



Microelectronics System Design Research Group
School of Electrical, Electronic & Computer Engineering
Merz Court
Newcastle University
Newcastle upon Tyne, NE1 7RU, UK

Copyright © 2011 Newcastle University

<http://async.org.uk/>

Task scheduling based on energy token model

Danil Sokolov, Alex Yakovlev

`danil.sokolov@ncl.ac.uk`, `alex.yakovlev@ncl.ac.uk`

NCL-EECE-MSD-MEMO-2011-001

January 2011

Abstract

Energy becomes the most critical aspect in the design of modern microelectronic devices, which is reflected in the increasing number of conflicting requirements to the power characteristics of the circuits. However, modern design tools do not consider dynamics of energy resource, do not have flexible notion of energy efficiency and limit power optimisation to few design-specific techniques. We propose to address these limitations through a fundamental token-based model of computation, extended with the concept of energy tokens to represent energy flows in the circuit. The energy token model will serve as a foundation for building systems in such a way that while maintaining functional equivalence, their operation will be altered to meet energy mode requirements under dynamically changing energy resource and operating conditions.

Motivation

Energy becomes one of the most critical aspects in the design of modern microelectronic devices. This is reflected in the increasing number of conflicting requirements to the power characteristics of the circuits, e.g. to reduce energy consumption for longer battery life of portable devices; to restrict peak power for efficient heat dissipation of computation-intensive electronics; to balance power signature for avoiding power analysis attacks in security applications; to operate on unstable harvested energy for medical implants, etc.

These requirements have been partially addressed within the well-established area of research called low-power design. A number of ad hoc techniques are used for this, such as clock gating, power gating, voltage and frequency scaling [1]. Existing architectural solutions also target low-power designs only and cannot satisfy the whole spectrum of energy-related requirements. Circuit synthesis for low-power is based on design-and-validate paradigm and suffers from increasing complexity of multi-mode/multi-voltage analysis [2]. In these circumstances new fundamental solutions are needed as the next generation of semiconductors will bring even more challenges to efficient utilisation of energy [3].

We propose to address the above limitations through an abstract token-based model of computation, extended with the concept of energy tokens. This technology independent model will serve as a foundation for building systems in such a way that while maintaining functional equivalence, their operation will be altered to meet energy utilisation requirements. We believe that in the future these energy-related requirements will be expressed explicitly in a form of new type of design constraints, alongside constraints for area and timing, which will help to extend the notion of energy efficiency beyond low-power.

Energy token model

Traditionally the data flow in the circuit is captured by token-based models such as Petri nets [4]. In this model the computation tasks are represented by transitions and the data dependency between them is captured by the directed arcs. The tokens represent data validity and their propagation through transitions captures the data flows, as illustrated in Figure 1(a).

We propose to extend this model by introducing energy resource explicitly as a special type of *energy tokens*, depicted as hollow circles in Figure 1(b). These tokens are produced by the *energy source transition* at *pwr* rate and get collected in the *energy accumulator place* with the maximum capacity *cap*. In this model, in addition to the validity of the input data, each task requires a certain amount of energy *en* to complete, which is depicted next to the dotted *energy arc*. The notion of the task transitions is extended with the attributes of the *execution time* and *power consumption*. These attributes can be visualised as the dimensions of a task transition box, thus associating the size of a task transition with the amount of energy it requires.

Task scheduling

The energy token model is very convenient for exploration of the scheduling possibilities targeting different power utilisation profiles. Even the basic example of Figure 2(a) permits many execution scenarios with different trade-offs. For example, if tasks *A* and *E* run first then there is not enough energy to execute task *C* at the same time. Next, when task *C* starts there is sufficient energy for tasks *D* and *F*, however the causality relation

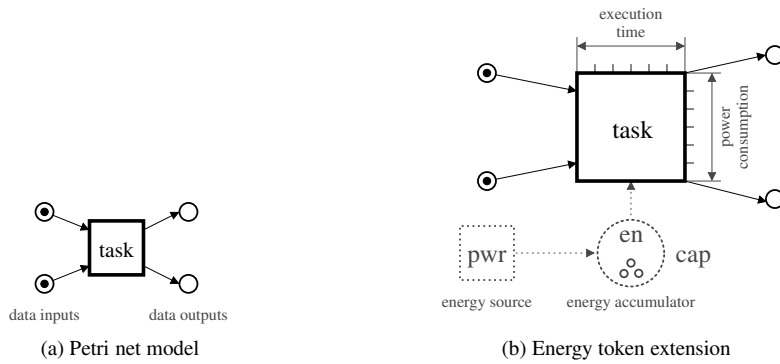


Figure 1: Token-based data flow model

does not allow them yet. The result is a sub-optimal power profile of the circuit depicted in Figure 2(b). A better power utilisation can be achieved if tasks *A* and *C* are enabled first, followed by *B* and *E*, and then by *D* and *F* - this execution scenario improves both the peak power and the speed of the modelled circuit, as shown in Figure 2(c).

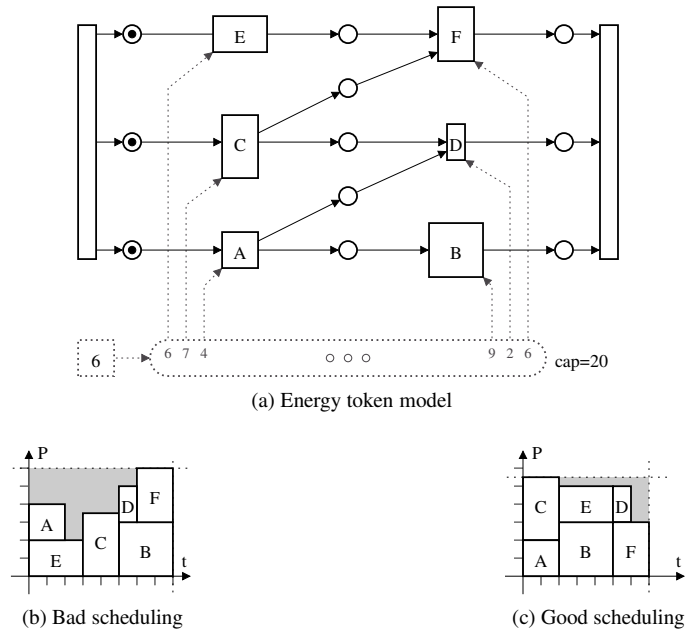


Figure 2: Basic scheduling example

Even more flexibility is added to the energy token model by the fact that the same task can be completed utilising different power control mechanisms of the computing components. The model in Figure 3(a) represents different voltage scaling modes for the components computing task *A*, decomposition options for task *B*, reduction of computation precision for task *C*, utilisation of different component implementations optimised for speed in case of task *E* or targeting power reduction for task *F*.

The scheduler can guide the circuit to reduce the peak power by decreasing the level of concurrency of

- *Fine-grain*, by using conventional power management mechanisms, such as clock and power gating, frequency and voltage scaling [1], as the knobs to fine-tune the circuit power profile. The controller can analyse the effects of power changes, e.g. by readings the temperature sensors or by monitoring the latency, and use the obtained information as a feedback to dynamically adjust the power control knobs.

Both these levels of power management granularity can be captured in the energy token model, which is convenient for analysis and justified choice of energy utilisation strategies to implement in the controller. The same model can be used for synthesis of the power control logic, possibly using a direct mapping approach for token-based graph models [7].

Challenges

The proposed model enables analysis of the scheduling scenarios targeting different energy utilisation requirements in a system. It is based on an abstract representation of power consumption and computation time of the tasks. In reality the things are more complicated - power consumption depends on input data, circuit operating conditions, physical location of the component on the chip, etc. It is a challenge to capture this multi-dimensional characteristic in an elegant form and still with sufficient level of detail. Furthermore, this is amplified by the complexity of modern systems with huge number of highly concurrent components. It is another challenge to make the energy scheduling decisions in run-time, based on the current state of the system and some approximations of its execution pattern in the future.

References

- [1] M.Keating, et al. "*Low power methodology manual: For system-on-chip design*", Series on Integrated Circuits and Systems, Springer, 2nd ed., 2008
- [2] S.Henzler. "*Power management of digital circuits in deep sub-micron CMOS technologies*", Springer-Verlag, 2006
- [3] [Kant10] K.Kant. "*Challenges in distributed energy adaptive computing*", ACM SIGMETRICS Performance Evaluation Review, 37(3), pp.3-7, 2010
- [4] R.David, H.Alla. "*Discrete, Continuous, and Hybrid Petri Nets*", Springer, 2004
- [5] M.Tartagni, et al. "*Energy autonomous systems: future trends in devices, technology and systems*", Report at Cluster for Application and Technology Research in Europe on NanoElectronics (CATRENE), 2009
- [6] D.Kinniment. "*Synchronization and arbitration in digital systems*", Wiley Publishing, 2008
- [7] D.Sokolov, et al. "*Direct mapping of low-latency asynchronous controllers from STGs*", IEEE Tran. on CAD, 2007