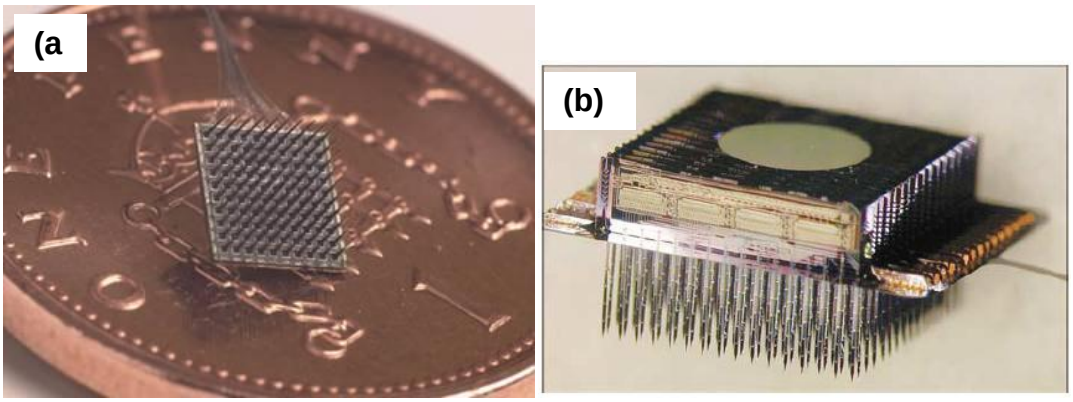


Figure 1: (a) A micro-electrode array containing around 100 electrodes. (b) An integrated IC and micro-electrode array system.

DESIGN FLOW

The goal of this research was to complete a front end design, meaning that the backend chip layout and corresponding re-verification of timing was not included in the scope of this research. The design flow is shown in figure *****. As is to be expected, most of the time in this research was in design and verification. Design includes designing the low power scanner in System Verilog, and verification was done by writing test benches in Verilog and using the Synopsys VCS tool to run the test benches on the scanner. Verification will be discussed in further detail in the section ***Verification.

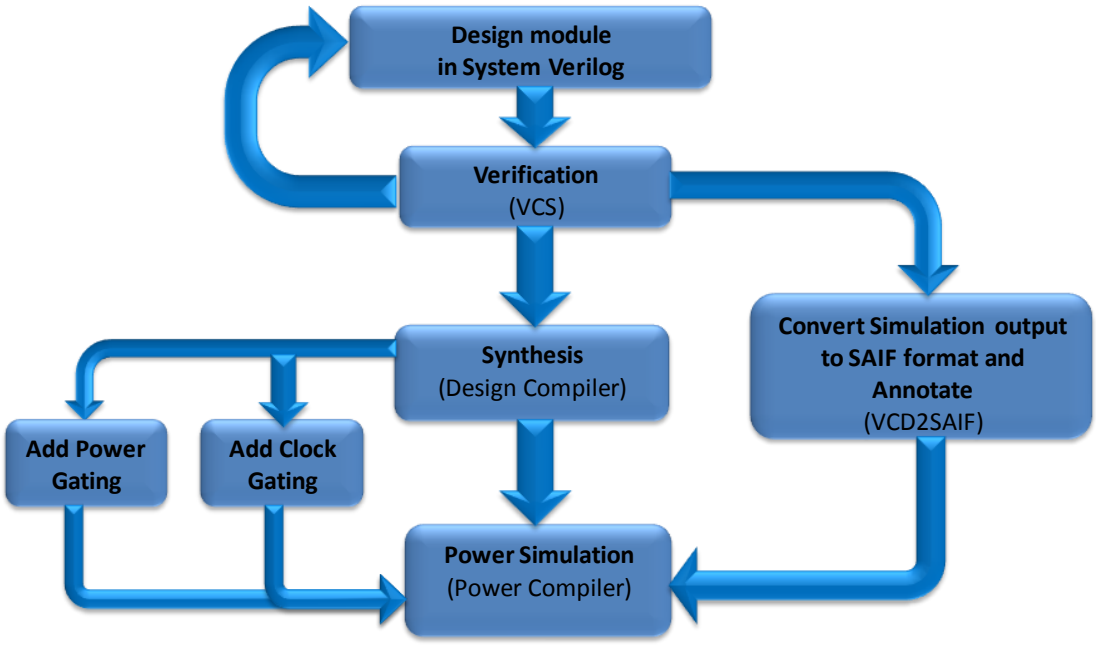


Figure 2: Design flow for front end of low power scanner.

When the test benches are run in VCS, a dump file is created that serves as a record of all the external signal values applied to the scanner under test, and the corresponding internal signal values of the scanner. This dump file can be converted to the Switching Activity Interchange Format (SAIF) that can later be annotated to the power simulation using the Synopsys Power Compiler tool. Annotating the SAIF file increases the accuracy of the power simulation since it bases the simulation on the switching activity that was simulated in the test bench.

Synthesis of the System Verilog scanner design into a gate level design was done using the Synopsys Design Compiler (DC) tool. A script was created in the Tool Command Language (Tcl) to setup the environment and options to run the DC tool. Clock gating is now an option that can be set when DC is executed and will cause DC to automatically apply clock gating to the design. In order to apply power gating in synthesis, a power controller is added to the scanner design in System Verilog and a script is Unified Power Format must be created that is loaded in DC when it is executed. Unfortunately, synthesis was not successful with the UPF, and a separate project to get the UPF working has been taken on by another student. Fortunately, a manual estimate of Leakage power under power gating allows for total energy calculations that should still be quite accurate and this will be discussed in the Power Simulation Results Chapter.

DESIGN OVERVIEW

The first step in the design process was defining basic specifications such as sampling rate and size of electrode array. The minimum nyquist rate for a neural signal is 10 kHz with 8-bit sampling suggested [**], and it was decided that the design would have 10 ADCs since a state of the art non-embedded design was designed with 16 ADCs [***]. It is assumed that each ADC could at least run at 15 MSPS, based on current offerings by vendors that support such sampling rates or higher at 8 bits per sample.

A minimum sampling rate of 10 kHz means that a sample from an electrode we would like to read the signal from must be taken every 100 microseconds. A burst mode method is used as a basis in the overall design strategy to reduce power of the scanner module without adding a lot of additional overhead. The objective of burst mode is to complete the required work quickly and then go into a sleep state. During the sleep state, no more dynamic power is consumed and leakage power can be reduced by using power gating. This strategy can be effectively implemented because the work load of the scanner can be known in advance and will only change depending on the mode of operation the user selects.

The workload of the scanner was an important early design decision based on a few different factors. It was decided early on that each ADC

would only have access to part of the electrode array, instead of each ADC having access to every electrode or having some overlap in access. This reduces complexity in interconnect connecting ADCs to electrodes. It was then decided that the each ADC should scan the same number of electrodes to keep timing and design simple. It was decided that each ADC would be able to scan 1000 electrodes and therefore the scanner could scan from 10,000 electrodes total.

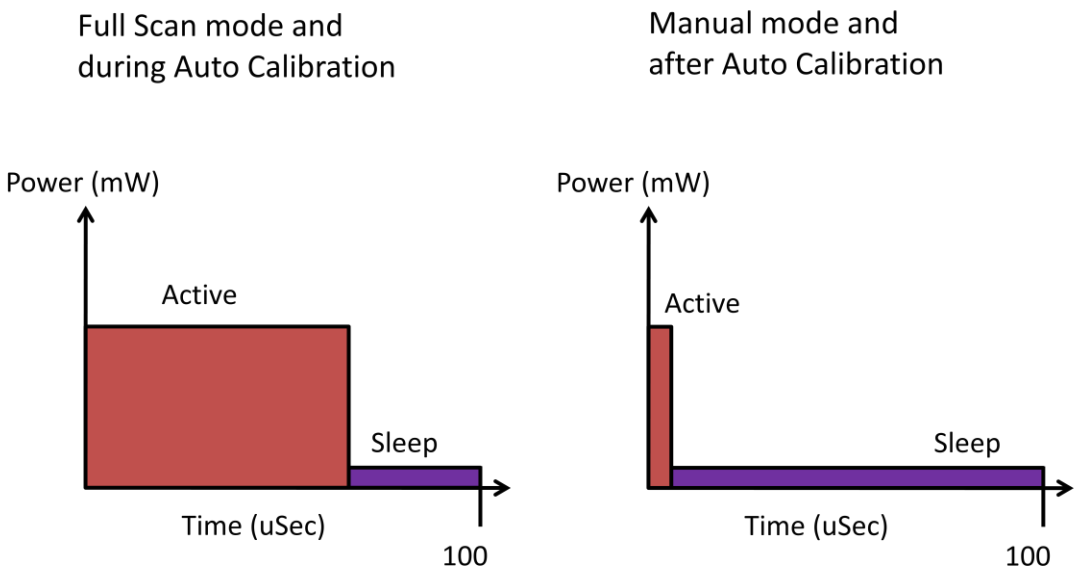


Figure 3: Illustration of power consumption using the burst mode strategy.

An overview of the state diagram of the Scanner is shown in Figure **. The design runs in three different modes, each of which have a sleep state. The complete state diagram for each mode is given in figures **, ** and **. Even though each mode has a sleep state in it, the amount of time spent in the sleep state is much shorter in Manual mode and in Auto Calibration mode

(after calibration period has ended) than in the Full Scan mode. Therefore, running the Scanner in Manual mode and Auto Calibration mode has a large power savings advantage at the cost of how many electrodes are scanned.

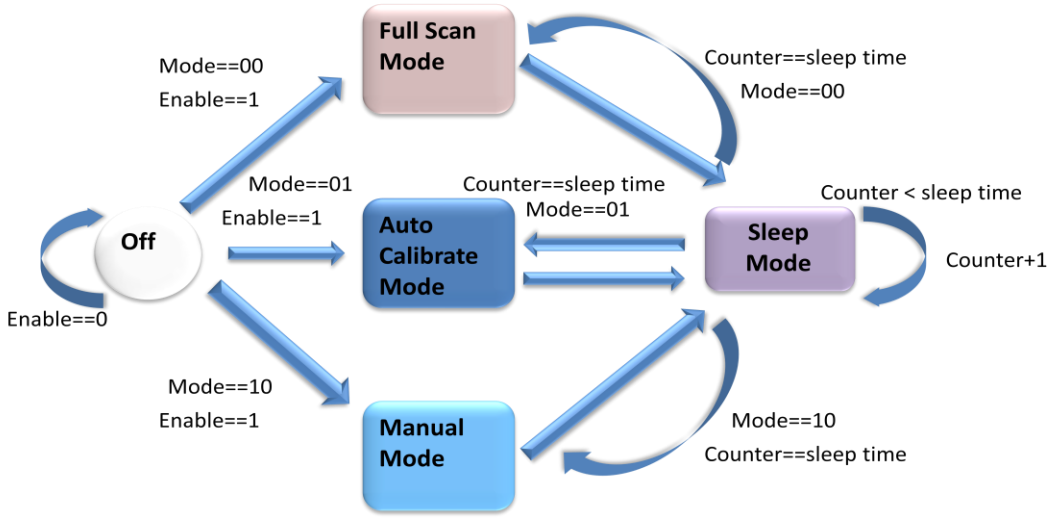


Figure 4: Overview of states in the low power scanner.

In Full Scan mode, every electrode location is read from in a period of operation. After all electrodes are scanned the Scanner enters sleep mode until the array must be scanned again.

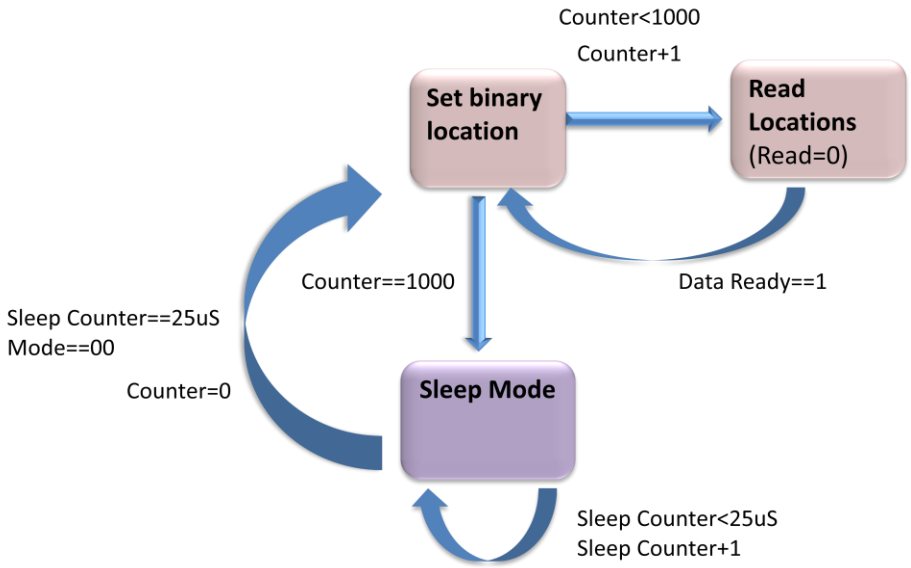


Figure 5: State diagram for the Full Scan mode.

In Manual mode, the user load 100 locations they would like to read and only those 100 locations are read from, meaning each ADC only reads 10 locations instead of 1000 in Full Scan mode. This is why manual mode is why the dynamic power in manual mode is so much smaller than in Full Scan mode which will be discussed in the Power Simulations Results.

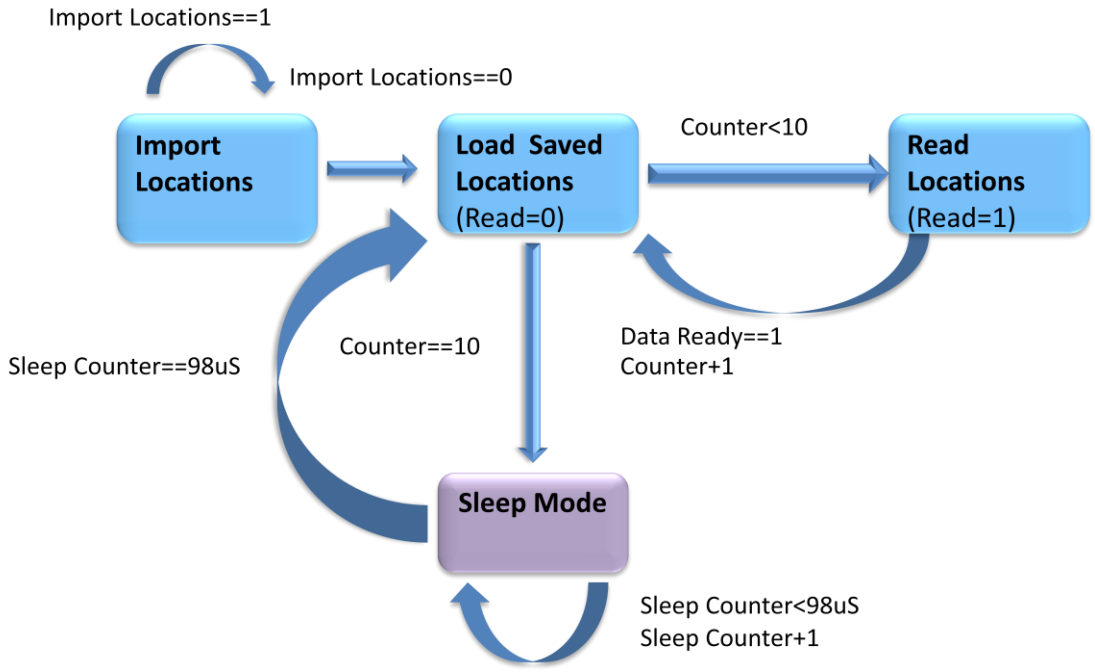


Figure 6: State diagram for Manual mode.

In Auto Calibration mode the whole array is scanned during the user set calibration period. The 100 by 100 array is split up into one hundred 10 by 10 regions. In each region the location with the highest measured signal is defined as the "best location" to read from. After each scan of the array, the new values in each region are compared with the value of the current best location, and if the new value is higher the best location will be replaced with the location of the new value and the new value will be saved. After the calibration period is over, only the best location in each region (100 locations total) are read from, reducing the power in comparison to full scan mode. This Auto Calibration mode is only one possible algorithm to reduce power by being selective about which electrodes are read from, and there is potential

for future research on this topic. Manual mode and Auto Calibration mode (after calibration period) spend the same amount of time setting locations for reading which is much shorter than in Full Scan mode, hence operation in Manual and Auto Calibration mode will allow the module to spend more time in the sleep state greatly reducing power.

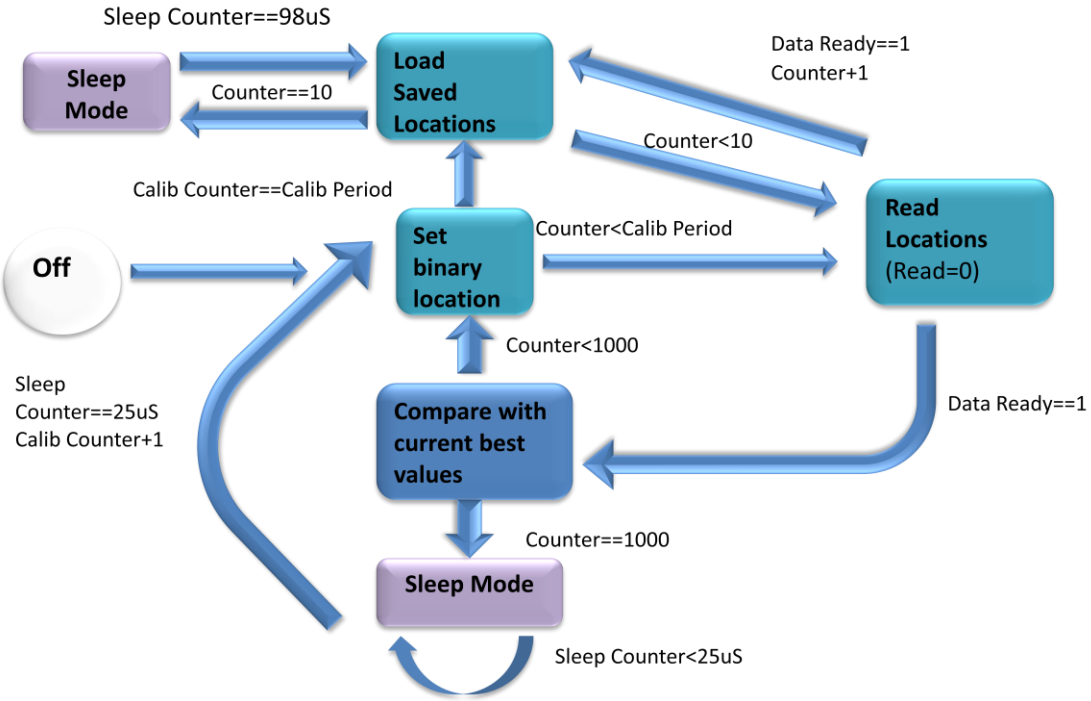


Figure 7: State diagram for Auto Calibration mode.

After the state diagram for each mode was created, the number of total states was reduced by finding shared states between modes and combining the next state conditions into the shared state. The final state diagram from

this process is shown in Figure **. The sleep state is one state drawn at different locations and the next state conditions are not shown for readability.

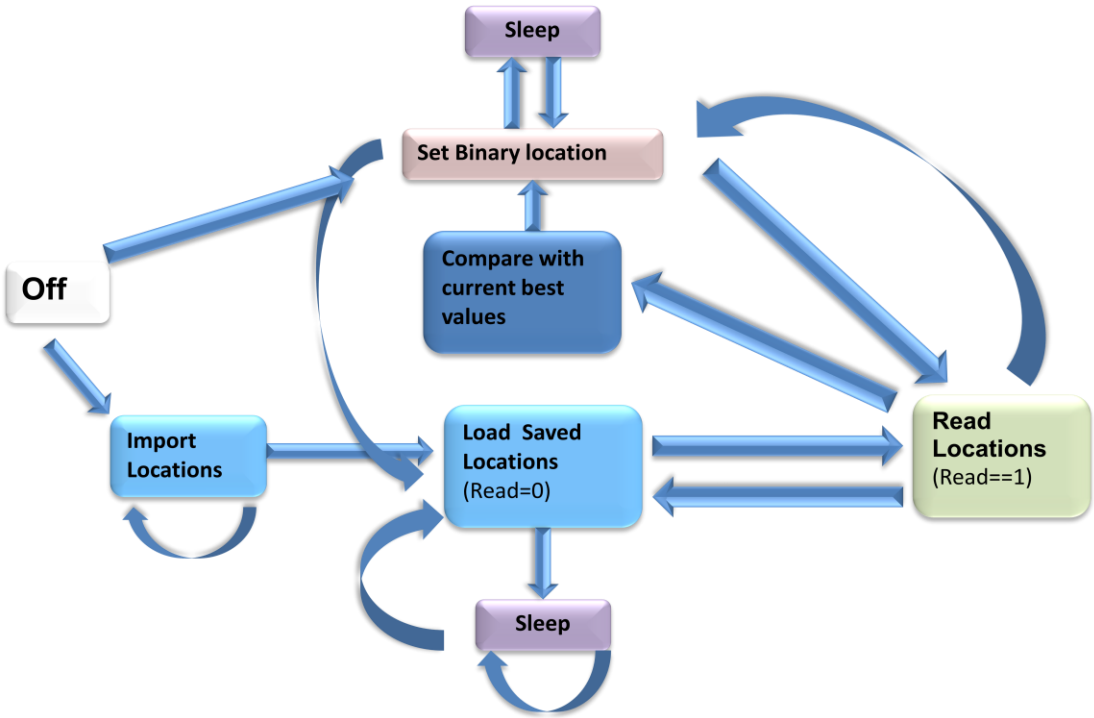


Figure 8: Reduced state Diagram for low power scanner design.

VERIFICATION

It has been said there is no such thing as design, only verification. This thesis research is certainly evidence of some truth to this statement, as the verification step required the most time in the design flow. A summary of the test benches simulated with the Low Power Scanner in VCS are given in Table 1. Three test benches tested the Scanner in different operating modes and one test bench tested the power controller. The final test bench tested the Scanner with power gating in the auto calibration thoroughly and will be discussed in detail.

Test Bench	Purpose
full_tb.v	Tests the controller in full scan mode
manual_tb.v	Tests the controller under manual mode (load 100 locations and read from those locations)
auto_tb.v	Tests the controller in auto calibrate mode
power_tb.v	A testbench for only the power gating controller
full_gating_tb.v	Tests the controller with power gating module in full scan mode

Table 1: Summary of Test Benches.

A “golden model” was created to simulate the operation of the Scanner in Auto Calibration mode. The Python scripting language was chosen for this application because of my familiarity with the language, and support for many MatLAB like features. The first objective in making the model was to create input vectors for test benches that test the Scanner in Auto Calibration mode, and these input vectors should loosely simulate signals that could be expected to come from the ADCs. This simulation was created not only for verification of the Low Power Scanner, but as a verification tool that could be adapted to verification for other designs as well. Therefore, it offers the user a lot of flexibility when running the simulation which will be further explained.

A three dimensional Sinc function was used as the basis for neural spike signal model. The sinc function is defined in (1) , and is defined as 1 at x=0 due to the limit of $\frac{\sin x}{x}$ as x goes to zero.

(1) $$sinc(x) = \begin{cases} 1\ , & for\ x = 0 \\ \frac{sinx}{x}\ , & otherwise \end{cases}$$

To make this function three dimensional sinc function Z, x is replace with R, which accomplishes rotation around the Z-axis.

(2) $$R = \sqrt{X^2 + Y^2}\ ;\ Z = \frac{\sin(R)}{\mathrm{R}}$$

This function is further manipulated to allow for variable amplitude (A), location (Xs and Ys), and spike width factor (W).

(3) $R = \sqrt{(X - Xs)^2 + (Y - Ys)^2}$; $Z = \frac{\mathrm{A} * \sin(R * W)}{\mathrm{R}}$

The Python script allows the user to enter the number of spikes they would like to simulate, the width factor, the minimum width between spikes, the number of frames to produce, and the size of the array (length of X and Y axis). After the user sets these parameters the simulation is executed.

First random locations are selected for each spike taking the user specified minimum spike separation distance into account and within the user specified X and Y boundaries. Each spike is then given a pseudorandom final amplitude and a pseudorandom delay associated with it (the range of these can be easily modified in the script). Assigning each spike a slightly different final amplitude and delay time for the spike to start growing creates more realistic input vectors, since it isn't expected that neurons will all fire at the same time and the exact same amplitude will be read from the electrodes. Each frame of the amplitude increases the amplitude of the spikes (when the individual spike delay is over) until the final amplitude is reached, at which point the amplitude in each from is decreased for those spikes. Therefore, some spikes will appear to increase in amplitude while others are disappearing. Each frame is exported as an image and also a text file that

records the amplitude value in binary at each X and Y location. This file can be imported into a test bench during verification in VCS. Figure *** illustrates an example of a frame from a simulation.

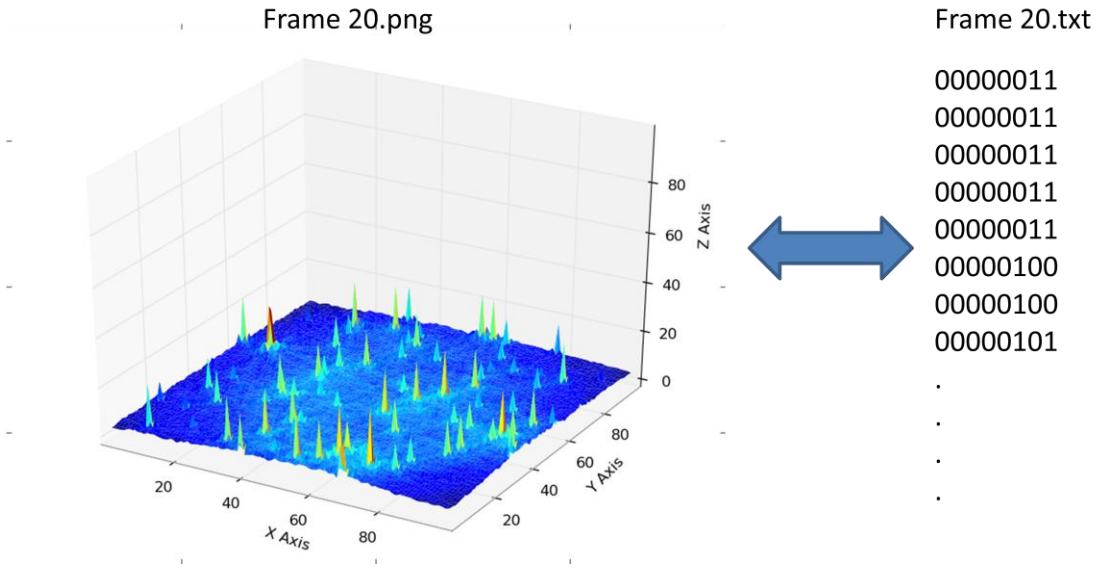
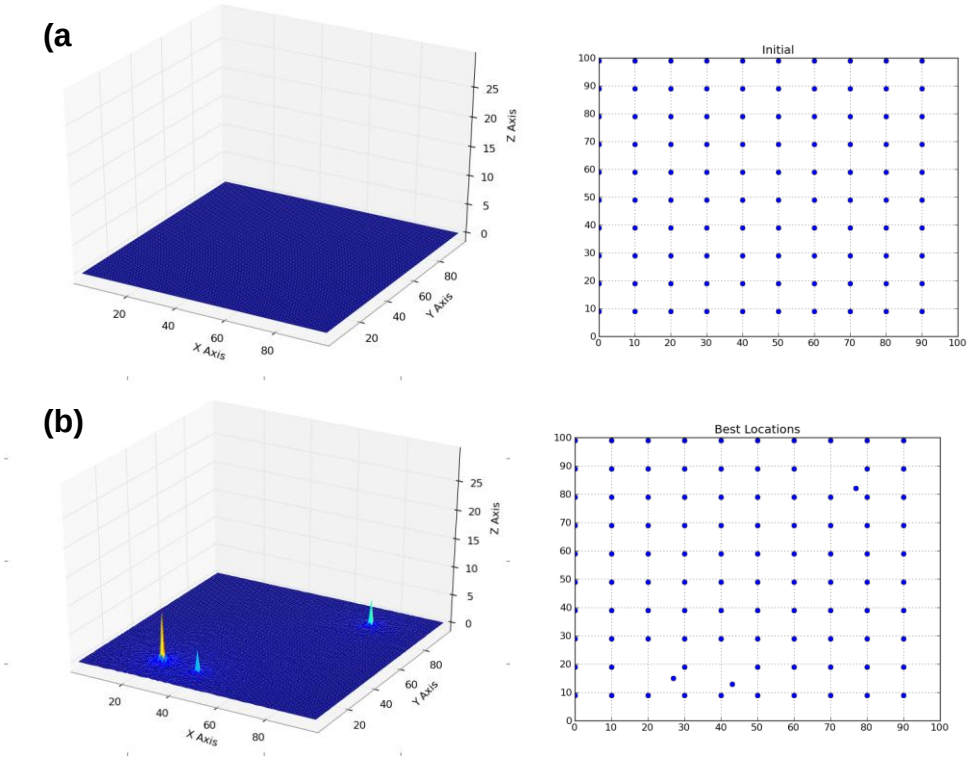


Figure 9: A frame of the simulation as an input vector for the VCS test bench.

In addition to creating input vectors for the test bench, the simulation also models the Scanner's Auto Calibration mode. An example of this is shown in Figure **. Before any spike activity has occurred in Figure **(a) the "best locations" (marked by blue dots on the graph on the right side) for each region are initialized to a default initial value seen in the. After some activity has occurred as in Figure **(b) it can be seen that some blue dots have shifted location in regions where activity has occurred. Figure **(c) shows

how the best locations move after a lot of activity has occurred. It is important to notice that there is only one blue dot in each ten by ten region due to the design of the best locations algorithm discussed in the Design Overview chapter.



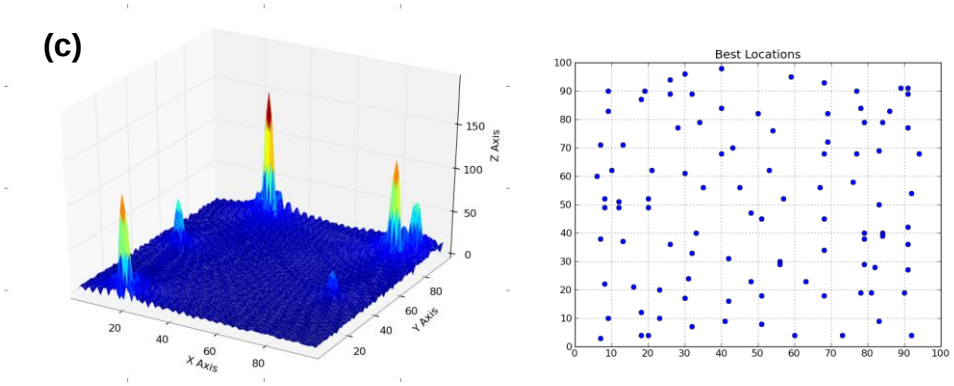


Figure 10: (a) In auto calibration mode before any activity has been recorded. (b) In auto calibration mode after activity has been recorded. (c) Another example of how the best location in each region is saved after activity is recorded.

CIRCUIT LEVEL LOW POWER METHODS

Power Consumption in CMOS

Power consumption by transistors in CMOS technology are given by equation (1). Dynamic power is the power consumed due to transistor switching. $\boldsymbol{V_{DD}}$ is the voltage supply, $\boldsymbol{C_L}$ is the load capacitance due to the physical structure of the MOSFET, $\boldsymbol{f}$ is the operating clock frequency, and $\boldsymbol{\alpha}$ (alpha) is the probability of switching. The value of alpha will vary transistor by transistor due to the logic of the IC.

(1) $$\boldsymbol{P = \alpha . f . C_L . V_{DD}{}^2 \; + \; f . I_{peak} V_{DD} \; + \; V_{DD} . I_{leakage}}$$

Dynamic Power Short-circuit Power Leakage Power

Dynamic power and leakage power can be reported using the Synopsys Power compiler tool. Short-circuit power is much smaller than dynamic power and leakage power, and is omitted from power analysis. Leakage power is also magnitudes smaller than dynamic power, however, leakage current is increasing greatly with transistor scaling, so leakage power is becoming quite significant. Furthermore, during periods of inactivity leakage power is the dominant power component. Since the scanner design presented has long periods of inactivity, it is important to account for and optimize both dynamic and leakage power.

Clock Gating

Clock gating is a method by which the alpha in the dynamic power equation (equation 1) can be reduced for the transistors inside flip flops. The normal configuration in Figure ** (a) have the clock connected directly to the flip flop. When the Enable signal is equal to one the flip flop latches a new input value and when enable is equal to zero the flip flip latches the same value that was held the previous clock cycle, which causes the flip flop to consume power unnecessarily. In Figure **(c) the enable signal and the clock signal become

inputs to an AND gate, and the output of this gate is connected to the flip flop. If the enable signal goes high a much smaller percentage of time than the clock, then the signal to the flip flop, ENCLK will toggle a much smaller percentage than the configuration in (a).

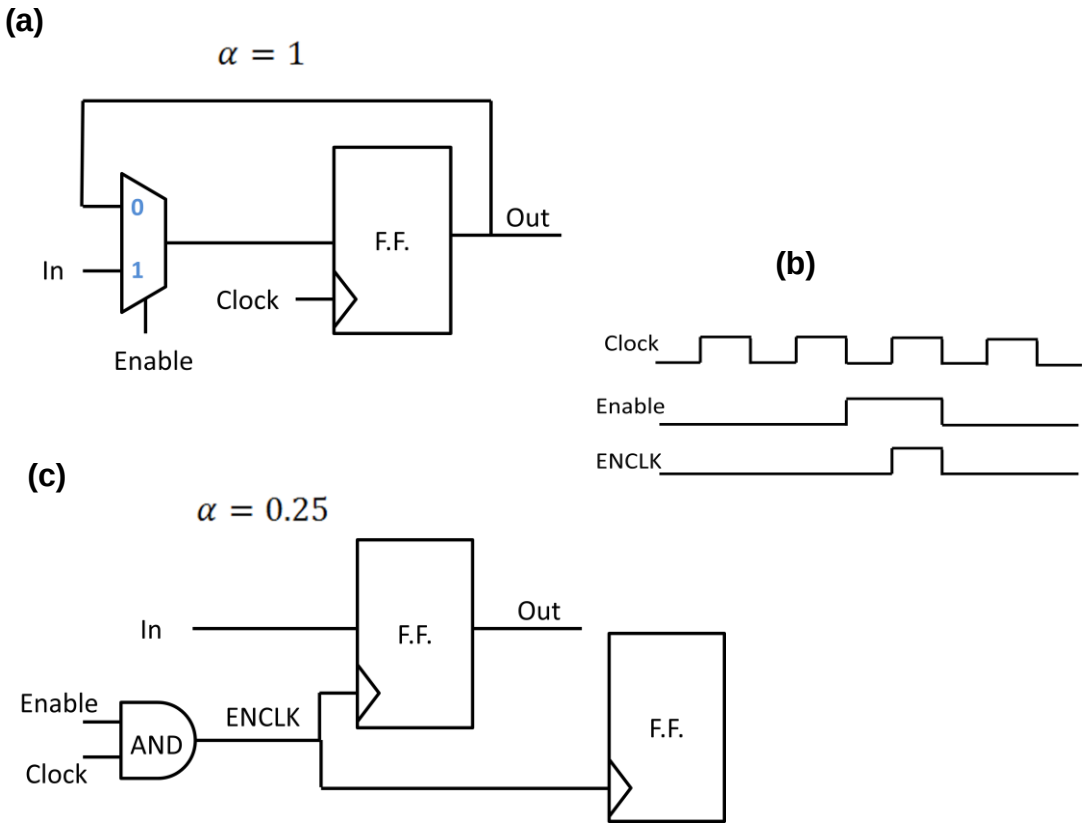
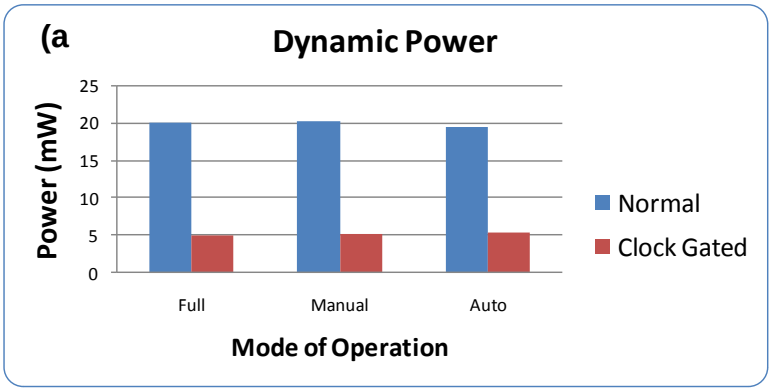


Figure 11: (a) Normal configuration. (b) Example of input waveforms. (c) Clock gating configuration.

In the example in figure **, applying clock gating in this case reduces the alpha for some transistors in the flip flops from 1 to 0.25. (b) shows the ENCLK connected to two flip flops because the same ENCLK signal tends to

be shared among many flip flops, and when this configuration is applied to a whole module or chip, large power savings can be .

Figure *** shows the power simulation results with and without power gating. The dynamic power is reduced by 70% in all three modes of operation. It should be noted that auto calibration mode was simulated during the calibration period, and after the calibration period is over the scanner is expected to consume the same amount of power as operation in manual mode. Figure **(b) shows that leakage power is barely effected one way or the other by clock gating, and it is important to observe that the units on the y-axis in the graph of dynamic power is in milli-watts while the units on the leakage power graph are in micro-watts. Therefore dynamic power is by far the dominant factor in power consumption when the scanner is not in sleep mode.



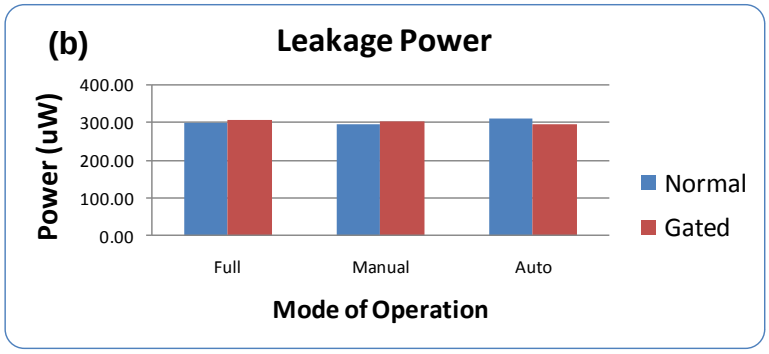


Figure 12: (a) Dynamic power results with and without clock gating (b) Leakage power results with and without clock gating.

Power Gating

Leakage power increases as transistors are scaled smaller and smaller. Equation (2) shows that transistor off (leakage) current can be considered sub-threshold current, which shows an exponentional dependence on $V_{GS} - V_T$. Therefore, if Vt decreases as a requirement of scaling, than the leakage current will increase exponentionally. This presents IC designers with the challenge of having to account for leakage power in newer designs, especially in battery powered applications.

$$(2)\quad I_{OFF} \propto I_{Subth} \propto \frac{W_{eff}}{L_{eff}} K_1 e^{(V_{GS}-V_T)}$$

A popular method of reducing leakage power, is to turn off blocks of logic temporarily when they are not active. The low power scanner is an excellent candidate for power gating since the activity of the block is highly

predictable and the block is designed for a large inactive period or “sleep” period as was discussed in the Design Overview chapter.

Since the goal is to turn off only specific blocks of logic, it is not practical to disconnect parts of a chip from the external power supply directly. Instead, internal power gating using on chip internal switches to disconnect logic blocks from the on-chip power supply VDD. When the logic block is in active mode the sleep transistors are on and Virtual VCC is connected to VDD and Virtual GND is connected to GND and the logic block operates with low Vt transistors. When the block goes to sleep high Vt PMOS or NMOS sleep transistors are turned off, disconnecting the logic block from the power supply. Figure ** shows sleep transistors for both VDD and GND, but only sleep transistors for one part of the supply, VDD or GND, are needed.

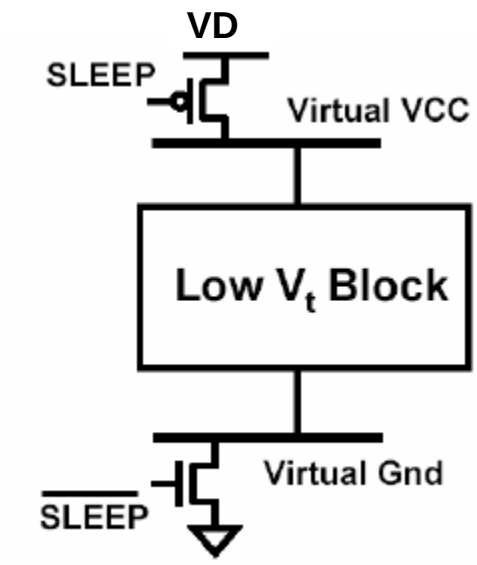


Figure 13: Illustration of power gating principle.

Figure *** shows how power gating is implemented for the Low Power Scanner. The power controller controls the sleep transistor for the Scanner

block. When the scanner module has finished scanning for one period it goes into its sleep state which signals the power controller it is ready to sleep. When the power controller receives the Sleep Ready signal it turns off the sleep transistor using the Power SWT signal. AT the same time the power controller also starts a timer, and when that timer reaches a designated value based on what mode the scanner is operating in (if it is in auto calibration mode it also depends on whether the calibration period is over or not), the sleep transistor is turned back on and power is restored to the Scanner block.

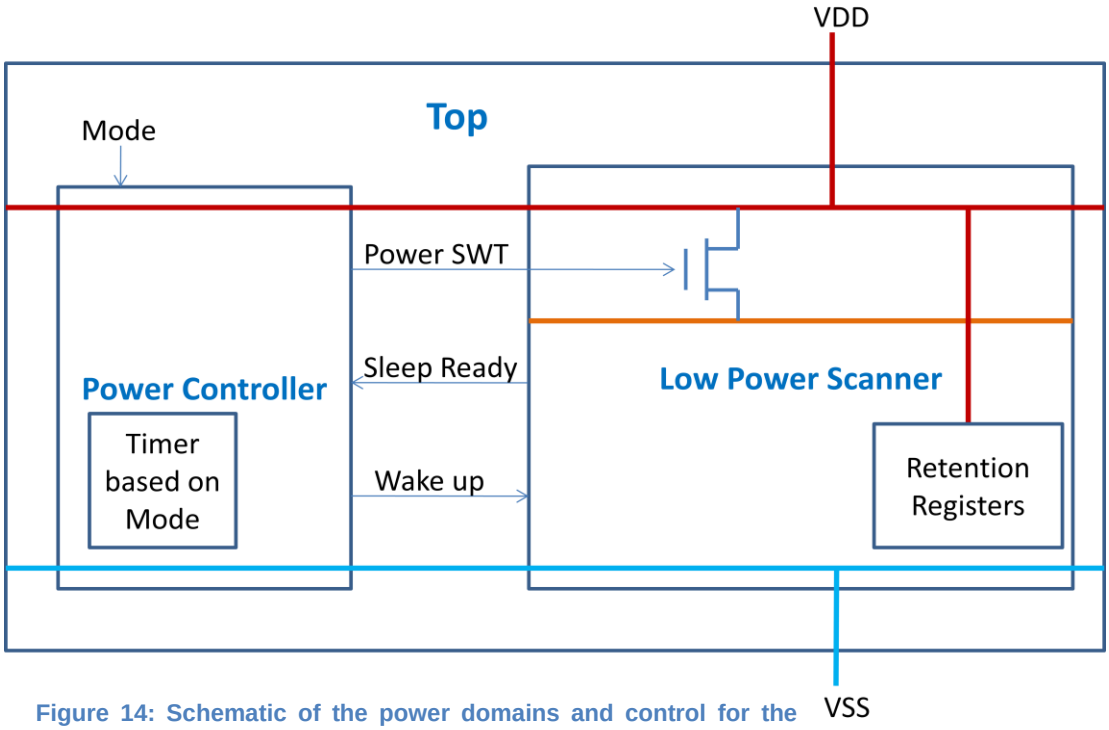


Figure 14: Schematic of the power domains and control for the Low Power Scanner.

The power controller has a delay between the time power is restored and the Wake up signal is sent to the Scanner module. This is in order to

account for the time it takes for the power to be fully restored to the Virtual VCC line. The delay is currently set to 50 clock cycles, but can be easily adjusted after this timing requirement is more accurately estimated during physical layout and parasitic extraction.

The power controller is simulated to consume 386 microwatts of dynamic power and 8 microwatts of leakage power when synthesized using clock gating. The dynamic power is due to the timer and is a little bit more than the leakage of the Scanner module which was typically simulated to be around 300 microwatts. However, without the power controller a timer would be needed to control the scanner reading start and stop time anyway, so the dynamic power of the controller is not considered as significant additional overhead for the Scanner design.

Typically in power gating, there are some flip flops that hold values that should not be lost during sleep mode. These registers should retain their values, and are thus designated as “retention registers”. The retention registers are not powered down during power gating.

The next chapter will present the power simulations for the Scanner in different operating modes and with and without synthesis using clock gating, power gating and the combination of both.

POWER SIMULATION RESULTS

The Low Power Scanner was synthesized using Design Compiler under normal, clock gated, power gated, and both power gated and clock gated conditions. After the design was synthesized into the gate level netlist, the gate level netlist was simulated for power. This power simulation was executed with annotated input, output and internal signal values from the test benches, which made the simulation more accurate to real operating conditions.

(3) $$Energy = (Dynamic\ Power * Active\ time\ in\ mode\ of\ operation) + (Leakage\ Power * Sleep\ time\ in\ mode\ of\ operation)$$

Equation (3) was used to calculate the energy consumed during a 100 microsecond period. A signal 100 microsecond period reflects the overall operation of scanner since it will repeat its behavior every period. The exception to this is rule is that in auto calibration mode the scanner will stop scanning the entire array after the user set calibration period is over and only scan the top 100 “best locations” defined in its algorithm. This switch essentially causes the scanner to start using the same amount of energy as in Manual mode. When power gating is applied leakage power in equation (3) is changed to 1nW, which is an estimate of leakage with power gating.

The results of power simulations are presented in figure **. Auto Calibration mode is defined as operation before the calibration has been reached. Energy consumption in Full Scan mode and Auto Calibration mode are expected to be nearly the same since in both modes the whole array is being scanned. Clock gating greatly reduced the energy consumed since it reduced the dynamic power so greatly as was previously discussed in the clock gating section.

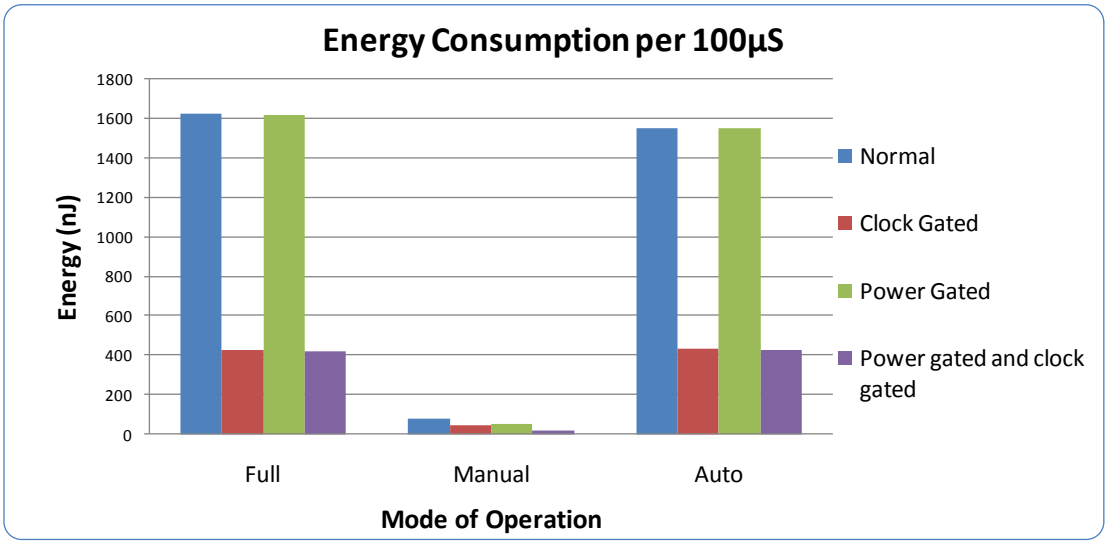


Figure 15: A graph showing Energy consumption of the low power scanner in different modes of operation over a 100uS period with different configurations applied.

At first look, power gating appears to have little effect on energy consumption. This is due to the fact that dynamic power is so much larger than leakage power. However, when the scanner is operating in manual mode, or in Auto Calibration mode after the calibration period has ended, the amount of time the Scanner is active is greatly reduced. Therefore, if manual

mode is looked at more closely, as in Figure **, It can be seen that power gating is more effective, since the Scanner is active for a shorter time and sleeping for a longer time, causing leakage power to become a more significant portion of energy consumption.

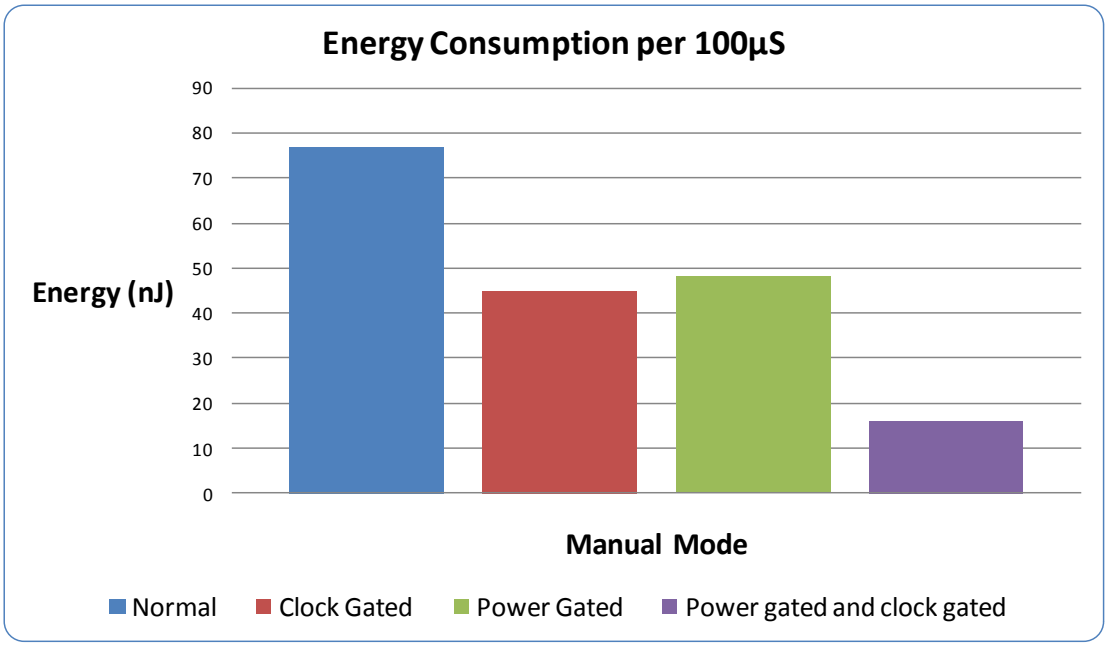


Figure 16: A graph of energy consumption of the low power scanner in Manual mode over a 100 uS period with different configurations applied.

Although figure 16 shows the effectiveness of power gating, it is not the complete picture. The design of the complete system includes ten ADCs, which can all be power gated as well. Equation 4 adjusts our original energy calculation equation to take this into account. ADC leakage power is estimated to be 50uW, which is a very rough, but conservative estimate. ADC leakage power is changed to 1nW in power gated mode, which is taken

from datasheets of ADCs in the performance category required for this application.

(4) $$Energy = (Dynamic\ Power * Active\ time\ in\ mode\ of\ operation) + (Leakage\ Power * Sleep\ time\ in\ mode\ of\ operation) + (ADC\ Leakage\ Power * Sleep\ time\ in\ mode\ of\ operation)$$

Figure ** Shows Energy consumption when the Leakage of power from the ten ADCs is taken into account. Now it can be seen visually in Figure ** that power gating has a larger effect in Full Scan mode and Auto Calibration mode (before calibration period is over).

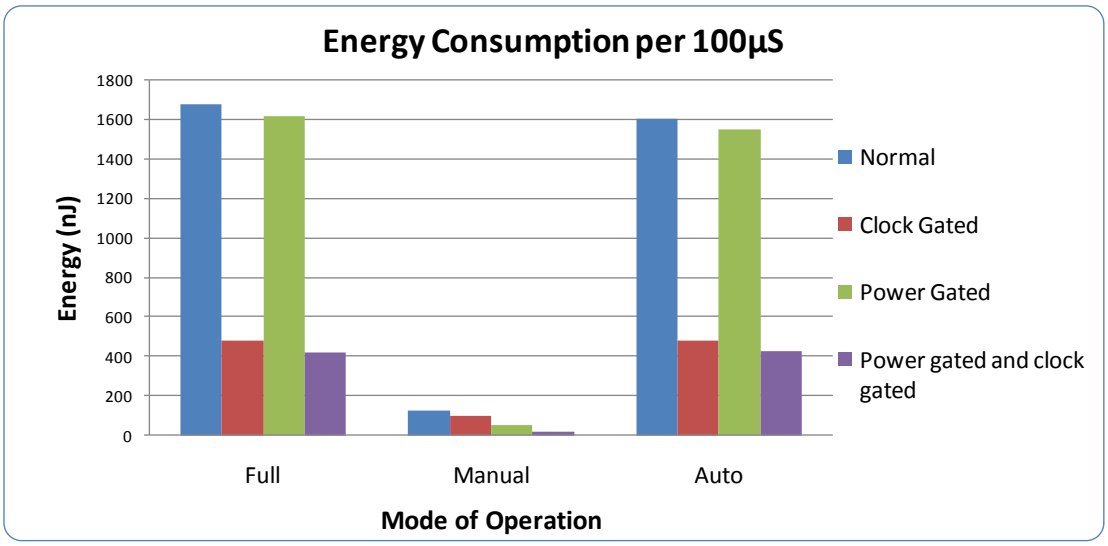


Figure 17: A graph showing Energy consumption including leakage of ADCs of the low power scanner in different modes of operation over a 100uS period with different configurations applied.

As before, if Manual Mode is looked at more closely in figure ** it can be seen that power gating has a large effect on energy consumption; in this

case it is even more effective than clock gating. When clock gating and power gating are applied to the Low Power Scanner design the new energy consumption is extremely low compared to the original design. These results overwhelmingly support the application of clock gating and power gating for the Low Power Scanner module to reduce truly make the Scanner low power.

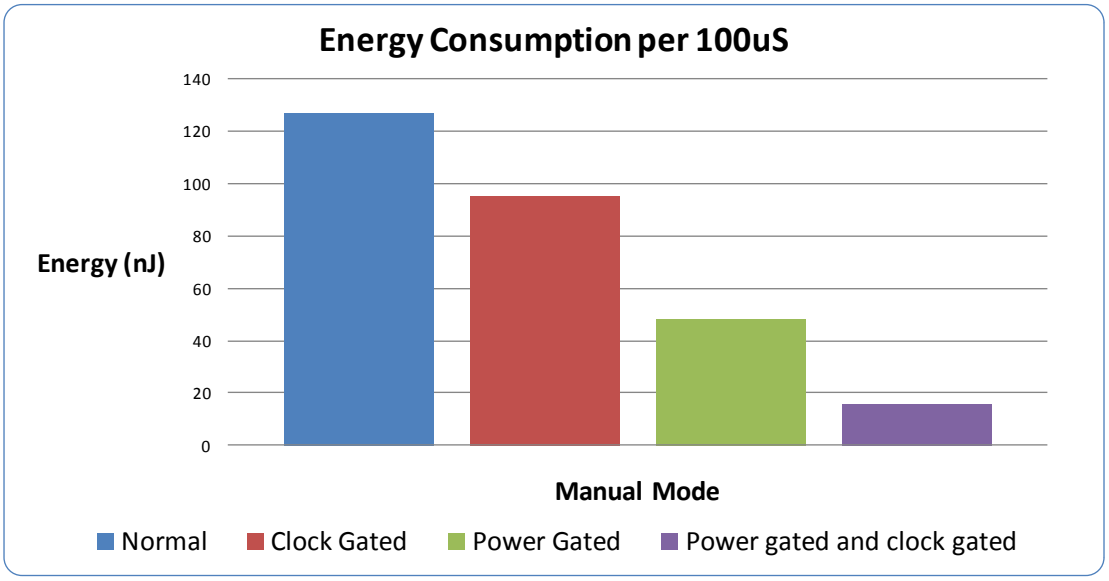


Figure 18: A graph of energy consumption including leakage of ADCs of the low power scanner in Manual mode over a 100 uS period with different configurations applied.

CONCLUSION

It is clear that at the circuit level, clock gating and power gating of the Scanner module and as well as power gating the ADCs can result in large power savings. Power savings due to power gating is due to the burst strategy for the scanner module, essentially creating a long period of inactivity when the Scanner can be put to sleep. Burst mode in conjunction with power gating is expected to be even more important as transistors are further scaled and leakage power consumes a larger and larger fraction of total energy consumed.

A low power design for scanning an electrode array has been presented, as well as verification methods and other important analysis. At the upper level of design, multiple operation modes offers the user flexibility for different research situations, with a focus on low power operation. The Scanner design was verified with multiple test benches and with a golden model created in python to verify auto calibration mode. This golden model may be useful for other designers and researchers to create input test vectors and could be adapted to test other features as well. Future work could include frequency scaling depending on operation mode, although such fine grain frequency scaling may not be practical other hardware could be included as well.

REFERENCES