Energy Minimization of System Pipelines Using Multiple Voltages

Gang Qu†, Darko Kirovski†, Miodrag Potkonjak†, and Mani B. Srivastava‡
† Computer Science Department, University of California, Los Angeles, CA 90095-1596
‡Electrical Engineering Department, University of California, Los Angeles, CA 90095-1596

Abstract

Modern computer and communication system design has to consider the timing constraints imposed by communication and system pipelines, and minimize the energy consumption. We adopt the recent proposed model for communication pipeline latency[23] and address the problem of how to minimize the power consumption in system-level pipelines under the latency constraints by selecting supply voltage for each pipeline stage using the variable voltage core-based system design methodology[11]. We define the problem, solve it optimally under realistic assumptions and develop algorithms for power minimization of system pipeline designs based on our theoretical results. We apply this new approach on the 4-stage Myrinet GAM pipeline, with the appropriate voltage profiles, we achieve **93.4%, 91.3%** and **26.9%** power reduction on three pipeline stages over the traditional design.

1 Introduction

System level pipelines are widely acknowledged as the most likely bottleneck of many computer systems [16, 20]. For example, a read miss in the system data or instruction cache blocks the application program until the entire block with requested data arrives [1, 22]. The trade-off is clear: longer blocks imply fewer misses, but also longer interrupt latency. Similarly, in high speed local and wide-area networks selecting properly block size to exploit intrinsic concurrency in communication pipelines is a key issue [2, 6, 25]. As the final example where communication pipelines dictate performances we mention path-oriented operating systems [17].

Therefore, it is not surprising that recently the question of how to improve the performance of a system pipeline received a great deal of attention in computer architecture, operating systems, and compilers communities. The essence of the problem is abstracted in recent work by Wang et al [23] where they discuss how to minimize the transmission latency by carefully packet fragmentation.

On the other hand, the increasing use of portable systems (such as personal computing devices, wireless communications and imaging systems) makes the power consumption one of the primary circuit and system design goals. The most effective method to reduce power consumption is to lower the supply voltage level, which exploits the quadratic dependence of power on voltage [4]. However, reducing the supply voltage increases circuit delay and decreases the clock speed. The resulting processor core consumes lower average power while meets the deadlines. Unfortunately, this technique becomes ineffective when tight deadlines are present in systems.

Recent progress in power supply technology along with custom and commercial CMOS chips that are capable of operating reliably over a range of supply voltages makes it possible to build processor cores with supply voltages that can be varied at run time according to the application latency constraints [18]. The variable voltage processor core is capable of operating at different optimal points along the power and speed curve in order to achieve high energy efficiency. In particular, with multiple supply voltages on the chip, the processor core can use high voltage for applications with tight deadlines while keep the voltage low for others to reduce the total energy consumption.

In this paper, we address the energy minimization problem in system-level pipelines under latency constraints. We use the recent advances in power supply technologies and the variable voltage design methodology to choose a voltage profile for each pipeline stage which optimally minimizes the energy consumption of the entire pipeline system.

A Motivational Example

To illustrate the key ideas behind our new approach, we consider a small communication system shown in Figure 1. The system consists of 3 store-and-forward stages operated by three identical processors. Assume stage 1 has the slowest transmission speed, and a packet of 4 equal-size fragment has to be sent through this system by a deadline T.

The transmission starts from time 0 and is completed at T. Figure 1 shows three different strategies. Each rectangle represents the transmission of one fragment at one of the three stages. The base of the rectangle is the time that the fragment stays at that stage, while the height can be considered as the supply voltage.

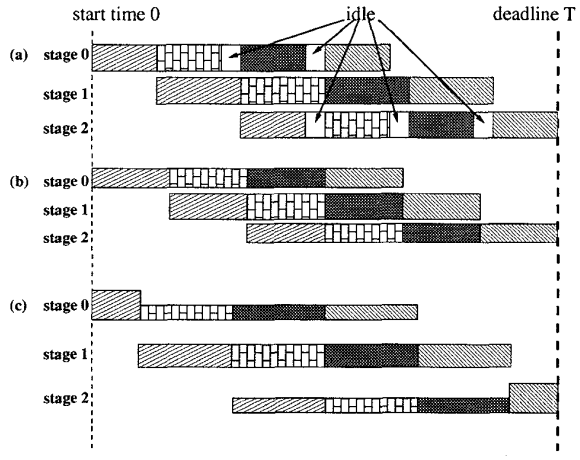

Figure 1: A 3-stage pipeline system transmits a 4-fragment packet.

Traditional processors run at a fixed supply voltage. The total energy consumption is minimized at the lowest possible supply voltage which guarantees a finishing time T. Further calculation results in solution (a) in Figure 1, where we can see processors on stages 0 and 2 have been idle due to stage 1's slow transmission speed. The total energy can be reduced by applying different voltages on different stages. As shown in (b), all stages are synchronized after reducing the supply voltages on stages 0 and 2. More

0-7803-5471-0/99/$10.00©1999 IEEE

energy efficiency is possible when we vary the supply voltage levels on each processor. Since the total energy consumption is dominated by stage 1, which requires the highest voltage, using high voltages on stage 0 for the first fragment and on stage 2 for the last fragment saves transmission time for stage 1. With more transmission time, stage 1 requires lower voltage and thus could reduce the total energy consumption. The concept is shown in (c).

The rest of the paper is organized as follows, we review the related work in communication pipeline and low power design techniques, then we define the problem in section 3. We solve the problem optimally in two cases: (i) each pipeline stage has a fixed voltage which varies from stage to stage; (ii) every stage can have variable supply voltages. We present the experimental results in section 6 and then conclude.

2 Related Work

The most relevant related work are efforts in communication pipeline design and evaluation, and low power design techniques. In particular, within the former domain fragmentation techniques for managing congestion control, packet buffering, packet losses, and the optimization techniques for improvement of distributed file systems and high-speed local area networks are directly relevant. Within the latter, we focus our survey on system-level power minimization techniques and variable voltage techniques.

In the introduction section, we already surveyed a number of communication-pipeline systems and research efforts for latency optimization of these systems. It is important to note that many application specific systems operate at the highest-level of abstraction as processing pipelines on blocks of input. Fragmentation has been used in the design of Internet for quite a long time. More recently, studies of how to exploit flexible block fragmentation to improve performances of DEC workstations has also been conducted [13]. More detailed survey of fragmentation techniques is given in [23].

Dynamically adapting voltage and therefore the clock frequency, to operate at the point of lowest power consumption for given temperature and process parameters was first proposed by Macken et al [14, 15]. Later, [12] described implementation of several digital power supply controllers based on this idea. Nielsen et al [19] extended the dynamic voltage adaptation idea to take into account data dependent computation times in self-timed circuits. Recently several researchers developed efficient DC-DC converters that allow the output voltage to be rapidly changed under external control [18]. Researchers at MIT [5, 10] have applied the idea of voltage adaptation based on data dependent computation time from [19] to synchronously clocked circuits.

In the software world, also there has been recent research on scheduling strategies for adjusting CPU speed so as to reduce power consumption. The existing work is in the context of non-real-time workstation-like environment. [24] proposed an approach where time is divided into 10-50 ms intervals, and the CPU clock speed (and voltage) is adjusted by the task-level scheduler based on the processor utilization over the preceding interval. [9] concluded that smoothing helps more than prediction in voltage changing. Finally, [27] described an off-line minimum-energy schedule and an average rate heuristic for job scheduling for independent processes with deadlines.

A great variety of system-level low power techniques has been proposed. For comprehensive surveys see [7, 26]. Energy efficient microprocessor design has been discussed in [3, 8]. Hong et al. [11] describes a design methodology for the real-time system-on-chip based on dynamically variable voltage processor cores.

3 Problem Formulation

The variable voltage is generated by the DC-DC switching regulators, the amount of time for the voltage to reach steady state at the new voltage is in the order of 10 cycles in a micro-processor [18]. In most part of this paper, we use the *ideal variable voltage processor* [21] where the supply voltage can be changed from 0 to ∞ instantaneously without any overhead. Although this ideal processor is not feasible the study of this model gives us insight view of the problem and more important, it provides the lower bound of energy consumption by using variable voltage processors.

As proposed in [23], we represent the communication system as a sequence of store-and-forward pipeline stages characterized by $\{n, g_i, T_i(v_{ref})\}$. There are n pipeline stages in the system, for each stage i, g_i is the fixed per-fragment overhead and $T_i(v_{ref})$ is the per-byte transmission time with a reference supply voltage. g_i can be considered as the context switch time. It may vary from stage to stage. $T_i(v_{ref})$ is proportional to the inverse of the bandwidth and high voltage implies a high transmission speed.

A packet of size B (e.g. in byte) has to be transmitted through the pipeline with latency constraint T. We send the packet in k fragments to utilize the pipeline, denote x_i the size of the ith fragment and $t_{i,j}$ the time that the ith fragment stays on stage j.

Let $v_j(t)$ be the voltage at which the jth processor operates at time t, then

$$E_j = \int_0^T P(v_j(t))\, dt \tag{1}$$

is the energy consumed by this processor, where $P(v)$ is the power dissipation at supply voltage v. We want to minimize $E = \sum_{j=0}^{n-1} E_j$ by finding the best voltage and fragment schemes.

The problem is formulized as:

Problem: Energy Minimization with Deadline on Variable Voltage Processor(EMDVVP).
Instance: A pipeline with parameters n, g_i and $T_i(v_{ref})$, a packet with size B and deadline T.
Question: Find the voltage scheme $v_j(t)$ for each processor and a fragment $\{x_0, x_1, \dots\}$ of the packet, such that the entire packet is transmitted within T and the total energy consumption $E = \sum_{j=0}^{n-1} \int_0^T P(v_j(t))\, dt$ is minimized.

Figure 2: Problem formulation.

4 Design of System Pipelines Using Multiple Voltages

To design application specific and energy efficient system pipelines with the variable voltage processors, we have to solve the EMDVVP problem based on the user-specified packet information (i.e., packet size B and transmission latency T) and the parameters of the system pipeline (number of pipeline stages n, per-fragment overhead g_i, transmission speed $T_i(v_{ref})$ as well as the power dissipation functions.).

Lemma 4.1 A necessary condition for the energy to be minimized is to finish the transmission exactly at the deadline T.

The intuition behind Lemma 4.1 is that the system will use as much time as possible to schedule the processors with low voltages and thus minimize energy consumption. On the other hand, from the convexity of the energy and voltage function [21], we have:

Lemma 4.2 On every stage, to minimize the energy, supply voltage changes on either the arrival of a new fragment or the accomplishment of sending the current fragment.

Recall that $t_{i,j}$ is the time that the ith fragment stays in the jth stage, which includes both the overhead g_j and the actual transmission time. for each single stage, the best strategy is to transmit a fragment immediately upon its reception or at the accomplishment of sending the previous fragment whichever comes later. This observation leads to the next lemma:

Lemma 4.3 In the optimal voltage and fragmentation schemes, for all $0 \leq i \leq k-2$ and $1 \leq j \leq n-1$, the following holds:

$$t_{i,j} = t_{i+1,j-1} \tag{2}$$

4.1 Fixed Voltage on the Same Stage

We first consider the simple case when the processor at each stage operates at a fixed voltage which can be arbitrary. The voltage scheme problem then becomes to finding a constant v_j for the processor at the jth stage, and $t_{i,j}$ can be expressed as:

$$t_{i,j} = g_j + T_j(v_j)x_i \tag{3}$$

Assume that the packet can only be fragmented into equal size fragments, then from (2),

Lemma 4.4 A voltage scheme $\{v_0, v_1, \ldots, v_{n-1}\}$ minimizes the energy consumption only if

$$t_{i,j} = constant \tag{4}$$

From (3), the processor at the stage that has the largest per-fragment overhead has to operate at a high voltage to achieve a small per-byte transmission time $T_j(v_j)$ due to (4). Therefore, this stage will consume more energy than other stages and we call such a stage *dominant stage* because it dominates the total energy consumption.

Theorem 4.1 Let stage d be the dominant stage, then there is a unique solution for the EMDVVP problem. The number of fragments is given by:

$$k = \sqrt{\frac{T}{g_d}(n-1)} - (n-1) \tag{5}$$

and the constant on the r.h.s. of (4) is $\frac{T}{n+k-1}$.

4.2 Variable Voltages on the Same Stage

Now we assume each fragment can have variable size and each processor can run at different level of voltage.

As formulated in Figure 2, a solution to the EMDVVP problem means a supply voltage function for each processor and a packet fragmentation.

Lemma 4.2 outlines the shape of the voltage functions, which are step functions with all possible break points at the time when new fragment arrives or current one leaves. Therefore we only need to determine the supply voltage $v_{i,j}$ for each processor to transmit each fragment, which reduces the problem from finding n functions to determining nk numbers, where k is the number of fragments.

Lemma 4.3 predicts a recursive relation among the time that fragments stay at each stage, from which $(n-1)(k-1)$ $v_{i,j}$'s can be easily determined.

Lemma 4.1 tells us the energy is minimized only when the entire transmission finishes at the deadline, so one more variable can be eliminated. Combining all these, we propose an approach to the optimal scheme in Figure 3 and draw the following conclusion:

Theorem 4.2 Given the number of fragments, the EMDVVP problem with variable-sized fragment and variable voltage at each stage is reduced to solving a nonlinear system (step 6 in Figure 3) of $n + 2k - 3$ free variables.

Numerical solution can be obtained from the empirical power and speed function. Furthermore, by repeating this approach for all possible values of k, we can solve the EMDVVP problem optimally.

1. Let $\{x_0, x_1, \ldots, x_{k-1}\}$ be a fragmentation, with x_{k-1} given by $B - \sum_{i=0}^{k-2} x_i$.
2. Let $\{v_{0,0}, v_{1,0}, \ldots, v_{k-1,0}\}$ be the voltage scheme for the processor at stage 0.
3. Let $\{v_{k-1,0}, v_{k-1,1}, \ldots, v_{k-1,n-1}\}$ be the voltages at each stage to transmit the last fragment of the packet, where $v_{k-1,n-1}$ is solved from the latency constraint $\sum_{i=0}^{k-1} t_{i,0} + \sum_{j=1}^{n-1} t_{k-1,j} = T$.
4. For each stage $j(1 \leq j \leq n-1)$, calculate its voltage scheme $\{v_{i,j} : 0 \leq i \leq k-2\}$ from (2),(3).
5. Total energy consumption: $E = \sum_{j=0}^{n-1} \sum_{i=0}^{k-1} P(v_{i,j})t_{i,j}$.
6. Solve all the variables in steps 1 2, and 3 from the system

$$\begin{cases} \frac{\partial E}{\partial x_i} = 0, & \text{for } 0 \leq i \leq k-2 \\ \frac{\partial E}{\partial v_{i,0}} = 0, & \text{for } 0 \leq i \leq k-1 \\ \frac{\partial E}{\partial v_{k-1,j}} = 0, & \text{for } 1 \leq j \leq n-2. \end{cases}$$

Figure 3: An approach to the optimal scheme.

5 Experimental Results

We report the results when apply our new energy minimization approach on the Myrinet GAM pipeline that Berkeley researchers use to study the packet fragmentation and to build the model for system pipeline evaluation[23].

Myrinet GAM pipeline consists of four stages, stage 0 copies data on the sender host; stage 1 is the sender host DMA; the next stage is an abstract pipeline stage of the network DMAs at both end hosts and a receiver host DMA; stage 3 is the copy on the receiver host. The parameters of this pipeline are given in Table 1[23]. The second column is the per-fragment overhead, the third column is the per-kilobyte transmission time at the reference supply voltage, the last column is the reference power for each stage at the reference supply voltage. It is clear that stage 2 is the dominant stage since it has the largest per-fragment overhead and the slowest transmission speed.

stage j	$g_j(\mu s)$	$T_j(\mu s$/KB)	P_j(Watt)
0	7.2	7.2	P_0
1	5.2	24.9	P_1
2	7.5	24.9	P_2
3	7.4	7.9	P_3

Table 1: Myrinet GAM pipeline parameters.

Suppose there is a 4KB-packet being transmitted via this pipeline with various user-specified latency constraints, we apply the variable voltage approach with fixed-size fragmentation to schedule the supply voltage for processors at each stage. The result is shown in Table 2.

The traditional energy minimization technique tries to find the minimal supply voltage and then apply it to the processors at all stages to meet the deadline constraint. In this case, this voltage is that in stage 2. Table 3 compares the energy consumption at each stage by our new approach vs. the traditional method. At both end hosts (stages 0 and 3), significant amount of energy (more than 90% in average) is saved due to the high transmission speed at these two stages. On stage 1, the average 26.9% energy reduction comes from its small overhead g_1.

6 Conclusion

Variable size packet fragmentation can reduce transmission latency and variable voltage processors are capable for power efficiency system design. We combine these techniques to address the problem of how to minimize the power consumption in system-level pipelines under latency constraints. We define the problem and solve it optimally based on the communication pipeline model[23]

Latency (μs)	Number of fragments	stage 0		stage 1		stage 2		stage 3	
		voltage (v)	power (P_0)	voltage (v)	power (P_1)	voltage (v)	power (P_2)	voltage (v)	power (P_3)
200	6	2.49	8.02e-02	4.96	0.97	5.52	1.40	2.63	9.97e-02
250	7	2.11	4.13e-02	3.97	0.45	4.32	0.61	2.22	5.04e-02
300	8	1.91	2.64e-02	3.43	0.27	3.68	0.35	1.99	3.19e-02
360	9	1.72	1.67e-02	2.96	0.16	3.13	0.19	1.79	1.99e-02
420	10	1.61	1.21e-02	2.67	0.11	2.81	0.13	1.67	1.43e-02

Table 2: Optimal voltage schemes for Myrinet GAM pipeline to transmit a 4KB packet.

Latency (μs)	power at stage 0 (P_0)			power at stage 1 (P_1)			power at stage 2 (P_2)			power at stage 3 (P_3)		
	traditional	new	saving	traditional	new	saving	traditional	new	saving	traditional	new	saving
200	1.40	8.02e-02	94.3%	1.40	0.97	30.5%	1.40	1.40	0%	1.40	9.98e-02	92.9%
250	0.61	4.13e-02	93.2%	0.61	0.45	25.2%	0.61	0.61	0%	0.61	5.04e-02	91.7%
300	0.35	2.64e-02	92.4%	0.35	0.27	22.3%	0.35	0.35	0%	0.35	3.19e-02	90.8%
360	0.19	1.67e-02	91.4%	0.19	0.16	19.0%	0.19	0.19	0%	0.19	1.99e-02	89.7%
420	0.13	1.21e-02	90.6%	0.13	0.11	17.1%	0.13	0.13	0%	0.13	1.43e-02	88.8%
Average	0.54	3.53e-02	**93.4%**	0.54	0.39	**26.9%**	0.54	0.54	**0%**	0.54	4.33e-02	**91.3%**

Table 3: Energy reduction on Myrinet GAM pipeline for transmission of a 4KB packet.

and variable voltage processor model[21]. Even when restricted to equal size fragmentation and fixed voltage on each processor core, we show that significant power reduction is possible without additional latency.

References

[1] T.E. Anderson, M.D. Dahlin, J.M. Neefe, D.A. Patterson, and others. *Serverless network file systems.* ACM Transactions on Computer Systems, Vol. 14, No. 1 pp. 41-79, 1996.

[2] N.J. Boden, D. Cohen, R.E. Felderman, A.E. Kulawik, and others. *Myrinet: a gigabit-per-second local area network.* IEEE Micro, Vol. 15, No. 1 pp. 29-36, 1995.

[3] T.D. Burd, R.W. Brodersen. *Processor design for portable systems.* Journal of VLSI Signal Processing, Vol. 13, No. 2-3, pp. 203-221, 1996.

[4] A. Chandrakasan, S. Sheng, and R.W. Broderson. *Low-power CMOS digital design.* IEEE Journal of Solid-State Circuits, Vol. 27, No. 4, pp. 473-484, 1992.

[5] A. Chandrakasan, V. Gutnik, T. Xanthopoulos. *Data driven signal processing: an approach for energy efficient computing.* International Symposium on Low Power Electronics and Design, pp. 374-352, 1996.

[6] B.N. Chun, A.M. Mainwaring, D.E. Culler. *Virtual network transport protocols for Myrinet.* IEEE Micro, Vol. 18, No. 1 pp. 53-63, 1998.

[7] L. Claesen, H. De Kuyper, R. Tits. *Low power applications at system level.* Low Power Design in Deep Submicron Electronics. Ed.: W. Nebel, Mermet, J. Dordrecht, Netherlands: Kluwer Academic Publishers, pp. 543-64, 1997.

[8] R. Gonzalez, M. Horowitz. *Energy dissipation in general purpose microprocessors.* IEEE Journal of Solid-State Circuits, Vol. 31, No. 9, pp. 1277-1284, 1996.

[9] K. Govil, E. Chan, and H. Wasserman. *Comparing algorithms for dynamic speed-setting of a low-power CPU.* ACM International Conference on Mobile Computing and Networking (MOBICOM'95), pp. 13-25, 1995.

[10] V. Gutnik, and A. Chandrakasan. *An efficient controller for variable supply-voltage low power processing.* Symposium on VLSI Circuits, pp. 158-159, 1996.

[11] I. Hong, D. Kirovski, G. Qu, M. Potkonjak, and M.B. Srivastava. *Power Optimization of Variable Voltage Core-Based Systems.* Proceedings of 35th Design Automation Conference (DAC'98), pp. 176-181, 1998.

[12] M. Horowitz. *Low power processor design using self-clocking.* Workshop on Low-power Electronics, 1993.

[13] H.A. Jamrozik et al. *Reducing network latency using subpages in a global memory environment.* International Conference on Architectural Support for Programming Languages and Operating Systems, SIGPLAN Notices, Vol. 31, No. 9 pp. 258-67, 1996.

[14] V. Von Kaenel, P. Macken, M. G. R. Degrauwe. *A voltage reduction technique for battery-operated systems.* IEEE Journal of Solid-State Circuits, Vol. 25, No. 5, pp. 1136-1140, 1990.

[15] P. Macken, M. Degrauwe, M. Van Paemel, H. Oguey. *A voltage reduction technique for digital systems.* 1990 IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers, pp. 238-239, 1990.

[16] R.P. Martin, A.M. Vahdat, D.E. Culler, T.E. Anderson. *Effects of communication latency, overhead, and bandwidth in a cluster rchitecture.* (24th Annual International Symposium on Computer Architecture. ISCA '97). Computer Architecture News, Vol. 25, No. 2 pp. 85-97, 1997.

[17] D. Mosberger, L.L. Peterson. *Making paths explicit in the Scout operating system.* Second USENIX Symposium on Operating Systems Design and Implementation (OSDI), pp. 28-31, 1996.

[18] W. Namgoong, M. Yu, T. Meng. *A high-efficiency variable-voltage CMOS dynamic dc-dc switching regulator.* 1997 IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers, pp. 380-381, 1997.

[19] L. S. Nielsen, C. Niessen, J. Sparso, K. van Berkel. *Low-power operation using self-timed circuits and adaptive scaling of the supply voltage.* IEEE Transactions on Very Large Scale Integration (VLSI) Systems, Vol. 2, No. 4, pp. 391-397, 1994.

[20] L.L. Peterson & B.S. Davie. *Computer networks : a systems approach* San Francisco, Calif. : Morgan Kaufmann Publishers, 1996.

[21] G. Qu. *Scheduling Problems for Reduced Energy on Variable Voltage Systems* Master Thesis, Computer Science Dept., Univ. of California, Los Angeles, 1998.

[22] G.M. Voelker et al. *Managing server load in global memory systems.* ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS 97), Performance Evaluation Review, Vol.25, No.1 pp. 127-138, 1997.

[23] R.Y. Wang, A. Krishnamurthy, R.P. Martin, T.E. Anderson, D.E. Culler. *Modeling Communication Pipeline Latency.* Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '98/PERFORMANCE'98), Performance Evaluation Review, Vol.26, No.1 pp. 22-32, 1998.

[24] M. Weiser, B. Welch, A. Demers, S. Shenker. *Scheduling for reduced CPU energy* USENIX Symposium on Operating Systems Design and Implementation (OSDI), pp. 13-23, 1994.

[25] M. Welsh, A. Basu, T. von Eicken. *ATM and fast Ethernet network interfaces for user-level communication.* Third International Symposium on High-Performance Computer Architecture, pp. 332-342, 1997.

[26] A. Wolfe. *Issues for low-power CAD tools: a system-level design study.* Design Automation for Embedded Systems, Vol. 1, No. 4 pp. 315-332, 1996.

[27] F. Yao, A. Demers, S. Shenker. *A scheduling model for reduced CPU energy.* IEEE Annual Foundations of Computer Science, pp. 374-382, 1995.