LIMITED DISTRIBUTION NOTICE
(NOT TO BE REMOVED FROM DOCUMENT):

This report has been submitted for publication outside of IBM and will probably be copyrighted if accepted for publication. It has been issued as a Research Report for early dissemination of its contents. In view of the transfer of copyright to the outside publisher, its distribution outside of IBM prior to publication should be limited to peer communication and specific requests. After outside publication, requests should be filled only by reprints or legally obtained copies of the article (e.g., payment of royalties).

Real-Time VLSI Chip Power/Performance Optimization

Christos J. Georgiou, Thor A. Larsen[1], and Eugen Schenfeld[2]

IBM T.J. Watson Research Center
P.O. Box 704
Yorktown Heights, NY 10598
(914) 784-7161 / geor@watson.ibm.com

Abstract

The performance of some CMOS chips (e.g., VLSI processors) may be increased by running the internal circuits at higher clock rates, but use of a higher clock rate is often limited by the heat-dissipation ability of the chip's package. In this paper, a method is described for estimating the total heat accumulated for dissipation at any given time. For the periods that the chip is idle, the clock rate is decreased to reduce heat generation. The heat (or energy) saved while the chip is idling is available for use later to increase the clock rate above the normal rate for a predetermined time such that the total heat generated does not exceed the heat-dissipation capacity of the package. This approach addresses the limitation of semiconductor chip performance which occurs as a consequence of a limited heat removal capability. More particularly, it relates to the maximum clocking rates which are imposed on chips fabricated in some of the popular technologies as a direct consequence of limitations in the rate at which heat can be removed from a chip. Analogous to the CMOS technology, this concept can be extended to other logic technologies (e.g., ECL, GaAs MESFET, etc.) by regulating circuit currents instead of clock rates.

---

[1]Currently with IBM Microelectronics Division, East Fishkill, New York

[2]Currently with NEC Research, Princeton, New Jersey. This work was done while on a post-doctoral assignment at T.J. Watson Research Center

1.0 Introduction

The performance of an increasing number of VLSI chips and multi-chip subsystems is bounded by the heat removal ability of their packages. The emerging large market for high performance, portable computer systems places new demands for very low power consumption while retaining maximum system performance. Limiting power dissipation to extend battery life or limit the maximum temperature of a chip or subsystem is handled by controlling the maximum clock frequency, in the case of CMOS designs, or by controlling the current, in the case of bipolar or GaAs MESFET VLSI circuits.

The focus of this paper is to describe techniques which could be used to optimize chip or system performance while, at the same time, remaining within the system constraints of maximum ***average*** chip or sub-system power dissipation and current drain.

One solution currently being pursued in this area is to cool the chip package using an external cooling device. Such a scheme has been used by Everex Systems, Inc. for the Intel 80486 microprocessor CPU chip that, although specified to run at a maximum of 25 MHz clock speed, achieves a clock rate of 50 MHz. The Everex scheme employs a product known as Icecap produced by Velox Computer Technology Inc., that cools the CPU chip to 0° C [1].

In the approach described in [2], a microprocessor controls the clock frequency based on the present rate of required microprocessor activity. However, the performance of the microprocessor is not improved. Variation of the clock frequency is used only to save power (i.e., battery life). Another approach is described in [3], in which a chip set intended for the laptop computer market includes power management features which are built into the microprocessor. In order to conserve battery life, every piece of logic can be powered down. Again, the focus is on battery life and there is no apparent performance gain as a consequence of any of the powering down.

Another approach is to statically allocate two clock frequencies to the tasks inside the chip [4]. Functions such as direct memory access run at the lower clock frequency while

massive computations run at the higher clock frequency. While the apparent focus here is on heat dissipation, the clocking rate of each circuit does not vary over time, so the maximum clocking rate for each circuit is still determined by the maximum heat dissipation rate from that circuit.

The objectives of the approach described in this paper are the following :

- To optimize the power-performance of CMOS VLSI chips by controlling, in real-time, the clock frequency distributed to the chip's functional units, so that the average chip power dissipation, or maximum chip current drain, remains within the prescribed limits.
- To utilize functional activity monitors as a means of identifying when to adjust the clock frequencies to the chip's functional units.
- To utilize a counter (or counters) as means of quantifying, in real time, the amount of heat (energy) being used on the chip and determining whether to speed up or slow down, and for how long, clocks being applied to the various functional islands.

Similar concepts can be utilized for GaAs MESFET or bipolar circuits by controlling the magnitude of the circuit operating current [5]. In the case of bipolar chips, inactive circuits can be turned on/off via current control, and thereby offer alternate performance logic paths. Thus, when there is an adequate energy reservoir, a high-performance, high-power logic path can be selected, instead of an equivalent lower-power path.

As mentioned above, VLSI chips generate heat that is proportional to clock rates in most synchronously operated circuits based on CMOS technologies (and clocked FET logic). According to the method described in this paper, heat production is controlled through changes in clock rate (i.e., by slowing down the clock rate when a circuit is idling). This makes it possible to speed up the clock rate when performing either special critical work or useful work in general, such that the total heat generated over time, indicated by the average power or average current consumed, remains within a pre-defined limit specified by thermal or battery life limits.

The clock rate is changed according to the particular activity that a circuit performs at any given time. When a circuit is switched to a lower clock rate, there are heat generation savings. Given a certain heat dissipation capacity (based on the characteristics of the chip and package), it is possible to switch the clock to a higher rate, if enough heat savings have been accumulated. While operating at this higher clock rate, the accumulated heat savings are decreased to account for the extra heat generated.

In summary, a circuit is either accumulating heat savings or using them (or neither, depending upon implementation) based upon the clock speed, which in turn depends upon the detected circuit status. Because of system delays, circuits demonstrate periods of inactivity, which would translate into heat savings. The objective is to run the circuit clocks faster when heat has been saved and, thus, realize an improved overall system performance, while maintaining the same average chip power dissipation or average current drain. This real-time dynamic scheme for chip or subsystem power management is more flexible than any of the prior art schemes.

The rest of the paper is organized as follows: In Section 2, we present a possible implementation of the proposed variable chip-clocking mechanism, in Section 3, we examine circuit implementation issues for technologies other than CMOS, and, finally, we offer a summary and conclusions.

2.0 Description of approach

Referring now to Figs. 1 and 2, a heat savings accumulator, which may be implemented simply as an up-down counter circuit, accumulates the gain and loss in terms of heat production units, while operating at a certain clock rate. The clock for the heat savings accumulator may be provided by an independent timer or more preferably from the same system clock generator that provides the clocking frequencies for the chip being controlled by the variable clocking mechanism. An upper bound is imposed on the maximum heat

savings which may be accumulated based on the given thermal characteristics of the particular chip package (i.e., even if the circuit were idle for one hour it does not mean it can be operated at twice the normal clock rate for the next hour) .

The activity monitor detects when each circuit, macro, or whole chip can be operated at a lower clock rate (e.g., when it is idling or waiting for external events to occur). The activity monitor may monitor the chip as a whole, as shown in Fig. 1, or may sense the activity of the various macros in the chip and change their corresponding clock speed individually. The particular activity that is monitored depends on the circuit design. For a CPU chip, this monitored activity can be the specific instructions executed at any given time. An instruction decoder can be used to detect instructions that take a long time to complete by an external unit (e.g., a floating point numerical coprocessor), during which the main CPU chip is idle waiting for the instruction to be finished by the external unit. For the time these instructions are being executed by an external unit, the main CPU chip can be switched to a minimal clock rate and the heat savings accumulated without negative impact on performance. Another type of activity monitoring might be to observe bus activity. The activity monitor might simply detect the idle periods on a bus (i.e., a DMA transfer, forcing the CPU into a "hold" state, or a "not ready" signal).

When it becomes beneficial to operate the chip, or a macro, at a higher clock rate (e.g., useful work is being done) and there are enough "heat savings in the bank" (i.e., the value of the up-down heat counter is high enough), a synchronous shift to a higher clock rate can take place in order to improve the chip's performance. The variable chip-clocking mechanism for this chip shown in Fig. 2 (a CPU) is implemented as follows:

The activity monitor is simply a logical "NOR" gate having as inputs the "wait" and "hold" lines of the CPU microprocessor bus. If either one of these lines is active, it means that the CPU is not doing any work (and should operate at a lower clock rate to save heat generation). Otherwise, the CPU is active and a high clock rate is used, if there are accumulated heat savings. If there are no accumulated heat savings (or when the

accumulated heat savings have been used up), a normal clock rate is used for the CPU when it is active.

The heat savings accumulator keeps track of heat savings and is implemented with a 20-bit up/down binary counter. The size of the counter means that it is possible to accumulate up to about one million (2^{20}) "saved clock cycles." The counter counts up or down according to the logic states shown in the right two columns of the Table 1, which in turn is determined by the two left-side input columns. This relationship is implemented by an Inverter gate and an "AND" gate. The counter does not overflow (meaning that it does not wrap-around to zero) or underflow (meaning that it does not go from 0 to 1 million). As shown in the table, the counter counts down when the clock rate is high and counts up when the clock rate is low. It should be apparent that with suitable logic modifications the sense (i.e., direction) of the counting could be reversed. Also note that the size of the counter is not fixed but, instead consistent with the provided clock rate, so that the total elapsed time of counting is equal to or less than the thermal time constant of the chip and package. As noted earlier, there is only a finite time that can be used to either accumulate or spend heat savings. The accumulated heat savings for a circuit would be the same regardless as to whether it is idle for a few milliseconds or for an hour, because the chip temperature would reach its rated limit in only a few milliseconds of over-activity.

The heat comparator, implemented by a 20-input "OR" gate, determines whether there are sufficient heat savings to be used. If there are any heat savings (i.e., the counter is different than zero), the "Savings" line will be high (a logic "1"). This signal is used by the clock rate selector, along with the "Active" line from the activity monitor, to determine the clock rate to the CPU chip, in accordance with the values in Table 1, as explained below.

The variable clock generator selects among three possible clock rates for the CPU chip. The clock rates are conveniently taken from the system clock generator: a) f_L is the lowest clock rate and is used to clock the CPU when there is no activity; b) f_N is the clock rate at which the CPU operates normally and when there are no heat savings (the counter is

zero); and c) f_H is the high clock rate that is used when the CPU is doing work and there are heat savings (the heat counter is greater than zero). f_c is the clock rate used to clock the 20-bit counter and it may or may not be the same as one of the other three clock rates.

The improvement in overall chip throughput depends on the characteristics of the particular application. For example, consider a CISC type processor made of one CPU VLSI chip and a second external chip working as a math coprocessor (e.g., the Intel 80386 and 80387). In a particular program, the main CPU may have to wait for the math coprocessor to finish a relatively long floating point operation before it can continue. During this period, heat generation can be saved by lowering the main CPU clock rate to the minimum allowed by the technology. After the floating-point operation is finished, the subsequent main CPU instructions can be executed with a higher clock rate, speeding up the overall performance. The ratio of the idling period to the active period, and the interleaving instruction ratio (i.e., the mix between idling and active periods, without having a long active period followed by a long idling one), will determine the overall gain in CPU performance.

The varying clock speed principle can be used in an even wider application. Consider a parallel computer architecture made of many "one-chip" CPUs. The interconnection network can be (and is in most cases) the performance bottleneck due to long latency times present for global information interchange. The cause for this time latency might be due to the interconnection network performance, or because of concurrently accessing common global data needed by a certain parallel program. Using the principle of varying the clock rates of each CPU, it is possible to make up for the lost waiting time, resulting in an overall performance improvement as if there was no time latency in accessing the global network.

Transaction processing is another good example. This application is characterized by short burst type operations, with idling periods in between. Increasing the clock rate while

processing the transactions is possible if the circuits can rest (i.e., produce little heat) while they are idling.

It should also be pointed out that while the description above utilizes three values for the circuit driving parameter which is varied, two (or four or more) values could be used instead. Also, the description above involves only drive parameter values which result in the circuit remaining in operation; however, a drive parameter value could be used which results in loss of operation of the circuit. This creates no problem when the circuit is idle, and even when the circuit is not idle the higher speed in the high speed mode might more than offset some enforced time in an idling inoperative mode while heat is being dissipated.

The variable clocking mechanism may be applied for use in any electronic circuit that is performance bounded by the heat-dissipation ability of its package, as long as the nature of the application using this circuit displays idling periods. Furthermore, the monitoring and control mechanism is so simple to build that it can be implemented on the same chip that it controls. That is, everything in Fig. 2 other than the system clock generator (and possibly that as well) could be moved onto the chip.

This concept may be applied to achieve a similar result in chip technologies in which the heat generation does not vary with clocking speed but instead with the operating current value of the circuits. Note that as the operating current level of the circuits increase, they switch faster. In all chip technologies there is a circuit driving parameter which can be varied in order to vary the heat being generated (e.g., drive current in ECL technology). In the next section, we examine the application of this technique to technologies other than CMOS.

3.0 Implementation Issues

The techniques described above, of monitoring circuit activity in a high performance system and reducing clock rates for circuits not active to conserve power, are not limited

only to CMOS technology. Reducing the power of inactive circuits could open up the possibility of having faster circuits operating at maximum power (maximum clock rates) and still realize the total chip power consumption to be within allowable limits. The discussion that follows describes a methodology for powering down silicon emitter coupled logic (ECL) circuits and GaAs and silicon FET NOR circuits when they are inactive. For GaAs and silicon FET NOR circuits, a "turbo-charged" high performance operation mode as well as the "normal" and power-down modes are possible.

The basic concept which enables the ECL and FET NOR circuits to be powered down is the ability to easily and rapidly reduce the operating current level of these circuits. Key to simplicity is a singular voltage node on each circuit which can be shared by many similar circuits. With a single node voltage changing, all common circuits to that node will be powered down. Secondly, the current and loading of this node need to be small so that many circuits can share node and simple, fast voltage generators can be designed. In ECL circuits, and the related DCS (differential current switch) and CML (current-mode logic) the current switch functions share a current source which is controlled by a voltage level. Similarly, the output emitter follower which generally utilizes a resistor as its current source, can also be modified to use an active current source which can be regulated by a voltage level.

The above ECL, DCS and CML circuits can be powered on and off rapidly, but cannot have their current level (and, consequently, their performance level) adjusted to several operating levels. Thus, high-performance, high-current macros and low-performance, low-power macros would be needed, performing the same function, which would be selected based on the value of the heat bank and the task at hand. The power level of FET NOR circuits is controlled by the current source in the load of the switching FET transistor. These current sources are generally depletion-load devices whose gate is generally shorted to the drain. Introducing a variable voltage source in the gate of the depletion load device can

regulate the current source magnitude and, consequently, the circuit power and performance level.

The predominant high-speed circuit used, albeit now at a considerable diminishing pace, is the classic ECL (emitter coupled logic) circuit [7], illustrated in Fig. 3. The power of the current switch function, T1, T2, T3, with R1, R2, and R3 is controlled by the current source (T5) connected to the common emitter. The value of the current is determined by the voltage across the resistor R_{bias1}. The power supply (-V2) is fixed, however, the voltage level at T5 input, V_{bias1}, is typically generated on chip. Thus, the current through T5 is $(V_{bias}-V_{be}(T5) - (-V2)) / R_{bias1}$. Since the current demand on V_{bias1} is very small, a very simple, emitter follower-type circuit could be used to generate this bias voltage. In fact, many ECL circuits will share the common node V_{bias1}, driven by bias source.

Although the standard ECL circuit generally has only a resistor in the emitter of the output device (T4), Fig. 3 assumes the version which has a transistor. This allows both the emitter current and the current-switch current to be turned on and off. Depending on design flexibility, V_{bias1} and V_{bias2} may be either the same or different in value. Thus, the operating current can be dramatically reduced or turned-off by the simple control of a bias voltage.

By introducing a switching function in the bias circuit, as shown in Fig. 4, one can provide, with the appropriate control logic, either V_{bias} or a more negative voltage to the inputs of the current sources. In Fig. 4, normally T1 is off and V_{bias} will be at its most positive level. This level is V2 + 1.6V (approx.) $-V_{be}(T3)$. Adding another series diode to Td1, Td2, will increase V_{bias} by another .8V. Applying a positive input (above V_{ref}) to T1 will turn T1 on. This switching bias circuit will be relatively fast (under 1 ns), with its speed being determined in part by its current level and the number of loads at its output.

Thus, by turning the bias on and off (or at least turning it to a low value) the ECL's current switch circuit will be provided with two operating modes: **on** (at high power) or **off** (on at low power or possibly totally off). Note, however, that at the low power or off-state,

the output levels (in-phase and out-of-phase) are both in the high (ONE) state. To assure no false data reaches any storage elements during the low power state of the function, the storage elements (such as, latches, etc.) have their data inputs disabled. For the applications considered, this limitations is not a problem.

Other high-speed bipolar circuits that can be powered down in a similar manner include the basic differential current mode (CML) (or differential current switch) circuits illustrated in Fig. 5. Switching V_{bias} in these circuits also powers down the circuits by reducing the operating current. These types of circuits are particularly attractive as latches, adders, etc. [8], i.e., complex functions which would be natural to power up and down as complete functional islands. The concept of a current-steering logic is also used in the GaAs technology. Specifically, the SCFL (Source-Coupled FET Logic) family, as discussed in [9], is analogous to the ECL circuit family and, as a result, can be power regulated by controlling the current source in a manner similar to that described above.

Consider the standard N-channel MOSFET enhance-deplete NOR circuit, or the equivalent GaAs MESFET enhance-deplete circuits, or the direct-coupled FET logic (known as DCFL) which are discussed in [10]. Their power level, and therefore their performance, is controlled by the current flowing through the load device, the depletion FET. These circuits are also able to be powered up and down by controlling the gate to source voltage of the load device, which in turn regulates the magnitude of current in the load device. In the case of the MOSFET, with its insulated gate, only a variable voltage source Vb is needed to regulate the current flow of the load device (Fig. 6). The current flow will be determined by the voltage difference between Vb and the source voltage (OUT mode), as mentioned earlier. In the case of the GaAs MESFET, since the gate becomes conducting when the value of the gate-to-source voltage exceeds a given drain-to-source value, depending on the design of the device, it is necessary to introduce a resistor between the voltage source Vb and the gate of the load device, to limit the current flow into the gate. It is, therefore, possible to regulate the operating current level and, consequently,

the performance of silicon MOSFET and GaAs MESFET circuits by a variable voltage source. The value of this voltage source can be determined in the same way as discussed above when defining the appropriate clock frequency.

4.0 Conclusions:

Current and future VLSI designs are increasingly constrained in realizing their optimum performance due to chip heat generation or current drain on the power supply batteries. The approach described in this paper provides a real-time power-performance optimization methodology which not only works well with CMOS circuits by varying the chip or chip macro clocks but can be extended, as well, to other technologies by varying the current levels of the logic circuits.

References:

[1] M. Fitzgerald, "Everex set to show 50-MHz PC," *Computerworld*, Vol. XXIV, No. 41, p. 4, October 8, 1990

[2] L.L. Sheets, "Electrical System Having Variable Frequency Clock," U.S. Patent #4,670,837, issued on June 2, 1987

[3] N. Krohn, "Intel to Introduce 20-MHZ Chip Set," *InfoWorld,* Vol. 12, Issue 42, p. 5, October 15, 1990

[4] Jim Nash, "Dual-clock 486SX: Bonanza for Intel?," *Computerworld,* p. 89, Aug. 5, 1991

[5] C.J. Georgiou, T.A. Larsen, E. Schenfeld, "High Performance Circuits with Power-down and Turbo-charged Operating Modes," *IBM Technical Disclosure Bulletin*, vol. 34, no. 10A, March 1992.

[6] C.J. Georgiou, T.A. Larsen, E. Schenfeld, "A Variable Chip Clocking Mechanism," U.S. Patent #5,189,314, issued on 2/23/93

[7] K. Ohno, H. Takeda, "High-Speed Bipolar Logic IC," *Fujitsu Sci. Tech. Journal,* Vol. 24, no.4, December, 1988, pp. 265-270.

[8] W. Heimsch, B. Hoffmann, R. Krebs, E. Mullner, B. Pfaffel, K. Ziemann, "Merged CMOS/Bipolar Current Switch Logic (MCSL)," *IEEE Journal of Solid-State Circuits,* Vol. 24, No. 5, October, 1989, pp.1307-1311.

[9] S. Shimizu, K. Yoshihara, T. Terada, K. Ishida, Y. Kitaura and C. Takubo, "An ECL-Compatible GaAs SCFL Design Method," *IEEE Journal of Solid-State Circuits,* Vol. 25, No.2, April, 1990, pp. 539-545.

[10] S. Long and S. Butner, "Gallium Arsenide Digital Integrated Circuit Design," McGraw Hill, 1990

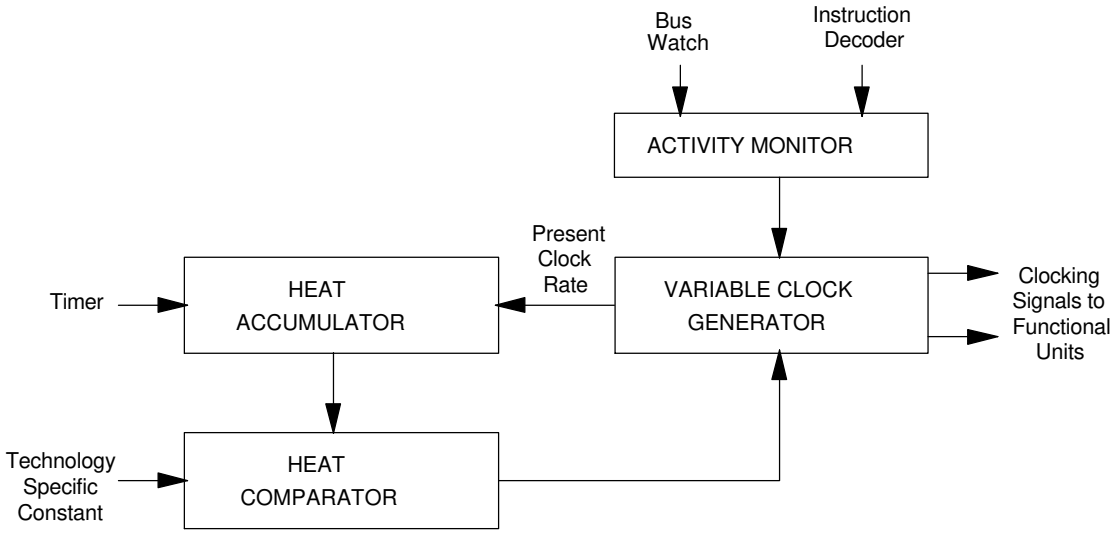
Bus
Watch
Instruction
Decoder
ACTIVITY MONITOR
Present
Clock
Rate
Timer
HEAT
ACCUMULATOR
VARIABLE CLOCK
GENERATOR
Clocking
Signals to
Functional
Units
Technology
Specific
Constant
HEAT
COMPARATOR

Figure 1:

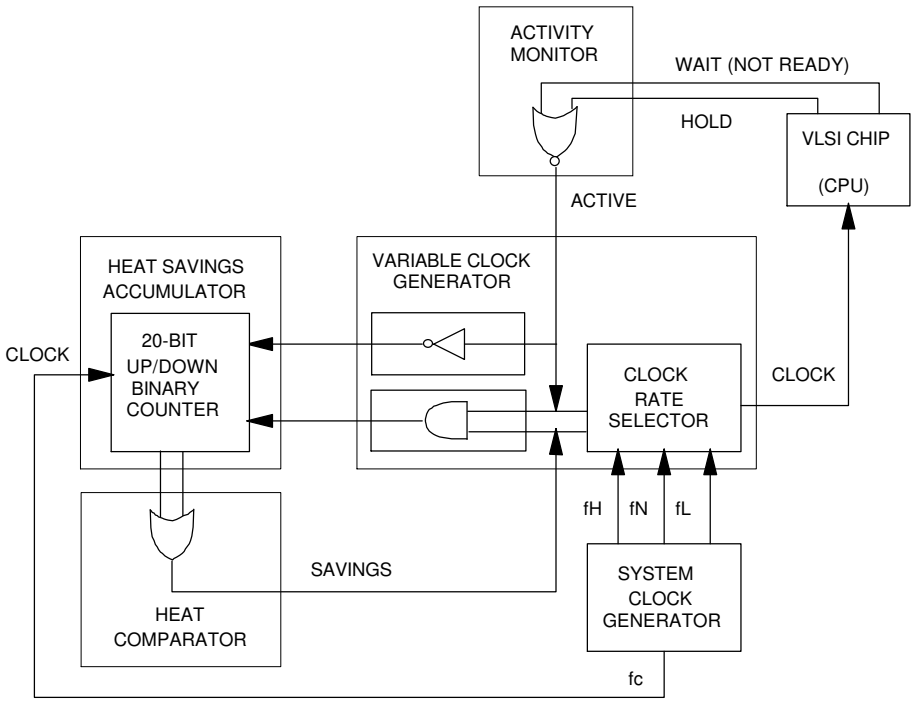

Figure 2

Active	Savings	Clock	Up	Down
1	1	f_H	0	1
1	0	f_N	0	0
0	X	f_L	1	0

Table 1:

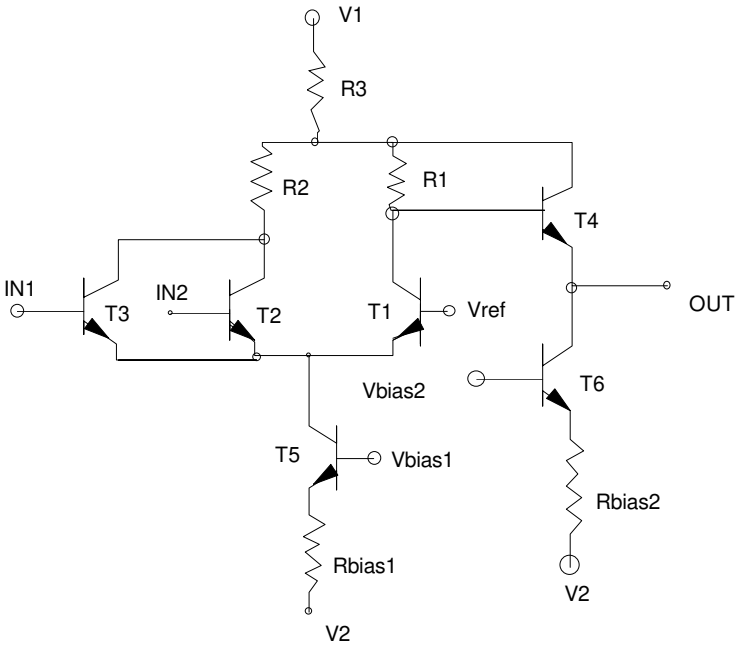

Figure 3

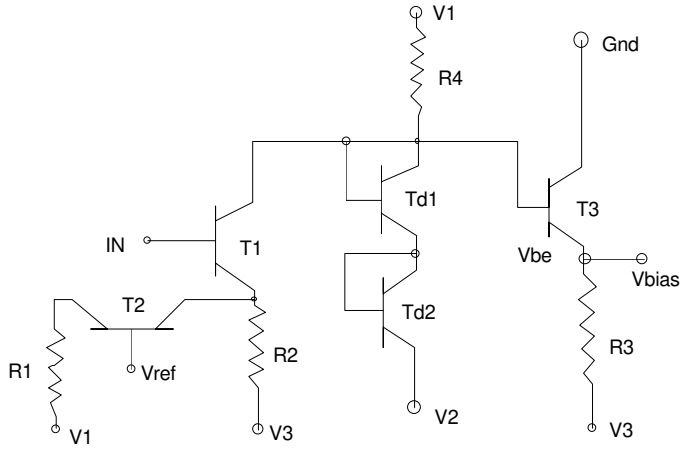

Figure 4

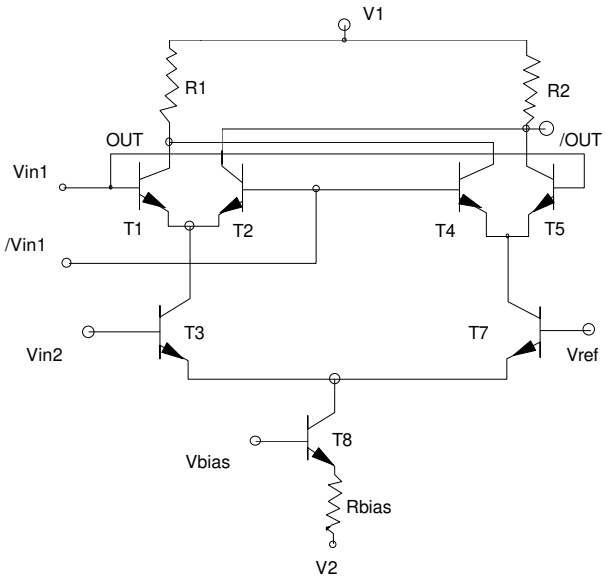

Figure 5

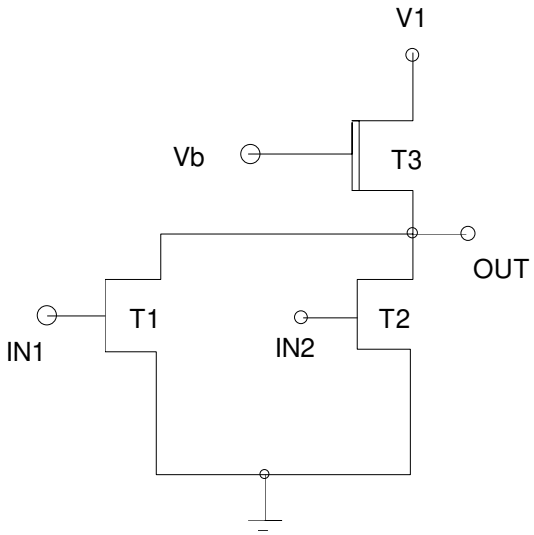

Figure 6