A Survey of Dynamic Power Optimization Techniques

Li-Chuan Weng[†]
[†]*School of Electronic Engineering*
Dublin City University
Dublin, Ireland
wengli@eeng.dcu.ie

XiaoJun Wang[†]

Bin Liu[§]
[§]*Department of Computer Science*
TsingHua University
P.R. China

Abstract

One of the most important considerations for the current VLSI/SOC design is power, which can be classified into power analysis and optimization. In this survey, the main concepts of power optimization including the sources and policies are introduced.

Among the various approaches, dynamic power management (DPM), which implies to change devices states when they are not working at the highest speed or at their full capacity, is the most efficient one. Our explanations accompanying the figures specify the abstract concepts of DPM. This paper briefly surveys both heuristic and stochastic policies and discusses their advantages and disadvantages.

1. Introduction

The average power consumption of CMOS can be modelled with the following equation [5, 23, 24, 4, 28].

$$P_{avg} = P_{Dynamic} + P_{Static} \quad (1)$$

The dominate part of P_{avg} is the dynamic power, $P_{Dynamic}$, caused by $P_{switching}$ and $P_{short-circuit}$. Here we emphasize on the minimization of $P_{switching}$ which can be expressed by the following equation:

$$P_{switching} = K \cdot C_{out} \cdot V_{dd} \cdot V_{swing} \cdot f \quad (2)$$

In most of the cases, $V_{swing} = V_{dd}$, therefore the equation 2 can be re-written as

$$P_{swithcing} \approx K \cdot C_{out} \cdot V_{dd}^2 \cdot f \quad (3)$$

From the equation 3, there are four parameters that we can consider to change [15]. i.e. the node transition activity factor K, load capacitance C_{out}, supply voltage V_{dd}, and working frequency f.

2. Strategies

2.1. by Sources

As indicated in equation 3, the switching power consumption decreases quadratically with the decrease of V_{dd} and decreases linearly with the decrease of the other parameters. Therefore, there are various approaches depending on different parameters.

Swing voltage V_{swing}: Because most of the time V_{swing} equals to V_{dd}, most researches concentrate on lowering supply voltage, although there are some researches try to decrease the swing voltage from the range of V_{gnd} to V_{ddH} to the range of V_{gnd} to V_{ddL} [22].

Supply voltage V_{dd}: The policies applied on this parameter including supply shutdown, variable voltage, and partial shutdown. Because of the different features of computations that can be either *event-driven*[1] in-nature or *continuous* [2] in nature, we use different approaches.

For example, we decrease the power consumption of an event-driven system by system shutdown when it is at inactivity periods, and reduce the power consumption of a continuous system by scaling the supply voltage according to the characteristic of unbalancing workload. The concerns of the shut down strategies are "how" and "when" to apply. While the first issue can be done by either stop the clock or switch off the supply voltage, the second one has to concern more complicated problems. On the other hand, though the scaling of supply voltage helps continuous system to minimize the power consumption, there are some overheads due to the increase of T_d, the CMOS circuit delay. In order to solve this problem, some researches use higher sup-

[1] *Event-driven* means the computation activities are triggered by external events [5]

[2] *Continuous* means the continuous computation e.g. the applications in speech coding [5].

Proceedings of The 3rd IEEE International Workshop on System-on-Chip
for Real-Time Applications ISBN 0-7695-1929-6/03 $17.00 © 2003 IEEE

ply voltage in the critical path and use lower voltage in others so that the system can finish the works within given time while save the power consumed during the process [31, 6, 7, 14, 32, 10, 11].

Effective capacitance C_{eff}=$K \cdot C_{out}$: There are some approaches try to lower the effective capacitance C_{eff} by lowering the transitions parameter K, or minimizing the load capacitor C_{out}. Lowering the switching activities of the capacitors in a system can be achieved by restructuring the computation, communication, memory storage and hierarchy, and changing the data encoding, etc. [30].

Clock frequency f: The observation that reducing the clock frequency can decrease power dissipation is very intuitive, but it also implies the working time is extended i.e. the performance is influenced. The techniques such as clock gating, variable frequency, and clock generator shutdown all belong to this approach. Further discussion can be found in [5, 20].

2.2. Power management

In general, there are two groups of power minimization techniques, *static* and *dynamic* [17]. The former is applied during the design time and the latter manages to save the power consumption by scheduling the runtime behavior. As the policy dynamically changes the system's states, we call them *Dynamic power management approaches, (DPM).*

The basic assumption of DPM is, though the systems are designed to be able to work at peak state, most of the time they are not working at their full capacity [3]. As a result, DPM could save the power consumption by selectively placing components into low power consumption states such as idle or sleep [30, 29]. Prediction accuracy and power savings are the two criteria used to assess if the policy is appropriate or not. The heart of DPM is power manager (PM) which does not have a fixed shape. PM is an abstract idea exists in the system; it can be implemented in either software or hardware. In software, it can be applied on Operating System (OS) or written in hardware description language; in hardware, it can be applied to a circuit module or a chip [3].

3. DPM techniques

The basic idea of DPM is that, since the workload of a system is usually not constant, therefore, if we can predict its status then we can reduce the power dissipation by some means[1]. Though leave the device "ON" when there is no workload wastes power, simply shutting off components as soon as they are not used might be counter productive as illustrated in figure 1. This occurs when the system needs the idled components again within very short time. It is due to the fact that it takes time to recover their states and sometimes the power saved is less than the power consumed during states transitions[3, 17].

The concept of break-even time, i.e. the minimum length of idle time to achieve power saving [17], plays an important role here. If we do not consider the drawback of wake-up delay but only see from the energy point of view, the system should shutdown only when idle period t is longer than T_{be}. Equation 4 explains the origin of T_{be}.

$$\begin{aligned} P_w \cdot t &\geq E_{sd} + E_{wu} + P_s \cdot (t - T_{sd} - T_{wu}) \\ t &\geq \frac{E_{sd} + E_{wu} - P_s \cdot (T_{sd} + T_{wu})}{P_w - Ps} \\ T_{be} &= \frac{E_{sd} + E_{wu} - P_s \cdot (T_{sd} + T_{wu})}{P_w - Ps} \end{aligned} \quad (4)$$

In short, because of the latency penalty for wake-up,

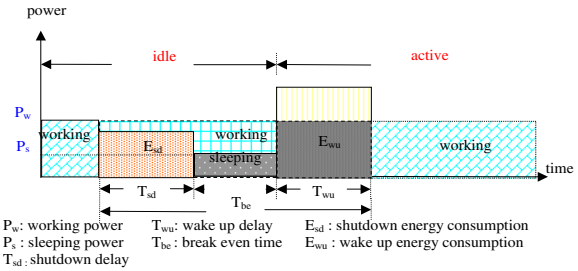

Figure 1. Break-even time

system should only be shutdown when $t \geq T_{be}$. For same idle time, longer break-even time T_{be} means less power saving due to the overhead for recovering state [1].

3.1. Heuristic policies

"When to shutdown" is the concern of heuristic policies [5], which can be classified into *time-out* and *predictive* approaches.

3.1.1. Timeout techniques

The time-out policies can be either *adaptive* or *non-adaptive*[27, 26]. In general, both policies change the system into idle state after the system is not performing any task for a period of time T_{pre}. In *non-adaptive* technique, T_{pre} is fixed, so it is also called fixed time-out policy. It is the simplest DPM policy [18] that uses

Proceedings of The 3rd IEEE International Workshop on System-on-Chip
for Real-Time Applications ISBN 0-7695-1929-6/03 $17.00 © 2003 IEEE

a time out value and assumes the system will continue to be idle for at least the break-even time after idle for the time out period [17], see figure 2. In the *adaptive* approach, T_{pre} can be adjusted by using the recent computation history to predict if the idle time will be longer than the break-even time or not.

Both *adaptive* and *non-adaptive* time-out techniques have two main disadvantages, one is the prediction must be accurate or it will lose rather than save power, and the other one is that there is still some power continuously dissipated during the waiting period [13, 1]. This method works when the idle time is long in average but does not perform very well on the other way round [13].

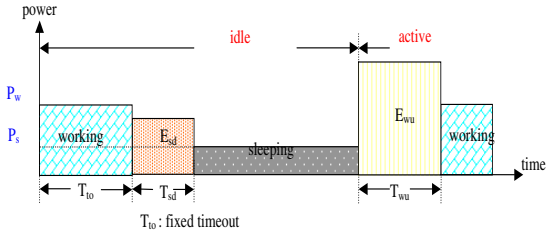

Figure 2. Fixed time-out

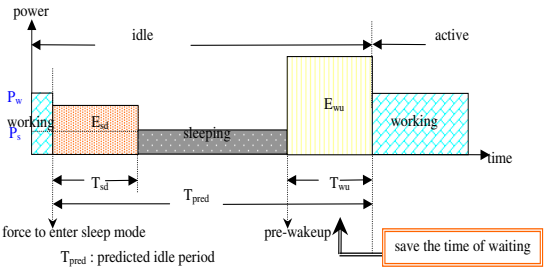

Figure 3. Predictive policy – correct prediction

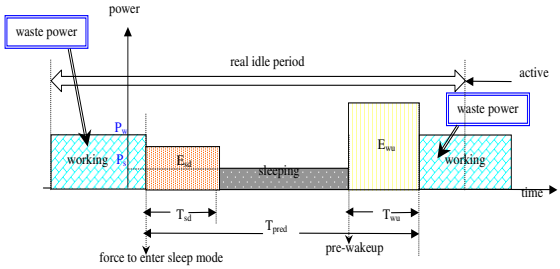

Figure 4. Predictive policy – wrong prediction

3.1.2. Predictive techniques

In stead of waiting for a fixed time period as time-out policies, the predictive techniques force the power state changes into low-power state as soon as it is predicted to stay idle for longer than break-even time, T_{be} [30]. Some policies try to avoid the waiting time of waking up as in fixed time-out policies by using prediction of the workloads [13], see figure 3.

However, from figure 4 we can see wrong predictions still waste power which might happen either when the component is forced to transit into idle too late or wake up too early [17]. Therefore, in order to avoid the overhead due to the wrong prediction, it becomes necessary to improve the hit ratio and reduce the delay overhead [13].

3.2. Stochastic techniques

In order to guarantee optimal result, there are some policies using stochastic process to model the the system and workload, i.e. the arrival of requests and device power-state changes [17, 29]. The current approaches are mostly based on Markov processes, including *stationary* and *non-stationary Markov models*, *discrete time* and *continuous-time Markov models*, *semi-Markov models* and *time-indexed semi-Markov models*.

Comparing to the stationary stochastic policy, the non-stationary stochastic policy generally has better power saving abilities when deal with the non-stationary wokloads due to its adaptive schemes [17]. The Discrete time Markov models evaluate the system at periodic time points, so it is not suitable for event-driven cases. Continuous time Markov models do not have the periodic limitation of the discrete time models but they require all the stochastic processes to be exponential. In order to relax the exponential distributions requirements, semi-Markov models are proposed, in which one of the stochastic process needs not to be of exponential distribution. Time-indexed semi-Markov models can have multiple non-exponential processes but it becomes even more complex.

Proceedings of The 3rd IEEE International Workshop on System-on-Chip
for Real-Time Applications ISBN 0-7695-1929-6/03 $17.00 © 2003 IEEE

4. Discussion and conclusion

While many studies focused on the high performance requirement, low-power system design now gets more and more attention because of the growing demands from portable computing and communication devices [30]. Detailed discussion of power optimization can be found in [9, 23, 2, 1, 17, 19, 3, 18].

The short discussion of resources of average power dissipation offers the overview of the concerns of power management. For systems with different characteristics we could apply either shutdown or scheduling techniques which are based on the consideration of supply voltage V_{dd} and clock frequency f. To efficiently apply shutdown, scheduling, and state transition, it is not trivial but complicated. The existing approaches are numerous and can be generally classified by the natures of the methods as heuristic and stochastic approaches. There is no known ideal solution yet, many of the methods are data or system dependent and some are too complex though their dependency are superior than the others.

This survey only discusses the optimization aspect but not the other important element, power analysis. [21, 25, 12, 8, 16] cover the induction of this aspect. The demand of power management is increasing and more support in both analysis and optimization are necessary.

References

[1] L. Benini, A. Bogliolo, and G. D. Micheli. A survey of design techniques for system-level dynamic power management. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 8(3):299–316, June 2000.

[2] L. Benini and G. de Micheli. System-level power optimization: techniques and tools. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 5(2):115–192, 2000.

[3] L. Benini, R. Hodgson, and P. Siegel. System-level power estimation and optimization. In *Proceedings 1998 international symposium on Low power electronics and design*, pages 173–178. ACM Press, 1998.

[4] L. Benini and G. D. Micheli. *Dynamic power management – design techniques and CAD tools*. Kluwer Academic Publishers, 1998.

[5] A. P. Chandrakasan and R. W. Brodersen. *Low power digital CMOS design*. Kluwer Academic Publishers, 1995.

[6] J. Chang and M. Pedram. Energy minimization using multiple supply voltages. In *Proceedings of the 1996 international symposium on Low power electronics and design*, pages 157–162. IEEE Press, 1996.

[7] J.-M. Chang and M. Pedram. Energy minimization using multiple supply voltages. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 5(4):436–443, 12 1997.

[8] R. Y. Chen, M. J. Irwin, and R. S. Bajwa. Architecture-level power estimation and design experiments. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 6(1):50–66, 2001.

[9] S. Devadas and S. Malik. A survey of optimization techniques targeting low power vlsi circuits. In *Proceedings of the 32nd ACM/IEEE conference on Design automation conference*, pages 242–247. ACM Press, 1995.

[10] S. Dhar and D. Maksimovic. Low-power digital filtering using multiple voltage distribution and adaptive voltage scaling (poster session). In *Proceedings of the 2000 international symposium on Low power electronics and design*, pages 207–209. ACM Press, 2000.

[11] M. Donno, L. Macchiarulo, A. Macii, E. Macii, and M. Poncino. Enhanced clustered voltage scaling for low power. In *Proceedings of the 12 th ACM Symposium on Great Lakes Symposium on VLSI*, pages 18–23. ACM Press, 2002.

[12] S. Gupta and F. N. Najm. Power macromodeling for high level power estimation. In *Proceedings of the 34th annual conference on Design automation conference*, pages 365–370. ACM Press, 1997.

[13] C.-H. Hwang and A. C.-H. Wu. A predictive system shutdown method for energy saving of event-driven computation. In *Proceedings of the 1997 IEEE/ACM international conference on Computer-aided design*, pages 28–32. IEEE Computer Society, 1997.

[14] M. C. Johnson and K. Roy. Datapath scheduling with multiple supply voltages and level converters. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 2(3):227–248, 1997.

[15] J. B. Kuo and J.-H. Lou. *Low-voltage CMOS VLSI circuits*. Wiley-Interscience, 1999.

[16] M. Lajolo, A. Raghunathan, S. Dey, and L. Lavagno. Cosimulation-based power estimation for system-on-chip design. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 10(3):253–266, 6 2002.

[17] Y.-H. Lu and G. D. Micheli. Comparing system-level power management. *IEEE Design and Test of Computers*, 18(2):10–19, March-April 2001.

[18] E. Macii. Dynamic power management of electronic systems. *IEEE Design and Test of Computers*, 18(2):6–9, March-April 2001.

[19] E. Macii, M. Pedram, and F. Somenzi. High-level power modeling, estimation and optimization. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 17(11):1061–1079, November 1998.

[20] T. Martin and D. Siewiorek. A power metric for mobile systems. In *Proceedings of the 1996 international symposium on Low power electronics and design*, pages 37–42. IEEE Press, 1996.

Proceedings of The 3rd IEEE International Workshop on System-on-Chip for Real-Time Applications ISBN 0-7695-1929-6/03 $17.00 © 2003 IEEE

[21] F. N. Najm. Towards a high-level power estimation capability. In *Proceedings 1995 international symposium on Low power design*, pages 87–92. ACM Press, 1995.

[22] J. Pangjun and S. S. Sapatnekar. Low-power clock distribution using multiple voltages and reduced swings. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 10(3):309–318, 6 2002.

[23] M. Pedram. Power minimization in ic design: principles and applications. *ACM Transactions on Design Automation of Electronic Systems (TODAES)*, 1(1):3–56, January 1996.

[24] J. M. Rabaey and M. Pedram. *Low power design methodologies*. Kluwer Academic Publishers, 1996.

[25] A. Raghunathan, S. Dey, and N. K. Jha. Register-transfer level estimation techniques for switching activity and power consumption. In *Proceedings of the 1996 IEEE/ACM international conference on Computer-aided design*, pages 158–165. IEEE Computer Society Press, 1996.

[26] D. Ramanathan. *High-level timing and power analysis of embedded systems*. PhD thesis, University of California, Irvine, 2000.

[27] D. Ramanathan and R. Gupta. System level online power management algorithms. In *Proceedings of the conference on Design, automation and test in Europe*, pages 606–611. ACM Press, 2000.

[28] K. Roy and S. C. Prasad. *Low-power CMOS VLSI circuit design*. John Wiley & Sons, Inc., 2000.

[29] T. Simunic. *Energy efficient system design and utilization*. PhD thesis, Stanford University, February 2001.

[30] M. B. Srivastava, A. P. Chandrakasan, and R. W. Brodersen. Predictive system shutdown and other architectural techniques for energy efficient programmable computation. *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, 4(1):42–55, March 1996.

[31] K. Usami and M. Horowitz. Clustered voltage scaling technique for low-power design. In *Proceedings 1995 international symposium on Low power design*, pages 3–8. ACM Press, 1995.

[32] K. Usami, K. Nogami, M. Igarashi, F. Minami, Y. Kawasaki, T. Ishikawa, M. Kanazawa, T. Aoki, M. Takano, C. Mizuno, M. Ichida, S. Sonoda, M. Takahashi, and N. Hatanaka. Automated low-power technique exploiting multiple supply voltages applied to a media processor. In *Custom Integrated Circuits Conference*, pages 131–134. IEEE, 1997.

Proceedings of The 3rd IEEE International Workshop on System-on-Chip for Real-Time Applications ISBN 0-7695-1929-6/03 $17.00 © 2003 IEEE