1 INTRODUCTION

Throughout the history of computing, digital signal processing (DSP) applications have pushed the limits of computer power, especially in terms of real-time computation. Although processed signals have broadly ranged from media-driven speech, audio, and video waveforms to specialized radar and sonar data, most calculations performed by signal processing systems have exhibited the same basic computational characteristics. The inherent data parallelism found in many DSP functions has made DSP algorithms ideal candidates for hardware implementation, leveraging expanding VLSI (very-large-scale integration) capabilities. Recently, DSP has received increased attention due to rapid advancements in multimedia computing and high-speed wired and wireless communications. In response to these advances, the search for novel implementations of arithmetic-intensive circuitry has intensified.

Although application areas span a broad spectrum, the basic computational parameters of most DSP operations remain the same: a need for real-time performance within the given operational parameters of a target system and, in most cases, a need to adapt to changing datasets and computing conditions. In general, the goal of high performance in systems ranging from low-cost embedded radio components to special-purpose ground-based radar centers has driven the development of application and domain-specific chip sets. The development and financial cost of this approach is often large, motivating the need for new ap-
proaches to computer architecture that offer the same computational attributes as fixed-functionality architectures in a package that can be customized in the field. The second goal of system adaptability is generally addressed through the use of software-programmable, commodity digital signal processors. Although these platforms enable flexible deployment due to software development tools and great economies of scale, application designers and compilers must customize their processing approach to available computing resources. This flexibility often comes at the cost of performance and power efficiency.

As shown in Figure 1, reconfigurable computers offer a compromise between the performance advantages of fixed-functionality hardware and the flexibility of software-programmable substrates. Like application-specific integrated circuits (ASICs), these systems are distinguished by their ability to directly implement specialized circuitry directly in hardware. Additionally, like programmable processors, reconfigurable computers contain functional resources that may be modified easily after field deployment in response to changing operational parameters and datasets. To date, the core processing element of most reconfigurable computers has been the field programmable gate array (FPGA). These bit-programmable computing devices offer ample quantities of logic and register resources that can easily be adapted to support the fine-grained parallelism of many pipelined DSP applications. With current logic capacities exceeding 1 million gates per device, substantial logic functionality can be implemented on each programmable device. Although appropriate for some classes of implementation,
FPGAs represent only one possible implementation in a range of possible reconfigurable computing building blocks. A number of reconfigurable alternatives are presently under evaluation in academic and commercial environments.

In this survey, the evolution of reconfigurable computing with regard to digital signal processing is considered. This study includes an historical evaluation of reprogrammable architectures and programming environments used to support DSP applications. The chronology is supported with specific case studies which illustrate approaches used to address implementation constraints such as system cost, performance, and power consumption. It is seen that as technology has progressed, the richness of applications supported by reconfigurable computing and the performance of reconfigurable computing platforms have improved dramatically. Reconfigurable computing for DSP remains an active area of research as the need for integration with more traditional DSP technologies such as PDSPs becomes apparent and the goal of automated high-level compilation for DSP increases in importance.

The organization of this chapter is as follows. In Section 2, a brief history of the issues and techniques involved in the design and implementation of DSP systems is described. Section 3 presents a short history of reconfigurable computing. Section 4 describes why reconfigurable computing is a promising approach for DSP systems. Section 5 serves as the centerpiece of the chapter and provides a history of the application of various reconfigurable computing technologies to DSP systems and a discussion of the current state of the art. We conclude in Section 6 with some predictions about the future of reconfigurable computing for digital signal processing. These predictions are formulated by extrapolating the trends of reconfigurable technologies and describing future DSP applications which may be targeted to reconfigurable hardware.

### 1.1 Definitions

The following definitions are used to describe various attributes related to reconfigurable computing:

- **Reconfigurable or adaptive**: In the context of reconfigurable computing, this term indicates that the logic functionality and interconnect of a computing system or device can be customized to suit a specific application through postfabrication, user-defined programming.
- **Run-time (or dynamically) reconfigurable**: System logic functionality and/or interconnect connectivity can be modified during application execution. This modification may be either data driven or statically scheduled.
- **Fine-grained parallelism**: Logic functionality and interconnect connectivity is programmable at the bit level. Resources encompassing multiple logic bits may be combined to form parallel functional units.
Specialization: Logic functionality can be customized to perform exactly the operation desired. An example is the synthesis of filtering hardware with a fixed constant value.

2 BACKGROUND IN DSP IMPLEMENTATION

2.1 DSP System Implementation Choices

Since the early 1960s, three goals have driven the development of DSP implementations: (1) data parallelism, (2) application-specific specialization, and (3) functional flexibility. In general, design decisions regarding DSP system implementation require trade-offs between these three system goals. As a result, a wide variety of specialized hardware implementations and associated design tools have been developed for DSP, including associative processing, bit-serial processing, on-line arithmetic, and systolic processing. As implementation technologies have become available, these basic approaches have matured to meet the needs of application designers.

As shown in Table 1, various cost metrics have been developed to compare the quality of different DSP implementations. Performance has frequently been the most critical system requirement because DSP systems often have demanding real-time constraints. In the past two decades, however, cost has become more significant as DSP has migrated from predominantly military and scientific applications into numerous low-cost consumer applications. Over the past 10 years, energy consumption has become an important measure as DSP techniques have been widely applied in portable, battery-operated systems such as cell phones, CD players, and laptops [1]. Finally, flexibility has emerged as one of the key differentiators in DSP implementations because it allows changes to system functionality at various points in the design life cycle. The results of

<table>
<thead>
<tr>
<th>Table 1</th>
<th>DSP Implementation Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Performance</td>
</tr>
<tr>
<td>ASIC</td>
<td>High</td>
</tr>
<tr>
<td>Programmable DSP</td>
<td>Medium</td>
</tr>
<tr>
<td>General-purpose processor</td>
<td>Low</td>
</tr>
<tr>
<td>Reconfigurable hardware</td>
<td>Medium</td>
</tr>
</tbody>
</table>
these cost trade-offs have resulted in four primary implementation options, including ASICs, programmable digital signal processors (PDSPs), general-purpose microprocessors, and reconfigurable hardware. Each implementation option presents different trade-offs in terms of performance, cost, power, and flexibility.

For many specialized DSP applications, system implementation must include one or more ASICs to meet performance and power constraints. Even though ASIC design cycles remain long, a trend toward automated synthesis and verification tools [2] is simplifying high-level ASIC design. Because most ASIC specification is done at the behavioral or register-transfer level, the functionality and performance of ASICs have become easier to represent and verify. Another, perhaps more important, trend has been the use of predesigned cores with well-defined functionality. Some of these cores are, in fact, PDSPs or reduced instruction set computer (RISC) microcontrollers, for which software has to be written and then stored on-chip. ASICs have a significant advantage in area and power, and for many high-volume designs, the cost-per-gate for a given performance level is less than that of high-speed commodity FPGAs. These characteristics are especially important for power-aware functions in mobile communication and remote sensing. Unfortunately, the fixed nature of ASICs limits their reconfigurability. For designs that must adapt to changing datasets and operating conditions, software-programmable components must be included in the target system, reducing available parallelism. Additionally, for low-volume or prototype implementations, the nonrecurring engineering (NRE) costs related to an ASIC may not justify its improved performance benefits.

The application domain of PDSPs can be identified by tracing their development lineage. Thorough summaries of programmable DSPs can be found in Refs. 3–5. In the 1980s, the first PDSPs were introduced by Texas Instruments. These initial processor architectures were primarily CISC (complex-instruction-set computer) pipelines augmented with a handful of special architectural features and instructions to support filtering and transform computations. One of the most significant changes to second-generation PDSPs was the adaptation of the Harvard architecture, effectively separating the program bus from the data bus. This optimization reduced the von Neumann bottleneck, thus providing an unimpeded path for data from local memory to the processor pipeline. Many early DSPs allowed programs to be stored in on-chip ROM and supported the ability to make off-chip accesses if instruction capacity was exceeded. Some DSPs also had coefficient ROMs, again recognizing the opportunity to exploit the relatively static nature of filter and transform coefficients.

Contemporary digital signal processors are highly programmable resources that offer the capability for in-field update as processing standards change. Parallelism in most PDSPs is not extensive but generally consists of overlapped data
fetch, data operation, and address calculation. Some instruction set modifications are also used in PDSPs to specialize for signal processing. Addressing modes are provided to simplify the implementation of filters and transforms and, in general, control overhead for loops is minimized. Arithmetic instructions for fixed-point computation allow saturating arithmetic, which is important for avoiding overflow exceptions or oscillations. New hybrid DSPs contain a variety of processing and input/output (I/O) features, including parallel processing interfaces, very-long-instruction-word (VLIW) function unit scheduling, and flexible datapaths. Through the addition of numerous, special-purpose memories, on-chip DSPs can now achieve high-bandwidth and, to a moderate extent, reconfigurable interconnect. Due to the volume usage of these parts, costs are reduced and commonly used interfaces can be included. In addition to these benefits, the use of a DSP has specific limitations. In general, for optimal performance, applications must be written to utilize the resources available in the DSP. Although high-level compilation systems which perform this function are becoming available [6,7], often it is difficult to get exactly the mapping desired. Additionally, the interface to memory may not be appropriate for specific applications, creating a bandwidth bottleneck in getting data to functional units.

The 1990s have been characterized by the introduction of DSP to the mass commercial market. DSP has made the transition from a fairly academic acronym to one seen widely in advertisements for consumer electronics and software packages. A battle over the DSP market has ensued primarily between PDSP manufacturers, ASIC vendors, and developers of two types of general-purpose processor, desktop microprocessors and high-end microcontrollers. General-purpose processors, such as the Intel Pentium, can provide much of the signal processing needed for desktop applications such as audio and video processing, especially because the host microprocessor is already resident in the system and has highly optimized I/O and extensive software development tools. However, general-purpose desktop processors are not a realistic alternative for embedded systems due to their cost and lack of power efficiency in implementing DSP. Another category of general-purpose processors is the high-end microcontroller. These chips have also made inroads into DSP applications by presenting system designers with straightforward implementation solutions that have useful data interfaces and significant application-level flexibility.

One DSP hardware implementation compromise that has developed recently has been the development of domain-specific standard products in both programmable and ASIC formats. The PDSP community has determined that because certain applications have a high volume, it is worthwhile to tailor particular PDSPs to domain-specific markets. This has led to the availability of inexpensive, commodity silicon while allowing users to provide application differentiation in software. ASICs have also been developed for more general functions
like MPEG decoding, in which standards have been set up to allow a large number of applications to use the same basic function.

Reconfigurable computing platforms for DSP offer an intermediate solution to ASICs, PDSPs, and general and domain-specific processors by allowing reconfigurable and specialized performance on a per-application basis. Although this emerging technology has primarily been applied to experimental rather than commercial systems, the application-level potential for these reconfigurable platforms is great. Following an examination of the needs of contemporary DSP applications, current trends in the application of reconfigurable computing to DSP are explored.

2.2 The Changing World of DSP Applications

Over the past 30 years, the application space of digital signal processing has changed substantially, motivating new systems in the area of reconfigurable computing. New applications over this time span have changed the definition of DSP and have created new and different requirements for implementation. For example, in today’s market, DSP is often found in human–computer interfaces such as sound cards, video cards, and speech recognition systems—application areas with limited practical significance just a decade ago. Because a human is an integral part of these systems, different processing requirements can be found, in contrast to communications front ends such as those found in DSL modems from Broadcom [8] or CDMA (code division multiple access) receiver chips from Qualcomm [9]. Another large recent application of DSP has been in the read circuitry of hard-drive and CD/DVD storage systems [10]. Although many of the DSP algorithms are the same as in modems, the system constraints are quite different.

Consumer products now make extensive use of DSP in low-cost and low-power implementations [11]. Both wireless and multimedia, two of the hottest topics in consumer electronics, rely heavily on DSP implementation. Cellular telephones, both GSM (global system for mobile communication) and CDMA, are currently largely enabled by custom silicon [12], although trends toward other implementation media such as PDSPs are growing. Modems for DSL, cable, local area networks (LANs), and, most recently, wireless all rely on sophisticated adaptive equalizers and receivers. Satellite set-top boxes rely on DSP for satellite reception using channel decoding as well as an MPEG decoder ASIC for video decompression. After the set-top box, the DVD player has now emerged as the fastest-growing consumer electronics product. The DVD player relies on DSP to avoid intersymbol interference, allowing more bits to be packed into a given area of disk. In the commercial video market, digital cameras and camcorders are rapidly becoming affordable alternatives to traditional analog cameras, largely supported by photo-editing, authoring software, and the Web.
Development of a large set of DSP systems has been driven indirectly by the growth of consumer electronics. These systems include switching stations for cellular, terrestrial, satellite and cable infrastructure as well as cameras, authoring studios, and encoders used for content production. New military and scientific applications applied to the digital battlefield, including advanced weapons systems and remote sensing equipment, all rely on DSP implementation that must operate reliably in adverse and resource-limited environments. Although existing DSP implementation choices are suitable for all of these consumer and military-driven applications, higher performance, efficiency, and flexibility will be needed in the future, driving current interest in reconfigurable solutions.

In all of these applications, data processing is considerably more sophisticated than the traditional filters and transforms which characterized DSP of the 1960s and 1970s. In general, performance has grown in importance as data rates have increased and algorithms have become more complex. Additionally, there is an increasing demand for flexible and diverse functionality based on environmental conditions and workloads. Power and cost are equally important because they are critical to overall system cost and performance.

Although new approaches to application-specific DSP implementation have been developed by the research community in recent years, their application in practice has been limited by the market domination of PDSPs and the reluctance of designers to expose schedule and risk-sensitive ASIC projects to nontraditional design approaches. Recently, however, the combination of new design tools and the increasing use of intellectual property cores [13] in DSP implementations have allowed some of these ideas to find wider use. These implementation choices include systolic architectures, alternative arithmetic (residue number system [RNS], logarithmic number system [LNS], digital-serial), word-length optimization, parallelizing transformations, memory partitioning, and power optimization techniques. Design tools have also been proposed which could close the gap between software development and hardware development for future hybrid DSP implementations. In subsequent sections, it will be seen that these tools will be helpful in defining the appropriate application of reconfigurable hardware to existing challenges in DSP. In many cases, basic design techniques used to develop ASICs or domain-specific devices can be reapplied to customize applications in programmable silicon by taking the limitations of the implementation technology into account.

3 A BRIEF HISTORY OF RECONFIGURABLE COMPUTING

Since their introduction in the mid-1980s, field programmable gate arrays (FPGAs) have been the subject of extensive research and experimentation. In this section, reconfigurable device architecture and system integration is investi-
gated with an eye toward identifying trends likely to affect future development. Although this summary provides sufficient background to evaluate the impact of reconfigurable hardware on DSP, more thorough discussions of FPGAs and reconfigurable computing can be found in Refs. 14–17.

3.1 Field Programmable Devices

The concept of a digital hardware device which supports programmable logic was originated in the early 1960s with the introduction of cellular arrays. These devices contained built-in logic structures whose functionality could be set either in the final stages of production or in the field. Early cellular arrays, such as the Maitra cascade [18], contained extremely simple logic cells and supported linear, near-neighbor interblock connectivity. Each cell could generally perform a single-output Boolean function of two inputs which was determined through a programmable mask set late in the device fabrication process. Field programmable technology became a reality in the mid-1960s with the introduction of cutpoint cellular logic [19]. Like Maitra cascades, these devices contained a fixed interconnection between cells, but the logic functionality of each cell could be programmed in the field. Customization was typically accomplished by blowing programmable cell fuses through the use of programming currents or photoconductive exposure [19]. A direct forerunner of today’s SRAM-based FPGA was a programmable array proposed and implemented by Wahlstrom [20] in 1967. Like today’s FPGA devices, the operation of each logic cell was controlled by a user-defined bit stream which determined both internal logic functionality and connectivity to adjacent intercell wires and buses. The array could be reprogrammed to implement a variety of logic circuits and to accommodate in-field operational faults. Extensions and analysis of Wahlstrom’s array were later documented in Ref. 21.

The modern era of reconfigurable computing was ushered in by the introduction of the first commercial SRAM-based FPGAs by Xilinx Corporation [22] in 1986. These early reprogrammable devices and subsequent offerings from both Xilinx and Altera Corporation [23] contain a collection of fine-grained programmable logic blocks interconnected via wires and programmable switches. Logic functionality for each block is specified via a small programmable memory, called a look-up table, driven by a limited number of inputs (typically less than five) which generates a single Boolean output. Additionally, each logic block typically contains one or more flip-flops for fine-grained storage. Although early FPGA architectures contained small numbers of logic blocks (typically less than 100), new device families have quickly grown to capacities of tens of thousands of look-up tables containing millions of gates of logic. As shown in Figure 2, fine-grained look-up table/flip-flop pairs are frequently grouped into tightly connected coarse-grained blocks to take advantage of circuit locality. Interconnection be-
Figure 2  Simplified Xilinx Virtex logic block. Each logic block consists of two 2-LUT (look-up table) slices. (From Ref. 26.)
between logic blocks is provided via a series of wire segments located in channels between the blocks. Programmable pass transistors and multiplexers can be used to provide both block-to-segment connectivity and segment-to-segment connections.

Much of the recent interest in reconfigurable computing has been spurred by the development and maturation of field programmable gate arrays. The recent development of systems based on FPGAs has been greatly enhanced by an exponential growth rate in the gate capacity of reconfigurable devices and improved device performance due to shrinking die sizes and enhanced fabrication techniques. As shown in Figure 3, reported gate counts [24–26] for look-up table (LUT)-based FPGAs, from companies such as Xilinx Corporation, have roughly followed Moore’s law over the past decade.* This increase in capacity has enabled complex structures such as multitap filters and small RISC processors to be implemented directly in a single FPGA chip. Over this same time period, the system performance of these devices has also improved exponentially. Whereas in the mid-1980s, system-level FPGA performance of 2–5 MHz was considered acceptable, today’s LUT-based FPGA designs frequently approach performance

* In practice, usable gate counts for devices are often significantly lower than reported data book values (by about 20–40%). Generally, the proportion of per-device logic that is usable has remained roughly constant over the years, as indicated in Figure 3.
levels of 60 MHz and beyond. Given the programmable nature of reconfigurable devices, the performance penalty of a circuit implemented in reprogrammable technology versus a direct ASIC implementation is generally a factor on the order of 5 to 10.

3.2 Early Reprogrammable Systems

The concept of using reprogrammable logic to enhance the functional capabilities of a computing system is generally credited to Gerald Estrin [27]. In a feasibility study performed in the early 1960s, a digital system is described that contains both a sequential processor and a programmable logic core which can change logic functionality on a per-application basis. Even though a functioning hardware system based on the concept was not built, the study outlined the potential of application-level specialization of system hardware. Estrin’s work motivated the later analysis of the use of cellular arrays for basic-block-level computation [28]. In this subsequent study, the potential of reconfigurability for use in design verification and algorithm development is addressed, setting the stage for contemporary multi-FPGA prototyping and development platforms.

Soon after the commercial introduction of the FPGA, computer architects began devising approaches for leveraging new programmable technology in computing systems. As summarized in Ref. 16, the evolution of reconfigurable computing was significantly shaped by two influential projects: Splash II [29] and Programmable Active Memories (PAM) [30]. Each of these projects addressed important programmable system issues regarding programming environment, user interface, and configuration management by applying pre-existing computational models in the areas of special-purpose coprocessing and statically scheduled communication to reconfigurable computing.

Splash II is a multi-FPGA parallel computer which uses orchestrated systolic communication to perform inter-FPGA data transfer. As shown in Figure 4, each board of multiboard Splash II systems contains 16 Xilinx XC4000 series FPGA processors (labeled with an X prefix), each with associated SRAM (labeled with an M prefix). Unlike its multi-FPGA predecessor, Splash [31], which was limited to strictly near-neighbor systolic communication, each Splash II board contains inter-FPGA crossbars for multihop data transfer and broadcast. Software development for the system typically involves the creation of VHDL (VHSIC hardware description language) circuit descriptions for individual systolic processors. These designs must meet size and performance constraints of the target FPGAs. Following processor creation, high-level inter-FPGA scheduling software is used to ensure that systemwide communication is synchronized. In general, the system is not dynamically reconfigured during operation. For applications with single instruction multiple data (SIMD) characteristics, a compiler [32] has been created to automatically partition processing across FPGAs and to synchronize interfaces to
local SRAMs. Numerous DSP applications have been mapped to Splash II, including audio and video algorithm implementations. These applications are described in greater detail in Section 5. Recently, FPGA-based systolic architectures based on the Splash II system have been developed by Annapolis Micro Systems [33]. The company’s peripheral component interface (PCI) based Wildforce system contains five Xilinx XC4000XL devices aligned in a systolic chain. A similar, VME-based Wildstar board contains four Xilinx Virtex devices.

As shown in Figure 5, Programmable active memory DECPeRLe-1 system [30] contain arrangements of FPGA processors (labeled X) in a two-dimensional mesh with memory devices (labeled M) aligned along the array perimeter. PAMs were designed to create the architectural appearance of a functional memory for a host microprocessor and the PAM programming environment reflects this. From a programming standpoint, the multi-FPGA PAM can be accessed like a memory through an interface FPGA, XI, with written values treated as inputs and read values used as results. Designs are generally targeted to PAMs through handcrafting of design subtasks, each appropriately sized to fit on an FPGA. The PAM
array and its successor, the Pamette [34], are interfaced to a host workstation through a backplane bus. Additional discussion of PAMs with regard to DSP applications appears in Section 5.

3.3 Reconfigurable Computing Research Directions

Over the past decade, interest in reconfigurable systems has progressed along four main paths [15]:

1. The proximity of reconfigurable hardware to a host CPU
2. The capability of hardware to support dynamic reconfiguration
3. Software support for high-level compilation and dynamic reconfiguration
4. The granularity of reconfigurable elements

Active research in these areas continues today in addition to a search for applications well-suited to the available architectural parameters.

As a result of the Prism I project [35], the first reconfigurable system which tightly coupled an off-the-shelf processor with an FPGA coprocessor was created. This project explored the possibility of augmenting the instruction set of a processor with special-purpose instructions that could be executed by an attached FPGA coprocessor in place of numerous processor instructions. For these instructions, the microprocessor would stall for several cycles while the FPGA-based coprocessor completed execution. More recently, the single-chip Napa [36] and OneChip [37] architectures have used similar approaches to synchronize pro-
cessing between RISC processors and FPGA cores. As chip integration levels have increased, interest in tightly coupling both processor and reconfigurable resources at multiple architectural levels has grown. Single-chip architectures, such as Garp [38], now allow interfacing between processors and reconfigurable resources, both through coprocessor interfaces and through a shared data cache. A second approach to integrating reconfigurable logic and microprocessors has explored integrating reconfigurable logic inside the processor as special-purpose functional units. Although early approaches in this area attempted to keep reconfigurable functional unit timing consistent with other nonconfigurable resources [39], newer reconfigurable functional units [40] allow multicycle operation synchronized by the microprocessor control path.

An important aspect of reconfigurable devices is the ability to reconfigure functionality in response to changing operating conditions and application datasets. Although SRAM-based FPGAs have supported slow millisecond reconfiguration rates for some time, only recently have devices been created that allow for rapid device reconfiguration at run time. Dynamically reconfigurable FPGAs, or DPGAs [41,42], contain multiple interconnect and logic configurations for each programmable location in a reconfigurable device. Often these architectures are designed to allow configuration switching in a small number of system clock cycles, measuring nanoseconds rather than milliseconds. Although several DPGA devices have been developed in research environments, only one has been developed commercially. The Context Switching FPGA [43], developed commercially by Sanders Corporation, can simultaneously hold up to four complete configuration contexts. A context switch for the device can be performed in a single clock cycle. During the context switch, all internal data stored in registers are preserved. To promote reconfiguration at lower hardware cost, several commercial FPGA families [26,44] have been introduced that allow for fast, partial reconfiguration of FPGA functionality from off-chip memory resources. A significant challenge to the use of these reconfigurables is the development of compilation software which will partition and schedule the order in which computation will take place and will determine which circuitry must be changed. Although some preliminary work in this area has been completed [45,46], more advanced tools are needed to fully leverage the new hardware technology. Other software approaches that have been applied to dynamic reconfiguration include the definition of hardware subroutines [47] and the dynamic reconfiguration of instruction sets [48].

Although high-level compilation for microprocessors has been an active research area for decades, development of compilation technology for reconfigurable computing is still in its infancy. The compilation process for FPGA-based system is often complicated by a lack of identifiable coarse-grained structure in fine-grained FPGAs and the dispersal of logic resources across many pin-limited reconfigurable devices on a single computing platform. In particular, because most reconfigurable computers contain multiple programmable devices, design
**partitioning** forms an important aspect of most compilation systems. Several compilation systems for reconfigurable hardware [49,50] have followed a traditional multidevice ASIC design flow involving pin-constrained device partitioning and individual device synthesis using RTL compilation. To overcome pin limitations and achieve full logic utilization on a per-device basis using this approach, either excessive internal device interconnect [49] or I/O counts [51] have been needed. In Ref. 52, a hardware virtualization approach is outlined that promotes high per-device logic utilization. Following design partitioning and placement, inter-FPGA wires are scheduled on interdevice wires at compiler-determined time slices, allowing pipelining of communication. Interdevice pipelining also forms the basis of several FPGA system compilation approaches that start at the behavioral level. A high-level synthesis technique described in Ref. 53 outlines inter-FPGA scheduling at the RTL level. In Refs. 54 and 55, functional allocation is performed that takes into account the amount of logic available in the target system and available interdevice interconnect. Combined communication and functional resource scheduling is then performed to fully utilize available logic and communication resources. In Ref. 56, inter-FPGA communication and FPGA-memory communication are virtualized because it is recognized that memory rather than inter-FPGA bandwidth is frequently the critical resource in reconfigurable systems. In Ref. 57, linear programming is used to partition MATLAB functions across sets of heterogeneous resources, including DSPs, RISC processors, and FPGAs. Scheduling, pipelining, and component-specific compilation are performed following partitioning to complete the mapping process.

4 THE PROMISE OF RECONFIGURABLE COMPUTING FOR DSP

Many of the motivations and goals of reconfigurable computing are consistent with the needs of signal processing applications. It will be seen in Section 5 that the deployment of DSP algorithms on reconfigurable hardware has aided in the advancement of both fields over the past 15 years. In general, the direct benefits of the reconfigurable approach for DSP can be summarized in three critical areas: functional specialization, platform reconfigurability, and fine-grained parallelism.

4.1 Specialization

As stated in Section 2.1, programmable digital signal processors are optimized to deliver efficient performance across a set of signal processing tasks. Although the specific implementation of tasks can be modified through instruction-configurable software, applications must frequently be customized to meet specific processor architectural aspects, often at the cost of performance. Currently,
most DSPs remain inherently sequential machines, although some parallel VLIW and multifunction unit DSPs have recently been developed [58]. The use of reconfigurable hardware has numerous advantages for many signal processing systems. For many applications, such as digital filtering, it is possible to customize irregular datapath widths and specific constant values directly in hardware, reducing implementation area and power and improving algorithm performance. Additionally, if standards change, the modifications can quickly be reimplemented in hardware without expensive NRE costs. Because reconfigurable devices contain SRAM-controlled logic and interconnect switches, application programs in the form of device configuration data can be downloaded on a per-application basis. Effectively, this single, wide program instruction defines hardware behavior. Contemporary reconfigurable computing devices have little or no NRE cost because off-the-shelf development tools are used for design synthesis and layout. Although reconfigurable implementations may exhibit a 5–10 times performance reduction compared to the same circuit implemented in custom logic, limited manual intervention is generally needed to map a design to a reconfigurable device. In contrast, substantial NRE costs require ASIC designers to focus on high-speed physical implementation often involving hand-tuned physical layout and near-exhaustive design verification. Time-consuming ASIC implementation tasks can also lead to longer time-to-market windows and increased inventory, effectively becoming the critical path link in the system design chain.

4.2 Reconfigurability

Most reconfigurable devices and systems contain SRAM-programmable memory to allow full logic and interconnect reconfiguration in the field. Despite a wide range of system characteristics, most DSP systems have a need for configurability under a variety of constraints. These constraints include environmental factors such as changes in statistics of signals and noise, channel, weather, transmission rates, and communication standards. Although factors such as data traffic and interference often change quite rapidly, other factors such as location and weather change relatively slowly. Still other factors regarding communication standards vary infrequently across time and geography, limiting the need for rapid reconfiguration. Some specific ways that DSP can directly benefit from hardware reconfiguration to support these factors include the following:

- **Field customization:** The reconfigurability of programmable devices allows periodic updates of product functionality as advanced vendor firmware versions become available or product defects are detected. Field customization is particularly important in the face of changing standards and communication protocols. Unlike ASIC implementations, reconfigurable hardware solutions can generally be quickly updated
based on application demands without the need for manual field upgrades or hardware swaps.

- **Slow adaptation**: Signal processing systems based on reconfigurable logic may need to be periodically updated in the course of daily operation based on a variety of constraints. These include issues such as variable weather and operating parameters for mobile communication and support for multiple, time-varying standards in stationary receivers.

- **Fast adaptation**: Many communication processing protocols [59] require nearly constant re-evaluation of operating parameters and can benefit from rapid adjustment of computing parameters. Some of these issues include adaptation to time-varying noise in communication channels, adaptation to network congestion in network configurations, and speculative computation based on changing datasets.

### 4.3 Parallelism

An abundance of programmable logic facilitates the creation of numerous functional units directly in hardware. Many characteristics of FPGA devices, in particular, make them especially attractive for use in digital signal processing systems. The fine-grained parallelism found in these devices is well matched to the high sample rates and distributed computation often required of signal processing applications in areas such as image, audio, and speech processing. Plentiful FPGA flip-flops and a desire to achieve accelerated system clock rates have led designers to focus on heavily pipelined implementations of functional blocks and interblock communication. Given the highly pipelined and parallel nature of many DSP tasks, such as image and speech processing, these implementations have exhibited substantially better performance than standard PDSPs. In general, these systems have been implemented using both task and functional unit pipelining. Many DSP systems have featured bit-serial functional unit implementations [60] and systolic interunit communication [29] that can take advantage of the synchronization resources of contemporary FPGAs without the need for software instruction fetch and decode circuitry. As detailed in Section 5, bit-serial implementations have been particularly attractive due to their reduced implementation area. However, as reconfigurable devices increase in size, more nibble-serial and parallel implementations of functional units have emerged in an effort to take advantage of data parallelism.

Recent additions to reconfigurable architectures have aided their suitability for signal processing. Several recent architectures [26,61] have included 2–4-kbit SRAM banks that can be used to store small amounts of intermediate data. This allows for parallel access to data for distributed computation. Another important addition to reconfigurable architectures has been the capability to rapidly change only small portions of device configuration without disturbing existing
device behavior. This feature has recently been leveraged to help adapt signal processing systems to reduce power [62]. The speed of adaptation may vary depending on the specific signal processing application area.

5  HISTORY OF RECONFIGURABLE COMPUTING AND DSP

Since the appearance of the first reconfigurable computing systems, DSP applications have served as important test cases in reconfigurable architecture and software development. In this section, a wide range of DSP design approaches and applications that have been mapped to functioning reconfigurable computing systems are considered. Unless otherwise stated, the design of complete DSP systems is stressed, including I/O, memory interfacing, high-level compilation, and real-time issues rather than the mapping of individual benchmark circuits. For this reason, a large number of FPGA implementations of basic DSP functions like filters and transforms that have not been implemented directly in system hardware have been omitted. Although our consideration of the history of DSP and reconfigurable computing is roughly chronological, some noted recent trends were initially investigated a number of years ago. To trace these trends, recent advancements are directly contrasted with early contributions.

5.1  FPGA Implementation of Arithmetic

Soon after the introduction of the FPGA in the mid-1980s, an interest developed in using the devices for DSP, especially for digital filtering which can take advantage of specialized constants embedded in hardware. Because a large portion of most filtering approaches involves the use of multiplication, efficient multiplier implementations in both fixed and floating points were of particular interest. Many early FPGA multiplier implementations used circuit structures adapted from the early days of large-scale integration (LSI) development and reflected the restricted circuit area available in initial FPGA devices [55]. As FPGA capacities have increased, the diversity of multiplier implementations has grown.

Since the introduction of the FPGA, bit-serial arithmetic has been used extensively to implement FPGA multiplication. As shown in Figure 6, taken from [Ref. 55], bit-serial multiplication is implemented using a linear systolic array that is well suited to the fine-grained nature of FPGAs. Two data values are input into the multiplier, including a parallel value in which all bits are input simultaneously and a sequential value in which values are input serially. In general, a data sampling rate of one value every $M$ clock cycles can be supported, where $M$ is the input word length. Each cell in the systolic array is typically implemented using one to four logic blocks similar to the one shown in
Figure 2. Bit-serial approaches have the advantage that communication demands are independent of word length. As a result, low-capacity FPGAs can efficiently implement them. Given their pipelined nature, bit-serial multipliers implemented in FPGAs typically possess excellent area–time products. Many bit-serial formulations have been applied to finite impulse response filtering [63]. Special-purpose bit-serial implementations have included the canonic signed digit [64] and the power-of-2 sum or difference [65].

Given the dual use of look-up tables as small memories, distributed arithmetic (DA) has also been an effective implementation choice for LUT-based FPGAs. Because it is possible to group multiple LUTs together into a larger fanout memory, large LUTs for DA can easily be created. In general, distributed arithmetic requires the embedding of a fixed-input constant value in hardware, thus allowing the efficient precomputation of all possible dot-product outputs. An example of a distributed arithmetic multiplier, taken from Ref. 55, appears in Figure 7. It can be seen that a fast adder can be used to sum partial products based on nibble look-up. In some cases, it may be effective to implement the LUTs as RAMs so that new constants can be written during execution of the program.

To promote improved performance, several parallel arithmetic implementations on FPGAs have been formulated [55]. In general, parallel multipliers implemented in LUT-based FPGAs achieve a speedup of sixfold in performance when compared to their bit-serial counterparts with an area penalty of 2.5-fold. Specific parallel implementations of multipliers include a carry-save implementation [66], a systolic array with cordic arithmetic [67], and pipelined parallel [63,68,69].

As FPGA system development has intensified, more interest has been given to upgrading the accuracy of calculation performed in FPGAs, particularly through the use of floating-point arithmetic. In general, floating-point operations are difficult to implement in FPGAs due to the complexity of implementation.
and the amount of hardware needed to achieve desired results. For applications requiring extended precision, floating point is a necessity. In Ref. 70, an initial attempt was made to develop basic floating-point approaches for FPGAs that met IEEE-754 standards for addition and multiplication. Area and performance were considered for various FPGA implementations, including shift-and-add, carry-save, and combinational multiplier. Similar work was explored in Ref. 71, which applied 18-bit-wide floating-point adders/subtractors, multipliers, and dividers to 2D fast Fourier transform (FFT) and systolic FIR (finite impulse response) filters implemented on Splash II. This work was extended to a full 32-bit floating point in Ref. 72 for multipliers based on bit-parallel adders and digit-serial multipliers. More recent work [73] re-examines these issues with an eye toward greater area efficiency.

5.2 Reconfigurable DSP System Implementation

Although recent research in reconfigurable computing has been focused on advanced issues such as dynamic reconfiguration and special-purpose architecture, most work to date has been focused on the effective use of application parallelization and specialization. In general, a number of different DSP applications have been mapped to reconfigurable computing systems containing one, several, and
many FPGA devices. In this subsection, a number of DSP projects that have been mapped to reconfigurable hardware are described. These implementations represent a broad set of DSP application areas and serve as a starting point for advanced research in years to come.

5.2.1 Image Processing Applications

The pipelined and fine-grained nature of reconfigurable hardware is a particularly good match for many image processing applications. Real-time image processing typically requires specialized datapaths and pipelining which can be implemented in FPGA logic. A number of projects have been focused in this application area. In Refs. 74 and 75, a set of image processing tasks mapped to the Splash II platform, described in Section 3.2, are outlined. Tasks such as Gaussian pyramid-based image compression, image filtering with 1D and 2D transforms, and image conversion using discrete fourier transform (DFT) operations are discussed. This work was subsequently extended to include the 2D discrete cosine transform (DCT) implemented on the Splash II platform in Ref. 76. The distributed construction of a stand-alone Splash II system containing numerous physical I/O ports is shown to be particularly useful in achieving high data rates. Because Splash II is effective in implementing systolic versions of algorithms that require repetitive tasks with data shifted in a linear array, image data can quickly be propagated in a processing pipeline. The targeted image processing applications are generally implemented as block-based systolic computations, with each FPGA operating as a systolic processor and groups of FPGAs performing specific tasks.

Additional reconfigurable computing platforms have also been used to perform image processing tasks. In Ref. 77, a commercial version of PAM, the turbochannel-based Pamette [34], is interfaced to a charge-coupled device (CCD) camera and a liquid-crystal polarizing filter is used to perform solar polarimetry. The activity of this application is effectively synchronized with software on an Alpha workstation. In Refs. 50 and 78, multi-FPGA systems are used to process 3D volume visualization data through ray casting. These implementations show favorable processing characteristics when compared to traditional microprocessor-based systems. In Ref. 79, a system is described in which a 2D DCT is implemented using a single FPGA device attached to a backplane bus-based processing card. This algorithm implementation uses distributed arithmetic and is initially coded in VHDL and subsequently compiled using RTL synthesis tools. In Ref. 80, a commercial multi-FPGA system is described that is applied to spatial median filtering. In Ref. 81, the application of a PCI-based FPGA board to 1D and 2D convolution is presented. Finally, in Ref. 82, a system implemented with a single-FPGA processing board is described that performs image interpolation. This system primarily uses bit-serial arithmetic and exploits dynamic reconfigu-
ration to quickly swap portions of the computation located in the reconfigurable hardware. Each computational task has similar computational structure, so reconfiguration time of the FPGA is minimal.

5.2.2 Video Processing Applications

Like image processing, video processing requires substantial data bandwidth and processing capability to handle data obtained from analog video equipment. To support this need, several reconfigurable computing platforms have been adapted for video processing. The PAM system [30], described in Section 3.2, was the first platform used in video applications. A PAM system programmed to perform stereo vision was applied to applications requiring 3D elevation maps such as those needed for planetary exploration. A stereo-matching algorithm was implemented that was shown to be substantially faster than programmable DSP-based approaches. This implementation employed dynamic reconfiguration by requiring the reconfiguration of programmable hardware among three distinct processing tasks at run time. A much smaller single-FPGA system, described in Ref. 83, was focused primarily on block-based motion estimation. This system tightly coupled SRAM to a single FPGA device to allow for rapid data transfer.

An interesting application of FPGAs for video computation is described in Ref. 84. A stereo transform is implemented across 16 FPGA devices by aligning two images together to determine the depth between the images. Scan lines of data are streamed out of adjacent memories into processing FPGAs to perform the comparison. In an illustration of the benefit of a single-FPGA video system, in Ref. 85 a processing platform is described in which a T805 transputer is tightly coupled with an FPGA device to perform frame object tracking. In Ref. 86, a single-FPGA video coder, which is reconfigured dynamically among three different subfunctions (motion estimation, DCT, and quantization), is described. The key idea in this project is that the data located in hardware do not move, but rather the functions which operate on it are reconfigured in place.

5.2.3 Audio and Speech Processing

Whereas audio processing typically requires less bandwidth than video and image processing, audio applications can benefit from datapath specialization and pipelining. To illustrate this point, a sound synthesizer was implemented using the multi-FPGA PAM system [30], producing real-time audio of 256 different voices at up to 44.1 kHz. Primarily designed for the use of additive synthesis techniques based on look-up tables, this implementation included features to allow frequency modulation synthesis and/or nonlinear distortion and was also used as a sampling machine. The physical implementation of PAM as a stand-alone processing system facilitated interfacing to tape recorders and audio amplifiers. The system
setup was shown to be an order-of-magnitude faster than a contemporary off-the-shelf DSP.

Other smaller projects have also made contributions in the audio and speech processing areas. In Ref. 87, a methodology is described to perform audio processing using a dynamically reconfigurable FPGA. Audio echo production is facilitated by dynamically swapping filter coefficients and parameters into the device from an adjacent SRAM. Third-party DSP tools are used to generate the coefficients. In Ref. 69, an inventive FPGA-based cross-correlator for radio astronomy is described. This system achieves high processing rates of 250 MHz inside the FPGA by heavily pipelining each aspect of the data computation. To support speech processing, a bus-based multi-FPGA board, Tabula Rasa [88], was programmed to perform Markov searches of speech phenomes. This system is particularly interesting because it allowed the use of behavioral partitioning and contained a codesign environment for specification, synthesis, simulation, and evaluation design phases.

5.2.4 Target Recognition

Another important DSP application that has been applied to Splash II is target recognition [89]. To support this application, images are broken into columns and compared to precomputed templates stored in local memory along with pipelined video data. As described in Section 3.2, near-neighbor communication is used with Splash II to compare pass-through pixels with stored templates in the form of partial sums. After an image is broken into pieces, the Splash II implementation performs second-level detection by roughly identifying sections of subimages that conform to objects through the use of templates. In general, the use of FPGAs provides a unique opportunity to quickly adapt target recognition to new algorithms, something not possible with ASICs. In another FPGA implementation of target recognition, researchers [90] broke images into pieces called chips and analyzed them using a single FPGA device. By swapping target templates dynamically, a range of targets may be considered. To achieve high-performance design, templates were customized to meet the details of the target technology. In Ref. 91, a description is given of a novel software system that is used to map a high-level description of a target recognition algorithm to a multi-FPGA system. This software tool set converts algorithmic descriptions previously targeted to the Khoros [92] design environment into a format which can be loaded into a Wildforce system from Annapolis Micro Systems [33].

5.2.5 Communication Coding

In modern communication systems, signal-to-noise ratios make data coding an important aspect of communication. As a result, convolutional coding can be used to improve signal-to-noise ratios based on the constraint length of codes
without increasing the power budget. Several reconfigurable computing systems have been configured to aid in the transmission and receipt of data. One of the first applications of reconfigurable hardware to communications involved the PAM project [30]. On-board PAM system RAM was used to trace through $2^{14}$ possible states of a Viterbi encoder, allowing for the computation of 4 states per clock cycle. The flexibility of the system allowed for quick evaluation of new encoding algorithms. A run-length Viterbi decoder, described in Ref. 93, was created and implemented using a large reconfigurable system containing 36 FPGA devices. This constraint length 14 decoder was able to achieve decode rates of up to 1 Mbit/sec. In Ref. 94, a single-FPGA system is described that supports variable-length code detection at video transfer rates.

5.3 Reconfigurable Computing Architecture and Compiler Trends for DSP

Over the past decade, the large majority of reconfigurable computing systems targeted to DSP have been based on commercial FPGA devices and have been programmed using RTL and structural hardware description languages. Although these architectural and programming methodologies have been sufficient for initial prototyping, more advanced architectures and programming languages will be needed in the future. These advancements will especially be needed to support advanced features such as dynamic reconfiguration and high-level compilation over the next few years. In this subsection, recent trends in reconfigurable computing-based DSP with regard to architecture and compilation are explored. Through near-term research advancement in these important areas, the breadth of DSP applications that are appropriate for reconfigurable computing is likely to increase.

5.3.1 Architectural Trends

Most commercial FPGA architectures have been optimized to perform efficiently across a broad range of circuit domains. Recently, these architectures have been changed to better suit specific application areas.

Specialized FPGA Architectures for DSP. Several FPGA architectures specifically designed for DSP have been proposed over the past decade. In Ref. 95, a fine-grained programmable architecture is considered that uses a customized LUT-based logic cell. The cell is optimized to efficiently perform addition and multiplication through the inclusion of XOR gates within LUT-based logic blocks. Additionally, device intercell wire lengths are customized to accommodate both local and global signal interconnections. In Ref. 96, a specialized DSP operator array is detailed. This architecture contains a linear array of adders and shifters connected to a programmable bus and is shown to efficiently implement
FIR filters. In Ref. 97, the basic cell of a LUT-based FPGA is augmented to include additional flip-flops and multiplexers. This combination allows for tight interblock communication required in bit-serial DSP processing. External routing was not augmented for this architecture due to the limited connectivity required by bit-serial operation.

Whereas fine-grained look-up table FPGAs are effective for bit-level computations, many DSP applications benefit from modular arithmetic operations. This need has led to an interest in reconfigurables with coarse-grained functional units. One such device, Paddi [98], is a DSP-optimized parallel computing architecture that includes eight ALUs and localized memories. As part of the architecture, a global instruction address is distributed to all processors, and instructions are fetched from a local instruction store. This organization allows for high instruction and I/O bandwidth. Communication paths between processors are configured through a communication switch and can be changed on a per-cycle basis. The Paddi architecture was motivated by a need for high data throughput and flexible datapath control in real-time image, audio, and video processing applications. The coarse-grained Matrix architecture [99] is similar to Paddi in terms of block structure, but it exhibits more localized control. Whereas Paddi has a VLIW-like control word which is distributed to all processors, Matrix exhibits more multiple instruction multiple data (MIMD) characteristics. Each Matrix tile contains a small processor, including a small SRAM and an ALU which can perform 8 bit data operations. Both near-neighbor and length-4 wires are used to interconnect individual processors. Interprocessor data ports can be configured to support either static or data-dependent dynamic communication.

The ReMarc architecture [100], targeted to multimedia applications, was designed to perform a SIMD-like computation with a single control word distributed to all processors. A 2D grid of 16-bit processors is globally controlled with a SIMD-like instruction sequencer. Interprocessor communication takes place either through near-neighbor interconnect or through horizontal and vertical buses. The MorphoSys architecture [101] was also designed for SIMD operation, but, unlike ReMarc, it offers support for efficient dynamic reconfiguration. Functional blocks in this architecture can perform either 8- or 16-bit ALU operations. A three-level hierarchy of interconnect provides for flexible interblock communication. The Chess architecture [102] is based on 4-bit ALUs and contains pipelined near-neighbor interconnect. Each computational tile in the architecture contains memory which can either store local processor instructions or local data memory. The Colt architecture [103] was specially designed as an adaptable architecture for DSP that allows interconnect reconfiguration. This coarse-grained architecture allows run-time data to steer programming information to dynamically determined points in the architecture. A mixture of both 1-bit and 16-bit functional units allows both bit and word-based processing.
Whereas coarse-grained architectures organized in a 2D array offer significant interconnect flexibility, often signal processing applications, such as filtering, can be accommodated with a linear computational pipeline. Several coarse-grained reconfigurable architectures have been created to address this class of applications. PipeRench [104] is a pipelined, linear computing architecture that consists of a sequence of computational stripes, each containing look-up tables and data registers. The modular nature of PipeRench makes dynamic reconfiguration on a per-stripe basis straightforward. Rapid [105] is a reconfigurable device based on both linear data and control paths. The coarse-grained architecture for this datapath includes multipliers, adders, and pipeline registers. Unlike PipeRench, the interconnect bus for this architecture is segmented to allow for nonlocal data transfer. In general, communication patterns built using Rapid interconnect are static, although some dynamic operation is possible. A pipelined control bus that runs in parallel to the pipelined data can be used to control computation.

**DSP Compilation Software for Reconfigurable Computing.** Although some high-level compilation systems designed to target DSP algorithms to reconfigurable platforms have been outlined and partially developed, few complete synthesis systems have been constructed. In Ref. 106, a high-level synthesis system is described for reconfigurable systems that promotes high-level synthesis from a behavioral synthesis language. For this system, DSP designs are represented as a high-level flowgraph and user-specified performance parameters in terms of a maximum and minimum execution schedule are used to guide the synthesis process. In Ref. 60, a compilation system is described that converts a standard ANSI C representation of filter and FFT operations into a bit-serial circuit that can be applied to an FPGA or to a field programmable multichip module. In Ref. 107, a compiler, debugger, and linker targeted to DSP data acquisition is described. This work uses a high-level model of communicating processes to specify computation and communication in a multi-FPGA system. By integrating digital-to-analog (D/A) and A/D converters into the configurable platform, a primitive digital oscilloscope is created.

The use of dynamic reconfiguration to reduce area overhead in computing systems has recently motivated renewed interest in reconfigurable computing. Although a large amount of work remains to be completed in this area, some preliminary work in the development of software to manage dynamic reconfiguration for DSP has been accomplished. In Ref. 108, a method of specifying and optimizing designs for dynamic reconfiguration is described. Through selective configuration scheduling, portions of an application used for 2D image processing is dynamically reconfigured based on need. Later work [46] outlined techniques based on bipartite matching to evaluate which portions of a dynamic application should be reconfigured. The technique is demonstrated using an image filtering example.
Several recent DSP projects address the need for both compile-time and run-time management of dynamic reconfiguration. In Ref. 109, a run-time manager is described for a single-chip reconfigurable computing system with a large FIR filter used as a test case. In Ref. 45, a compile-time analysis approach to aid reconfiguration is described. In this work, all reconfiguration times are statically determined in advance and the compilation system determines the minimum circuit change needed at each run-time point to allow for reconfiguration. Benchmark examples which use this approach include arithmetic units for FIR filters which contain embedded constants. Finally, in Ref. 62, algorithms are described that perform dynamic reconfiguration to save DSP system power in time-varying applications such as motion estimation. The software tool created for this work dynamically alters the search space of motion vectors in response to changing images. Because power in the motion estimation implementation is roughly correlated with search space, a reduced search proves to be beneficial for applications such as mobile communications. Additionally, unused computational resources can be scheduled for use as memory or rescheduled for use as computing elements as computing demands require.

Although the integration of DSP and reconfigurable hardware is just now being considered for single-chip implementation, several board-level systems have been constructed. GigaOps provided the first commercially available DSP and FPGA board in 1994 containing an Analog Devices 2101 PDSP, 2 Xilinx XC4010s, 256KB of SRAM, and 4MB of DRAM. This PC-based system was used to implement several DSP applications, including image processing [110]. Another board-based DSP/FPGA product line is the Arix-C67 currently available from MiroTech Corporation [111]. This system couples a Xilinx Virtex FPGA with a TMS320C6701 DSP. In addition to supporting several PC-bus interfaces, this system has an operating system, a compiler, and a suite of debugging software.

6 THE FUTURE OF RECONFIGURABLE COMPUTING AND DSP

The future of reconfigurable computing for DSP systems will be determined by the same trends that affect the development of these systems today: system integration, dynamic reconfiguration, and high-level compilation. DSP applications are increasingly demanding in terms of computational load, memory requirements, and flexibility. Traditionally, DSP has not involved significant run-time adaptivity, although this characteristic is rapidly changing. The recent emergence of new applications that require sophisticated, adaptive, statistical algorithms to extract optimum performance has drawn renewed attention to run-time reconfigurability. Major applications driving the move toward adaptive computation in-
clude wireless communications with DSP in hand-sets, base stations and satel-

lites, multimedia signal processing [112], embedded communications systems found in disk drive electronics [10] and high-speed wired interconnects [113], and remote sensing for both environmental and military applications [114]. Many of these applications have strict constraints on cost and development time due to market forces.

The primary trend impacting the implementation of many contemporary DSP systems is Moore’s law, resulting in consistent exponential improvement in integrated circuit device capacity and circuit speeds. According to the National Technology Roadmap for Semiconductors, growth rates based on Moore’s law are expected to continue until at least the year 2015 [115]. As a result, some of the corollaries of Moore’s law will require new architectural approaches to deal with the speed of global interconnect, increased power consumption and power density, and system and chip-level defect tolerance. Several architectural approaches have been suggested to allow reconfigurable DSP systems to make the best use of large amounts of VLSI resources. All of these architectures are characterized by heterogeneous resources and novel approaches to interconnection. The term system-on-a-chip is now being used to describe the level of complexity and heterogeneity available with future VLSI technologies. Figures 8 and 9 illustrate various characteristics of future reconfigurable DSP systems. These are not mutually exclusive and some combination of these features will probably emerge based on driving application domains such as wireless handsets, wireless base stations, and multimedia platforms. Figure 8, taken from Ref. 116, shows an architecture containing an array of DSP cores, a RISC microprocessor, large amounts of uncommitted SRAM, a reconfigurable FPGA fabric, and a reconfigurable interconnection network. Research efforts to condense DSPs, FPGA logic, and memory on a single substrate in this fashion are being pursued in the Pleiades project [116,117]. This work focuses on selecting the correct collection of functional units to perform an operation and then intercon-

![Figure 8](image_url)  
Figure 8  Architectural template for a single-chip Pleiades device. (From Ref. 116.)
necting them for low power. An experimental compiler has been created for this system [116] and testing has been performed to determine appropriate techniques for building a low-power interconnect. An alternate, adaptive approach [118] that takes a more distributed view of interconnection appears in Figure 9. This figure shows how a regular tiled interconnect architecture can be overlaid on a set of heterogeneous resources. Each tile contains a communication switch which allows for statically scheduled communication between adjacent tiles. Cycle-by-cycle communications information is held in embedded communication switch SRAM (SME).

The increased complexity of VLSI systems enabled by Moore’s law presents substantial challenges in design productivity and verification. To support the continued advancement of reconfigurable computing, additional advances will be needed in hardware synthesis, high-level compilation, and design verification. Compilers have recently been developed which allow software development to be done at a high level, enabling the construction of complex systems including significant amounts of design reuse. Additional advancements in multicompliers [119] will be needed to partition designs, generate code, and synchronize interfaces for a variety of heterogeneous computational units. VLIW compilers [120] will be needed to find substantial amounts of instruction-level parallelism in DSP code, thereby avoiding the overhead of run-time parallelism extraction. Finally, compilers that target the codesign of hardware and software and leverage techniques such as static interprocessor scheduling [56] will allow truly reconfigurable systems to be specialized to specific DSP computations.

A critical aspect of high-quality DSP system design is the effective integration of reusable components or cores. These cores range from generic blocks like RAMs and RISC microprocessors to more specific blocks like MPEG decoders.

![Figure 9](image.jpg)

**Figure 9** Distributed single-chip DSP interconnection network. (From Ref. 118.)
and PCI bus interfaces. Trends involving core development and integration will continue and tools to support core-based design will emerge, allowing significant user interaction for both design-time and run-time specialization and reconfiguration. Specialized synthesis tools will be refined to leverage core-based design and to extract optimum efficiency for DSP kernels while using conventional synthesis approaches for the surrounding circuitry [1,121].

Verification of complex and adaptive DSP systems will require a combination of simulation and emulation. Simulation tools like Ptolemy [122] have already made significant progress in supporting heterogeneity at a high level and will continue to evolve in the near future. Newer verification techniques based on logic emulation will emerge as effective mechanisms for using reconfigurable multi-FPGA platforms to verify DSP systems are developed. Through the use of new generations of FPGAs and advanced emulation software [123], new emulation systems will provide the capability to verify complex systems at near real-time rates.

Power consumption in DSP systems will be increasingly important in coming years due to expanding silicon substrates and their application to battery-powered and power-limited DSP platforms. The use of dynamic reconfiguration has been shown to be one approach that can be used to allow a system to adapt its power consumption to changing environments and computational loads [62]. Low-power core designs will allow systems to be assembled without requiring detailed power optimizations at the circuit level. Domain-specific processors [116] and loop transformations [124] have been proposed as techniques for avoiding the inherent power inefficiency of von Neumann architectures [125]. Additional computer-aided design tools will be needed to allow high-level estimation and optimization of power across heterogeneous architectures for dynamically varying workloads.

The use of DSP in fields such as avionics and medicine have created high-reliability requirements that must be addressed through available fault tolerance. Reliability is a larger system goal, of which power is only one component. As DSP becomes more deeply embedded in systems, reliability becomes even more critical. The increasing complexity of devices, systems, and software all introduce numerous failure points which need to be thoroughly verified. New techniques must especially be developed to allow defect tolerance and fault tolerance in the reconfigurable components of DSP systems. One promising technique which takes advantage of FPGA reconfiguration at various grain sizes is described in Ref. 126.

Reconfiguration for DSP systems is driven by many different goals: performance, power, reliability, cost, and development time. Different applications will require reconfiguration at different granularities and at different rates. DSP systems that require rapid reconfiguration may be able to exploit regularity in their algorithms and architectures to reduce reconfiguration time and power consump-
tion. An approach called dynamic algorithm transforms (DAT) \cite{127,128} is based on the philosophy of moving away from designing algorithms and architectures for worst-case operating conditions in favor of real-time reconfiguration to support the current situational case. This is the basis for reconfigurable ASICs (RAS-ICs) \cite{129}, where just the amount of flexibility demanded by the application is introduced. Configuration cloning \cite{130}, caching, and compression \cite{131} are other approaches to address the need for dynamic reconfiguration. Techniques from computer architecture regarding instruction fetch and decode need to be modified to deal with the same tasks applied to configuration data.

In conclusion, reconfiguration is a promising technique for the implementation of future DSP systems. Current research in this area leverages contemporary semiconductors, architectures, computer-aided design tools, and methodologies in an effort to support the ever-increasing demands of a wide range of DSP applications. There is much work still to be done, however, because reconfigurable computing presents a very different computational paradigm for DSP system designers as well as DSP algorithm developers.

REFERENCES


69. BV Herzen. Signal processing at 250 MHz using high performance FPGAs. Pro-
86. J Villasenor, B Schoner, C Jones. Video communications using rapidly reconfig-