

“Power Optimization of SoC using GALS”

Prerna Pardhy, Vivek Khetade

Dept of Electronics Engineering
Shri Ramdeobaba Kamla Nehru Engineering College
Rashtrasant Tukdoji Maharaj, Nagpur University
Nagpur, India
prerna.pardhy@gmail.com, khetadev@rk nec.edu

Dr. S.S. Limaye

Department of Electronics Engineering
Jhulelal Institute of Technology
Rashtrasant Tukdoji Maharaj Nagpur University
Nagpur, India
shyam.limaye@hotmail.com

Abstract— The power dissipation of modern processors has been rapidly increasing along with increasing transistor count and clock frequencies. As clock frequency increases and feature size decreases, clock distribution and wire delays present a growing challenge to the designers of singly-clocked, globally synchronous systems. Power consumption is more in globally synchronous system due to the large global clock distribution network. To mitigate this problem we are suggesting Globally Asynchronous Locally Synchronous system for power reduction. GALS removes the need for global clock net and also provides efficient means for managing complexity and reuse of large architecture. Apart from power reduction, GALS has benefits such as multiple clock frequencies for IP blocks and dynamic frequency/voltage scaling, it supports concurrent data transfer, clock tree removal, IP core reusability.

We are proposing a method that evaluates benefits of GALS together with clock gating to minimize power consumption. The result shows that power reduction upto 73% can be achieved with negligible area and performance overhead.

I. INTRODUCTION

In earlier generations of IC design technologies, the main parameters of concern were timing and area. EDA tools were designed to maximize the speed while minimizing area. Power consumption was a lesser concern. CMOS was considered a low-power technology, with fairly low power consumption at the relatively low clock frequencies used at the time, and with negligible leakage current.

In recent years, however, device densities and clock frequencies have increased dramatically in CMOS devices, thereby increasing the power consumption dramatically. At the same time, supply voltages and transistor threshold voltages have been lowered causing leakage current to become a significant problem. As a result, power consumption levels have reached their acceptable limits, and power has become as important as timing or area.

High power consumption can result in excessively high temperatures during operation. This means that expensive

ceramic chip packaging must be used instead of plastic, and complex and expensive heat sinks and cooling systems are often required for product operation. Laptop computers and hand-held electronic devices can become uncomfortably hot to the touch. Higher operating temperatures also reduce reliability because of electromigration and other heat-related failure mechanisms. High power consumption also reduces battery life in portable devices such as laptop computers, cell phones, and personal electronics. As more features are added to a product, power consumption increases and the battery life is reduced, requiring a larger, heavier battery or shorter life between charges. Battery technology has lagged behind the increased demands for power. Another aspect of power consumption is the sheer cost of electrical energy used to power millions of computers, servers, and other electronic devices used on a large scale, both to run the devices themselves and to cool the machines and buildings in which they are used. Even a small reduction in power consumption of a microprocessor or other device used on a large scale can result in large aggregate cost savings to users and can provide significant benefits to the environment as well.[1][2]

IRTS Table:

Power	Technology		
	90nm	65nm	45nm
Dynamic power/ cm ²	1X	1.4X	2X
Static Power/ cm ²	1X	2.5X	6.5X
Total power consumption/ cm ²	1X	2X	4X

As per the above IRTS table, power density increases as the technology decreases. So for all application, reducing the power consumption by SoC is essential in order to add the performance and features.

Also, with recent rise in the System-on-a-Chip (SoC) design approach, problems arising due to issues such as clock skew, and power consumption caused by global clock distribution have become more significant. The distributions of low-skew, high frequency clock signals are responsible for a considerable share of power consumption and die area. Figure1 shows the power breakdown in high performance CPU.As can be seen the clock is the largest power

consuming component. This includes the clock generator, the clock drivers, the clock distribution tree, the latches, and the clock loading due to all the clocked elements [3].

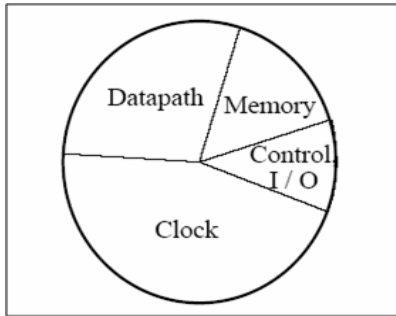


Figure 1 : Power Breakdown in High Performance CPU.

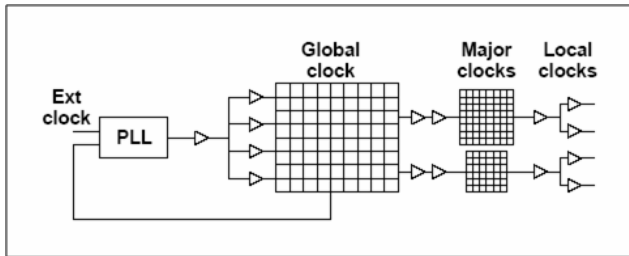


Figure2: Hierarchical Clock Distribution Network

In most systems, a phase lock loop (PLL) generates a high frequency clock signal from a slower external clock. Fig 2, shows an example of a clock distribution network. Several major clocks are derived from the global clock grid, and local clocks are in turn derived from the major clocks [3].

Power consumption in CMOS circuits can be static or dynamic. Leakage current or other current drawn continuously from the power supply causes static power dissipation. The dynamic power (*switching power*) dissipated by a chip is $C \cdot V^2 \cdot f$, where C is the capacitance being switched per clock cycle, V is voltage, and f is the switching frequency. Dynamic dissipation occurs during output switching because of short-circuit current, and charging and discharging of load capacitance. Dynamic power consumption depends on the clock frequency and switching activity. For existing CMOS technology, dynamic power is the dominant source of power consumption, although this might change for future high-scale integration.

During our survey, we have gone through some power some power reduction techniques, such as Dynamic Voltage Scaling, Dynamic Frequency Scaling, Clock Gating, GALS Unfortunately, there is no single power reduction technique that is able to meet all the system requirements for energy minimization. The trick is to effectively combine different techniques to intelligently develop energy-efficient semiconductor designs.

II. POWER REDUCTION TECHNIQUES

1. Dynamic Voltage Scaling

The most basic way to reduce power is to reduce supply voltage. Power usage is proportional to the square of the supply voltage. As shown in figure 3 a 50 percent reduction in supply voltage results in 50 percent reduction in current and 75 percent reduction in power.

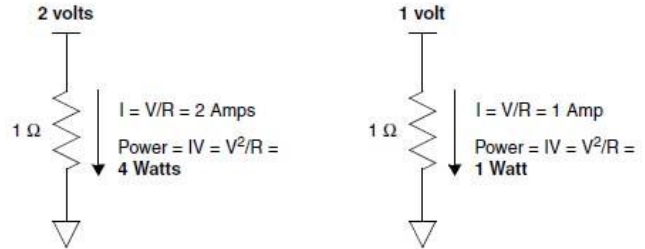


Figure 3 : Power Versus Supply Voltage example

As illustrated in Figure 4. The figure shows the power, feature size, voltage, frequency, and relative die size of some recent CPUs. Starting from a 5V part, there was an initial decrease in power when moving to a smaller technology at 3.3V. However, the power came back to original levels when the frequency was increased. The Pentium® Pro with its aggressive microarchitecture, saw an additional increase in power, even at 3.3V and a smaller technology.

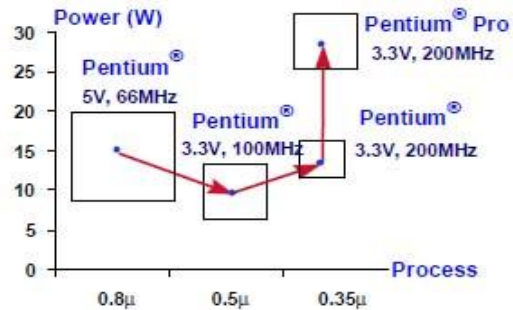


Figure 4: Power and Vcc of some recent CPU

2. Dynamic Voltage and Frequency :

The principle of multivoltage operation can be extended to allow the voltage to be changed during operation of the chip to match the current workload. For example, a math processor chip in a laptop computer might operate at a lower voltage and lower clock frequency during simple spreadsheet computations, thereby saving power; and then at a higher voltage and higher clock frequency during 3-D image rendering when the highest performance is needed. The changing of supply voltage and operating frequency during operation to meet workload requirements is called dynamic voltage and frequency scaling. The chip and voltage supply can be designed to use a number of established levels, or even a continuous range. Dynamic voltage scaling requires a multilevel power supply and a logic block to determine the best voltage level to use for a

given task. Design, implementation, verification, and testing of the device can be especially challenging because of the ranges and combinations of voltage levels and operating frequencies that must be analyzed and accommodated.[2]

3. Clock Gating

The clock is the largest contributor to the CPU power. Clock gating is a dynamic power reduction method in which the clock signals are stopped for selected register banks during times when the stored logic values are not changing. One possible implementation of clock gating is shown in figure. Clock gating is particularly useful for registers that need to maintain the same logic values over many clock cycles. Shutting off the clocks eliminates unnecessary switching activity that would otherwise occur to reload the registers on each clock cycle. The main challenges of clock gating are finding the best places to use it and creating the logic to shut off and turn on the clock at the proper times. Clock gating is a well-established power-saving technique that has been used for years. Synthesis tools such as Power Compiler can detect low-throughput datapaths where clock gating can be used with the greatest benefit, and can automatically insert clock-gating cells in the clock paths at the appropriate locations. Clock gating is relatively simple to implement because it only requires a change in the netlist. No additional power supplies or power infrastructure changes are required.

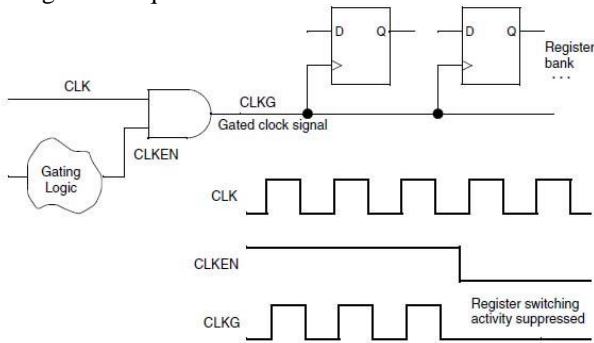


Figure 5: clock gating Example

4. Multiple Clock Domain

Multiple Clock Domain (MCD) processor in which the chip is divided into several clock domains, within which independent voltage and frequency scaling can be performed.[4][5]An average energy-delay product improvement of 20% with MCD can be achieved compared to a modest 3% savings from frequency and voltage scaling in a single clock system. In an MCD microprocessor each functional block operates with a separately generated clock and synchronizing circuits ensure reliable inter domain communication. Thus, fully synchronous design practices are used in the design of each domain. An MCD microprocessor offers a number of potential advantages over a singly clocked design:

- The global clock distribution network is greatly simplified, requiring only the distribution of the externally generated clock
- The designers of each domain are no longer constrained by the critical paths in the other domains.
- Using separate voltage inputs, external voltage regulators and controllable clock frequency circuits in each domain allows finer grained dynamic voltage and frequency scaling.

A multiple clock domain (MCD) microarchitecture, uses a globally-asynchronous, locally-synchronous (GALS) clocking style.[7]

III. BLOCK DIAGRAM

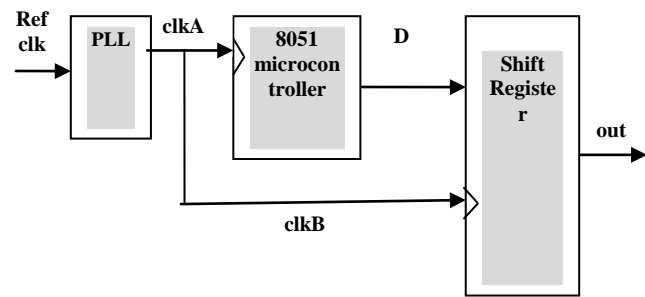


Figure6: Block Diagram of a system :

Figure6 shows the block diagram of a system. Here we are using PLL as a clock generator. First we are using global clock network. The single clock is used for microcontroller and shift register. We are operating both the blocks on 120 MHz frequency. Following fig shows the simulation waveform.

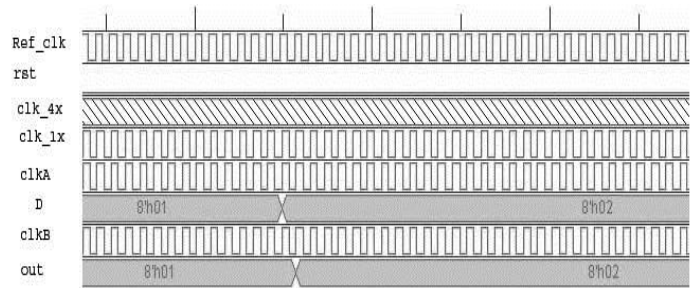


Figure: Simulation waveform

Now our aim is to use GALS architecture to reduce power consumption of the system. So below is the block diagram of a proposed architecture where we are using multiple clock domain.

Figure7 shows the block diagram of a proposed system. Here we are using PLL as a clock generator. It generates multiple frequencies. The clkA is driving a microcontroller and clkB is driving Shift Register. Since both the blocks are operating at two different clock frequencies, a synchronizer is used. The synchronizer is a two stage synchronizer. The

microcontroller is operated at 120MHz and shift register is operated at 30 MHz.

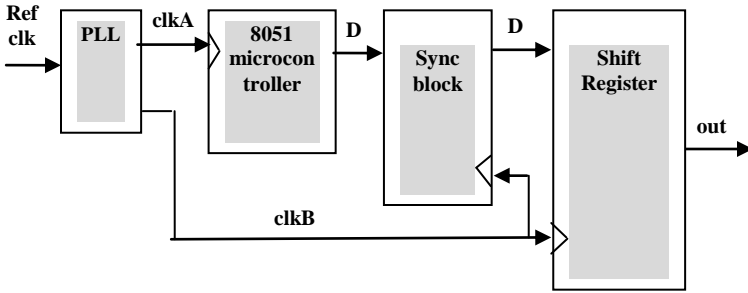


Figure 7: Block Diagram of Proposed System :

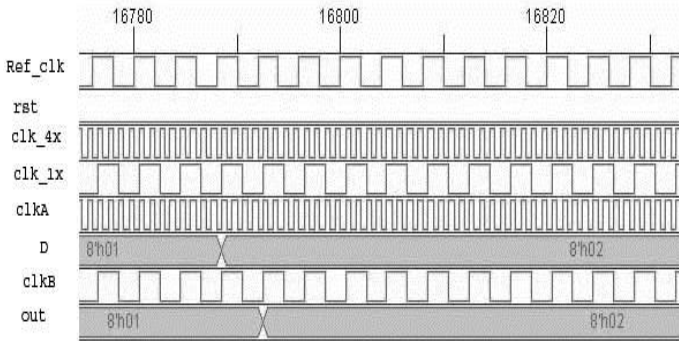


Figure: Simulation Waveform of Proposed system

IV. RESULTS

Table shows the comparative study of power consumption of **shift register block** for both the systems.

Frequency (MHz)	Cell Internal Power (μ W)	Net Switching Power (μ W)	Total Dynamic Power (μ W)
120	19.32	87.96	19.41
30	5.15	23.45	5.17

So the power reduction of approximately 70% is obtained in shift register block if we used GALS system. Power consumption of Synchronizer block is 1.69μ W, which is negligible.

1. PLL Block:

PLL block is used as a clock generator. The PLL is clock multiplier circuit that can generate a stable, high-speed clock from a slower clock signal.

The PFD output applied to the Charge pump (CP), and consecutively to Low pass filter (LPF), which produces a control voltage for setting the VCO frequency. If the PFD produces an UP signal, then the VCO frequency should increase. A DN signal decreases the VCO frequency. If the CP receives an UP signal, current is driven into the LPF. Conversely, if it receives a DN signal, current is drawn from the LPF. The VCO output feeds up to two post-scale counters (Output dividers). These post-scale counters allow

a number of harmonically related frequencies to be produced within the PLL. The PLL is designed to generate 30 MHz (CLK_1X), 60 MHz (CLK_2X) and 120 MHz (CLK_4X) clock signals from a lower frequency 30 MHz clock reference signal (REF_CLK). Typical input clock source is crystal oscillator. Analog power (AVDD) is provided.

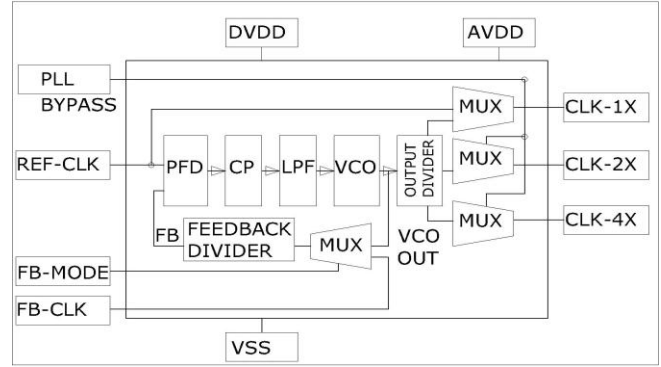


Figure 8: Block Diagram of PLL

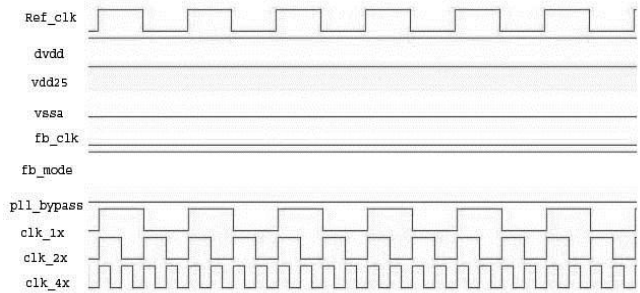


Figure : simulation Waveform of PLL

2. Synchronization Circuit

The problem generally faced in MCD are Setup and hold time violation and Metastability

The metastability occurrences can be predicted by using the mean time between failures (MTBF) formula:

$$MTBF = \frac{e^{C2 * t_{MET}}}{C1 * f_{clk} * f_{data}}$$

Where $C1$ and $C2$ are constants that depends on the technology used to build the flip-flop; t_{MET} is the duration of the metastable output; and f_{clk} and f_{data} are the frequencies of the synchronous clock and the asynchronous input, respectively.

To reduce the effects of metastability, designers most commonly use a multiple stage synchronizer in which two or more flip-flops are cascaded to form a synchronizing circuit shown in Figure 9 .

The asynchronous output (src_data_out) from the source module working on src_clk is fed to the first synchronizer flip-flop. The signal dest_data1 (output of first synchronizing flip-flop) goes metastable but resolves to a stable state before it is sampled by the second flip-flop in the synchronizer circuit. Thus the signal dest_data2 (output of the second synchronizer flipflop) is synchronized to the dest_clk used by the destination module.

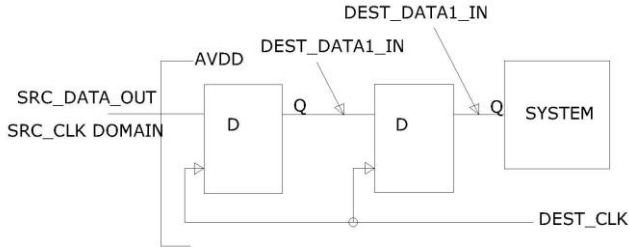
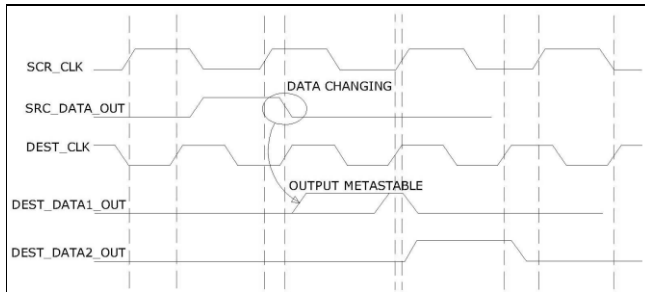


Figure 9 :Two Stage Synchronization Circuit



Timing Diagram of Two Stage Synchronization Circuit

V. GALS ARCHITECTURE

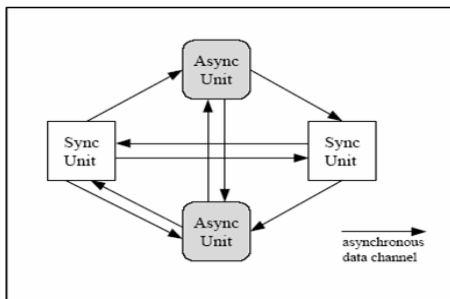


Figure 10: GALS Architecture

With the recent rise in the demand for System-on-a-Chip (SoC) designs, global clock distribution and power dissipation due to clock distribution are inevitable. In order to reduce/eliminate the effects of the global clock in synchronous designs and large area overhead in asynchronous designs, an alternative approach would be to utilize GALS design techniques. Not only do GALS designs eliminate the issue of using a global clock, they also have smaller area overhead when compared to purely asynchronous designs. The Globally Asynchronous Locally Synchronous (GALS) clocking style separates processing

blocks such that each part is clocked by an independent clock domain[6].In GALS design, asynchronous interfaces are added to locally synchronous (LS) modules, where-in the interiors of a specific module are encapsulated by the interfaces; and it is possible for each module to use its own clock, power supply voltage, or even to collapse the power supply for the synchronous part without affecting any other parts of the chip. In GALS designs, each distinct module can thus generate its own clock signal, and communication between modules is asynchronous. Therefore, the problem of power consumption due to global clocking, and clock skew are avoided; and these issues are limited to the local synchronous modules.

VI. CONCLUSION :

As compared to purely synchronous circuit upto 70 % power reduction is possible by using GLAS.And further power reduction is possible using clock gating or voltage scaling to individual synchronous block. In a GALS we are having a liberty of using modules of different frequencies, we can achieve power optimization. By using synchronizer, we are able to send data from fast system to slower system.

VII. REFERENCES

- [1] Stephen H. Gunther, Desktop Platforms Group, Intel Corp.Frank Binns, Desktop Platforms Group, Intel Corp. “Managing the Impact of Increasing Microprocessor Power Consumption”
- [2] Synopsys Low Power Flow User Guide
- [3] Akarsh Reddy Ravi B.E. Bangalore University, 2001 “Globally Asynchronous Locally Synchronous Wrapper Configuration for point to point and multipoint data communication.”
- [4] “Integrated circuits and system design Power and timing modelling” by Vassillis Paliouras,JohanVounckx,Diederik Verkest
- [5] Paper by MAssoud Pedram “Design Technologies for Low Power VLSI”.
- [6] Frank Emmett and Mark Biegel, Automotive Integrated Electronics Corporation “Power Reduction Through RTL Clock Gating”
- [7] Greg Semeraro, Grigorios Makklis,Rajeev alasubramonian,David H. Albonesi, Sandhya Dworkadas and Michael L. Scott , Departments of Computer Science and of Electrical and Computer Engineering University of Rochester “Energy-Efficient Processor Design Using Multiple Clock Domains with Dynamic Voltage and Frequency Scaling”
- [8] Sourdip Sarkar”Multiple Clock Domain Synchronization for Network On Chips”.
- [9] Grigorios Magklis, Michael L. Scott ,Greg Semeraro, David H. Albonesi, and Steven Dropsho Dept of Computer Sciencs, Department of Electrical and computer engg University of Rochester, Rochester, NY 14627 “Profile-based Dynamic Voltage and Frequency Scaling for a Multiple Clock Domain Microprocessor”
- [10] Vivek Tiwari, Deo SinghSuresh Rajgopal, Gaurav Mehta,Rakesh Patel,Franklin Baez Intel Corporation “Reducing power in high performance Microprocessor”