Low-Power Silicon Architectures for Wireless Communications

Jan M. Rabaey

Berkeley Wireless Research Center
University of California at Berkeley
tel: (510) 666 3111
e-mail: jan@eecs.berkeley.edu

Abstract — Wireless communication and networking is experiencing a dramatic growth, and all indicators point to an extension of this growth in the foreseeable future. This paper reflects on the demands and the opportunities offered with respect to the integrated implementation of these applications in the "systems-on-a-chip" era.

I. INTRODUCTION

Progress in wireless information systems, as envisioned by the 3rd generation and beyond, puts tremendous pressure on the underlying implementation technologies. Silicon technology has witnessed an unprecedented growth cycle in the last few decades, which will by-all-probability extend into the next few as well. The increasing complexity of silicon implementations closely follows a path prescribed by Moore's law. Yet, its is commonly observed that the complexity of wireless systems outpaces the technology evolution predicted by Moore's law. As a result, the implementation of future wireless systems will require novel thinking with respect to the implementation architectures and their relation to the applications at hand.

The wireless infrastructure and mobile components pose a set of stringent implementation requirements that are often considered to be contradictory:

- Reducing the cost in a multi-standard, evolutionary and adaptive environment requires flexible and evolvable solutions, which are often translated into software-programmable architectural implementations
- Yet these solutions often do not offer the performance and energy-efficiency that is required for the state-of-the art wireless algorithms and networks. The latter are more easily offered by semi-custom solutions, but these tend to lack the required flexibility.

As a result, the architectural community has been exploring a variety of novel architectures and implementation approaches, such as VLIW processors, processors with accelerators and reconfigurable processors.

This paper will first discuss the opportunities offered by the progress in semiconductor technology. This is followed by an exploration of the needs of future wireless systems. It will then progress into an overview of the different solutions, and provide clear metrics in how to evaluate and compare the options. Actual values for metrics such as flexibility, cost, performance and energy-efficiency will be offered with the aid of some actual benchmark examples.

II. THE SYSTEM-ON-A-CHIP

The continued progress of semiconductor technology has enabled the "system-on-a-chip" to become a reality, combining a wide range of complex functions on a single die. Integrated circuits that merge core processors, DSPs, embedded memory, and custom modules have been reported by a wide range of companies. Projections of future integration densities suggest that this trend will surely continue in the next decade.

One of the most dramatic consequences of this super-integration is a fundamental change in the use of computation in our society. In the pc-era, computation was mostly confined to the desktop or server box. The possibility of performing millions (or billions) of operations in a module of a cubic cm has broken that lock, and complex computation is now emerging in a ubiquitous fashion, ranging from set-top boxes, consumer media components, appliances, and sensor and monitor devices. This revolution in the way in computation is performed is extremely exciting, and opens the door for many novel opportunities, some of which we cannot even envision today.

Yet at the same time, the "system-on-chip" approach introduces some major challenges, that have to be addressed for the technology to become a viable undertaking. First of all, one has to address the economics of the production and realization process of these complex designs that can integrate billions of transistors on a single die:

- The cost of production facilities and mask making has increased significantly. This has lead to a major increase in the NRE (non-recurring engineering) costs of a design, and translates into a higher entry barrier for a new design start to be economically attractive.
- A larger number of physical effects (such as capacitive and inductive coupling introduced by interconnect wires) have come to the foreground in the design of these deep-submicron devices. This negatively impacts the design process. Realizing a correct design in a reasonable amount of time is becoming increasingly harder.
- While the complexity of the designs is increasing from year-to-year, the productivity of the integrated-circuit designer is not keeping pace.

All these observations, combined with an intense pressure to reduce the time-to-market, are causing a shift in the design paradigm. If one can define a common hardware

denominator that can be shared across multiple applications, volume increases and overall costs may eventually be (much) lower than in the case when the chip is customized for the application. This approach has been dubbed **platform-based design** [ASV99]. The applicability and hence the reusability of a single platform is inherently determined by its **flexibility**, i.e. its capability of covering a range of functionalities and applications [Rab97].

The most flexible platforms are the software programmable components. Here, flexibility is achieved using a "machine" capable to perform a set of instructions specified by the instruction set architecture (ISA). Microprocessors and, to a certain degree, DSPs, are examples of such. A systems engineer might be naturally drawn to these software-programmable platforms because of their ultimate flexibility. Yet, the embedded applications, targeted by these SOCs, impose a different set of design metrics and constraints than the pure performance goals, set by the traditional software platforms. For instance, in the world of wireless networks, cost and energy dissipation are the dominant factors determining the potential success of a given implementation. These broader metrics as well as the specific properties of the addressed applications are the main causes for the differentiation in the architectural space that we witness today.

The definition of the system platform can be considered the result of a push-and-pull process between the application engineers — who desire independence of the implementation architecture while meeting stringent implementation metrics — and the architecture designers and/or semiconductor companies — who want maximum reuse of their designs. This is illustrated in Figure 1. The position of the equilibrium point depends upon the tightness of the design constraints as well as cost determining factors such as volume and market.

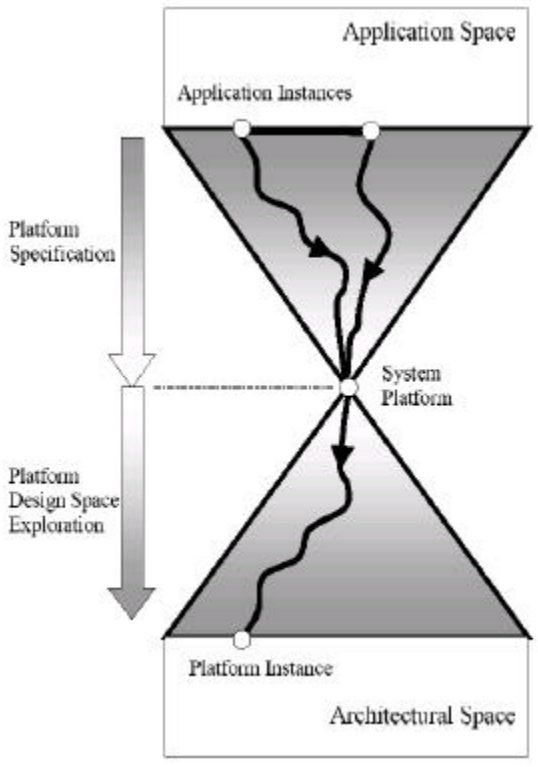

Figure 1 The system pl;atform concept creates a common solution between the constraints of the application and architectural abstractions [ASV99].

III. TRENDS IN WIRELESS SYSTEMS

The world of wireless communications and networking has experienced an unprecedented growth in the last decade. The number of wireless subscribers has grown exponentially between 1990 and 1998, matching remarkably well a Fibonacci sequence as shown in Figure 2 This trend is expected to continue well into the next decade. While most of the volume in the wireless traffic is due to voice communications at present, a shift towards data traffic is slowly but surely emerging, a shift which is expected to materialize in full force with the deployment with the so-called 3rd generation of wireless systems. In fact, the growth of wireless networking closely mirrors and complements the evolution towards fully-distributed and ubiquitous computation systems, already discussed in the introduction of this paper. It is my conjecture that the major growth in wireless systems, after the deployment of a wide-area 3rd generation network, will be in the home, office, and building environment, providing connectivity between the myriad of distributed sensor, information processing, computation, and monitoring devices, as well as providing access to the existing high-bandwidth networks and the internet. Wireless technology is clearly the preferred approach towards solving the **"last-meter" problem** in information access and distribution.

An analysis of the characteristics of these next-generation systems reveals some dominant requirements that have to be met for the technology to be viable:

1. The abundance of wireless services with varying data-rate and quality of service requirements will create a spectrum shortage. This can only be addressed by increasing the spectral efficiency of modems (in bits/sec/Hz/m^3), and by reconsidering the way spectrum is allocated to services — in

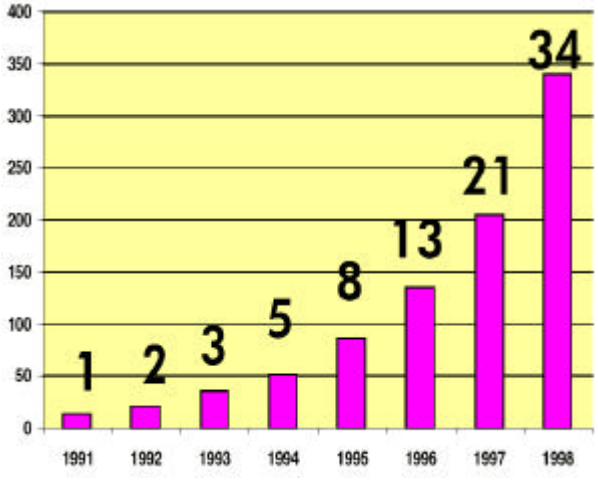

Figure 2 Growth in wireless subscribers over the world (source: Goldman-Sachs).

	Wide-band CDMA			FDMA Multiple Antenna
	Matched Filter	Blind MMSE	Exact Decorrelator	SVD
Performance Bits/sec/Hz	1	2	2	6
Multiplications	124	496	230,000	736
Memory	248	1240	640,000	2120
ALU	124	502	240,000	800
Word Length	8-bit	12-bit	16-bit	16-bit

Figure 3 Computational complexity of some advanced wireless communication algorithms (from [])

other words, the standardization process. Fortunately, advances in communication theory provide us with solutions to both of these problems. Techniques such as multi-antenna diversity, beamforming, and multi-user detection and cancellation can provide dramatic increases in spectral efficiency, while opening the door for a peaceful co-existence of uncoordinated wireless services in the same band without effecting the quality-of-service. This approach, which we have dubbed the universal spectrum-sharing radio, does come at a price in computational complexity. Figure 3 illustrates the computational costs of some of the mentioned algorithms. This growing appetite for computational cycles is confirmed in the picture of Figure 4, which shows the growth in computational complexity over the generations of wireless systems. In fact, the complexity has grown faster than the computational capabilities of silicon, as predicted by Moore's law.

It may hence be stated that the wireless implementation platforms of the future have to provide tremendous real-time computational capabilities. Fortunately, most of the complexity resides in a few structured and regular computational kernels that are easily parallelized and that are amenable to stream-based processing.

2. Achieving the deep penetration of wireless services, as envisioned in the ubiquitous wireless scenario, requires a steep reduction in cost and energy dissipation of the wireless modem. To be successful, modem costs below 5$ and energy consumption below 1 mW have to be targeted. While integration is certainly an important help in that direction, further improvements in both implementation and energy efficiency are needed.

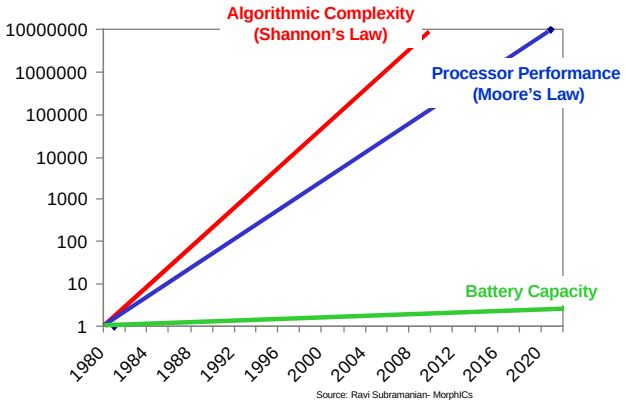

Figure 4 Algorithmic complexity as required by various generations of wireless systems outstrips Moore's law.

In summary, the implementation platform for the wireless system of the future requires a seamless combination of the contradictory requirements of high performance, flexibility, low energy consumption, and low cost.

IV ARCHITECTURES FOR WIRELESS SYSTEMS

The definition of such a platform requires an in-depth understanding of the modem functions and their properties. As shown in Figure 5, a radio receiver (or transmitter) typically combines a data pipe, which gradually transforms the bit-serial data stream coming from the A/D converter into a set of complex data messages, and a protocol stack, that controls the operation of the data pipe. Data-pipe and protocol stack differ in the types of computations to be performed, and in the communication mechanisms between the functional modules. In short, they exhibit fundamentally different models of computation. In addition, the different modules of the data and protocol stacks operate on time and data granularities that vary over a wide range. For instance, the physical layer of the radio manipulates individual bits at rates in the high MHz range, while the application layer processes complex data structures at rates that are in the low Hz range.

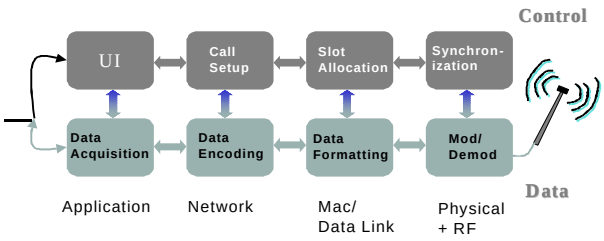

Figure 5 Functional components of wireless transceiver.

Devising a homogeneous architecture that efficiently deals with this range of computational models and data and time granularities is extremely hard, if not impossible. It is hence safe to predict that the architectural platform for wireless application will combine a heterogeneous collection of computational modules, each of which targets one specific computational model and data granularity.

This raises immediately the question of how to select the constituent components that make up this heterogeneous fabric, given the wide range in architectural options. Ultimately, this process boils down to an elaborate trade-off in the flexibility-efficiency-performance-cost space. Only an extensive profiling of the applications or algorithms can help to determine the required bounds on performance and flexibility, or, even more importantly, to outline the dominant computational patterns and the model of computation. While this effort might seem to be excessive, the pay-back can be dramatic, as is illustrated in Figure 6. Based on a number of benchmark studies, this plot compares some architectural alternatives in the energy efficiency/flexibility space. The impact of providing full flexibility (in the style of the Von Neuman machine, exemplified by the embedded microprocessor) is quite staggering, and results in three orders of magnitude in "inefficiency" compared to the fixed custom

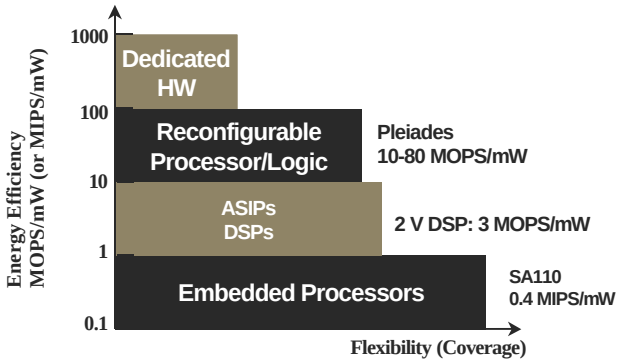

Figure 6 Analysis of energy efficiency versus flexibility trade-off for a number of common architectural styles (for a 0.25 micron CMOS process).

implementation of the same function. Another interesting observation is that the gap between the two can be closed by the introduction of "domain-specific" architectures, such as DSPs and ASIPs (Application-Specific Instruction Processors).

It is fair to state that one of the most interesting and unexpected results of the "system-on-a-chip" has been a renaissance in the field of processor architecture. Rather than being a pure "board-on-a-chip", the tight integration of computation and communication, combined with the redefined metrics of embedded applications, has led to a wide range of new (or revised) architectural approaches that attempt to exploit the specific properties of a limited application domain. Examples of these are the configurable instruction-set processor, the embedded very-long instruction set processor (VLIW), the very-short instruction set processor (or the vector processor), and the reconfigurable processor.

The latter is perhaps the most innovative entry in the computer architecture field in the last decade. Rather than the "multiplexing-in-time" concept, advocated in the instruction-set processor, the reconfigurable processor relies on a "programming-in-space" approach. Using programmable wiring, parametrizable functional units are connected to form a dedicated computing engine. For tasks that are semi-stationary in time, this approach has the advantage of combining flexibility and efficiency. Various approaches have been published addressing different levels of data, operator, and reconfiguration rate granularity. Figure 7 shows an abstract model of the Berkeley Pleiades approach, which belongs to the architectural class of task-based reconfiguration at the arithmetic operator level. It combines a traditional

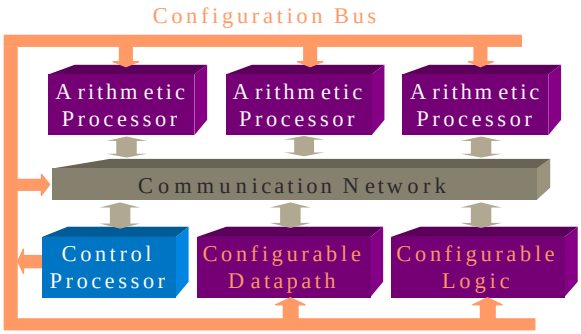

Figure 7 Berkeley Pleiades Reconfigurable Architecture

embedded processor with an array of heterogeneous computational modules, connected by a reconfigurable communication network. The approach efficiently combines two models of computation — communicating processes and data flow — with the latter being implemented on the co-processor array. As shown in Figure 6, the reconfigurable approach effectively succeeds in bridging the gap between the DSP and the custom design approach in terms of energy efficiency - by trading in some flexibility.

The advent of the SOC has furthermore focused the spotlight on the importance of a well-founded interconnect strategy. Most often, all interest in the design process is focused on the selection and conception of the computational elements. How to connect them together might be just as important. The traditional bus approach tends to be inefficient and unpredictable, definitely in the presence of real-time constraints. THe "communication-based design" methodology might very well be one of the most important breakthroughs in design-methodology for SOC.

V. Summary

In this paper, we have collected a number of observations regarding the needs of the next-generation wireless systems in light of the emerging system-on-a-chip approach. Based on this analysis, we venture to make a projection on a plausible implementation platform for the wireless modem of the future, which is shown in Figure 8. While its computational core is formed by a group of embedded processors (general purpose and/or DSP), all data and control intensive functions will be performed by an array of configurable data and control processors, each of which is optimized for a particular model of computation, and a specific data and time granularity. The analog component are kept as simple as possible with digital processing compensating for the resulting imperfections.

References

[ASV99] A. Ferrari and A. Sangiovanni-Vincentelli, "System Design: Traditional Concepts and New Paradigms", Proceedings ICCD 1999.

[Rab97] J. Rabaey et al, "Heterogeneous Reconfigurable Systems", Proceeding VLSI Signal Processing Workshop, November 1997, Manchester, England.

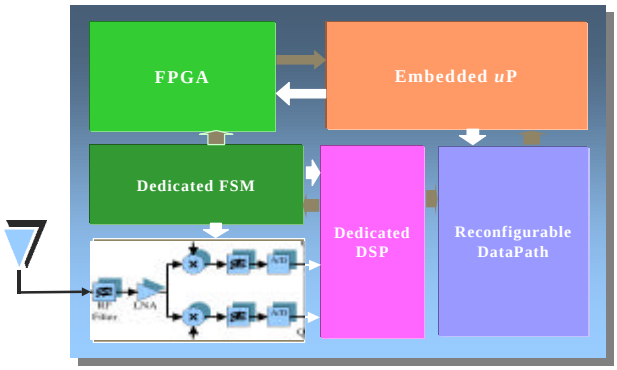

Figure 8 Plausible implementation fabric for "software-enabled" radio.