The Pervasiveness of Reconfigurable Computing

The von Neumann Paradigm loosing its Dominance


Abstract. Reconfigurable Computing, the second RAM-based machine paradigm offers a drastic reduction of the electric energy budget and speedup factors by several orders of magnitude - compared to using the von Neumann paradigm, now loosing its dominance. Coming along with a growing configware industry the new discipline of configware engineering is developing as the counterpart of Software Engineering. This paper intends to be a wake-up call discussing the impact of this primarily non-instruction-stream-based fundamentally different mind set on the skill requirements of the IT job market, and on CS and related curricula.

1. PREFACE

Currently the dominance of the basic computing paradigm is gradually wearing off with the growing pervasiveness of Reconfigurable Computing (RC) - bringing profound changes to the practice of both, scientific computing and ubiquitous embedded systems, as well as new promise of disruptive new horizons for affordable very high performance computing. Due to RC the desk-top personal supercomputer is near. To obtain the payoff from RC we need a new understanding of computing and supercomputing. For bridging the translational gap, the software / configware chasm, we need to think outside the box.

Sceptic about the significance of RC, some colleagues from CS pointed toward the rise and fall of hardware / software co-design (HS codesign). Sure, this has been obscured by renaming conference series and changing slogans of the EDA industry: co-design, H/S co-design, CODES, High-Level Synthesis, System Synthesis, ESDA (electronic system design automation), ESL (electronic system-level design). However, the truth is, that hardware / software codesign is a long-lasting success story within the (also undersurface) embedded systems success story [2] [3] - despite troublesome experiences with EDA industry products. A fall happened inside the CS curricula because the currently still dominant CS culture mainly failed to cure the hardware / software chasm. This is a reason, why embedded software is often implemented by hardware people. The embedded systems scene now is running its own curriculum development effort, since typical CS graduates are not really qualified and tend to miss this most important job market1.

Inside the embedded Systems scene at first glance the use of reconfigurable devices like FPGAs has looked more like a variety of hardware design, but on a strange platform. Now we have 2 reconfigurable computing scenes (fig. 2). Meanwhile FPGAs are also used everywhere for high performance in scientific computing, where this is really a new computing culture - not at all a variety of hardware design. Instead of HS codesign we have here software / configware co-design (SC co-design), which is really a computing issue. This major new direction of developments in science will determine how academic computing will look in 2015 or even earlier. The instruction-stream-based mind set will loose its monopoly-like dominance and the CPU will quit its central role - to be more an auxiliary clerk, also for software compatibility issues.

This new direction has not yet drawn the attention of the curriculum planners within the embedded systems scene. For computer science this is the opportunity of the century, of decampment for heading toward new horizons. This should be a wake-up call to CS curriculum development. Each of the many different application domains has only a limited view of computing and takes it more as a mere technique than as a science on its own. This fragmentation makes it very difficult to bridge the cultural and practical gaps, since there are so many different actors and departments involved. Only Computer Science can take the full responsibility to merge Reconfigurable Computing into CS curricula for providing Reconfigurable Computing Education from its roots. CS has the right perspective for a transdisciplinary unification in dealing with problems, which are shared across many different application domains. This new direction would also be helpful to reverse the current downward trend of CS enrolment.

2. THE PERVASIVENESS OF THE FPGA

The FPGA (field-programmable gate array) is an array of gate level reconfigurable elements (rE) embedded in a reconfigurable interconnect fabrics [6]. Its configware code (reconfiguration code [7]; fig 8) is stored in a distributed hRAM memory (hRAM for "hidden RAM"), hidden in the background of the FPGA circuitry. Comparable to booting a computer the configware has to be loaded after each power-on. FPGAs are with 6 billion


Fig. 2. Two different reconfigurable computing cultures.
US-Dollars the fastest growing segment of the semiconductor market. Complex projects can be implemented on FPGAs, commodities off the shelf (COTS), without needing very expensive customer-specific silicon. The growth of the number of design starts is predicted from 80,000 in 2006 to 115,000 in 2010 [Dataquest]. Impressive are the hit rates by Google [1] [8] upon the embedded systems community is demonstrated by the combination of topic area keywords and FPGA [1] illustrates, that FPGAs massively go in a wide variety of application areas.

The Strategic Significance of Reconfigurable Computing. The area of embedded systems is unthinkable without FPGAs [9]. This has been the driving force behind the commercial break-through of FPGAs. The pervasiveness of FPGAs within the embedded systems community is demonstrated by the number of hits by Google [1] [8] upon FPGA combined with application areas like embedded (3,280,000), wireless (1,490,000), automotive (915,000), multimedia (731,000), signal processing (647,000), music (398,000), image processing (272,000) and others. About 90% of all software is implemented for embedded systems [5] (Fig. 3) dominated by FPGAs usage, where frequently hardware / configware / software partitioning problems have to be solved. The quasi monopoly of the von Neumann mind set in most of our CS curricula prohibits this dichotomic qualification of our graduates, needed for the requirements of the contemporary and future job market. At a summit meeting of US state governors Bill Gates has drastically criticized this Situation in CS education.

FPGAs in Scientific Computing. The Pervasiveness of FPGAs is not limited to embedded systems, but is also spread over practically all areas of scientific computing, where high performance is required and access to a supercomputing center is not available or not affordable. Some examples are: medical (710,000), physics (508,000), chemical (247,000), mathematics (171,000), fluid dynamics (162,000), astrophysics (158,000), bio (140,000), weather (118,000) and other mostly non-embedded scientific applications.

FPGAs and the EDA industry. The pervasiveness of FPGAs also reaches the EDA (Electronic Design Automation) industry, where all major firms spend a substantial effort to offer a variety application development tools and environments for FPGA-based product development. Also FPGA vendors have cooperations with firms in the EDA industry and offer such tools and development environments. Since this is a highly complex market his paper does not go into detail because of a lack of space.

Configware Industry. After switch-on of the supply power the configuration code has to be downloaded to the FPGA’s hRAM, which is a kind booting like known from the VLSI processor. But the source of this code for FPGAs1 is not software: it definitely does not program instruction streams. The advent of FPGAs provides a second RAM-based fundamental paradigm: the Kress-Kung machine [11], which, however, is not instruction-stream-based. Instead of organizing the schedule for instruction executions the compilation for FPGAs has to organize the resources by placement and routing and, based on the result, to implement the data schedules for preparing the data streams moving through these resources (Fig. 4d). FPGAs or, the Kress-Kung machine, respectively, has no „instruction fetch“ at run time. Not to confuse students and customers with the term „software“ another term is used for these non-procedural programming sources of RC: the term configware. Not only FPGA vendors offer configware modules to their customers. Also other commercial sources are on the market: a growing configware industry - the little sister of the software industry.

3. THE RECONFIGURABLE COMPUTING PARADOX

Compared to software implementations sensational speed-up factors have been reported for software to configware migrations by using FPGAs. Fig. 1b shows a few speedup factors picked up from literature, reporting a factor of 7.6 in accelerating raytracing calculations [12], a factor of 10 for FFT (fast Fourier transform), a speedup factor of 35 in traffic simulations [13]. For a commercially available Lanman/NTLM Key Recovery Server [14] a speedup of 50 - 70 is reported. Another cryptology application reports a factor of 1305 [16]. A speedup by a factor of 304 is reported for a R/T spectrum analyzer [18]. In the DSP area [19] for MAC [19] operations a speedup factor of 100 has been reported compared to the fastest DSP on the market (2004) [20]. Already in 1997 versus the fastest DSP a speedup of 50 - 46 has been obtained [21]. In Biology and genetics (also see [22][22]) a speedup of up to 30 has been shown in protein identification [24], by 133 [25] and up to 500 [26] in gene 1, in this context the term FPGA does not mean non-FPGA on-chip-modules like processor cores etc., which are usually embedded in modern so-called platform-FPGAs.
FPGA’s bad efficiency. The trend line in fig. 1b, obtained by linear extrapolation, indicates that FPGA performance doubles every year. By the year 2005, 2 decades after the market introduction of the first little FPGA, this yields a lead of a factor of substantially more than 10,000 over the 8080-compatible microprocessor. These unbelievable performance margins contrast against the very bad technological parameters of the FPGA, like area efficiency, integration density, clock frequency, and, compared to hardwired ASICs: power dissipation [44]. The effective integration density (transistors per chip) of FPGAs (fig 1a) is substantially more than 4 orders of magnitude (more than 10,000) behind the Gordon Moore curve. Three categories of overhead are contributing: wiring overhead, reconfigurability overhead, and routing congestion. Due to wiring overhead, the physical integration density, i. e. the real number of transistors, is down by about 2 orders of magnitude because wiring patterns take most of the chip area. The logical integration density is reduced by another 2 orders of magnitude, since, roughly only 1 of about a hundred transistors directly deserves the application, whereas the other 99 transistors are needed for reconfigurability. A third overhead effect, growing with the size of the FPGA, is routing congestion, because of local excess demands of routing resources not all inEs can be connected. FPGAs have more bad parameters. The FPGA clock frequency with around 500 MHz is about almost an order of magnitude lower that of 8080-compatible newest microprocessors with around 3 GHz. Compared to non-reconfigurable ASICs the power dissipation of FPGA-based solutions is substantially higher. All these parameters look very bad. Why are finally the performance and even electricity consumption results so good? This is the Reconfigurable Computing Paradox.

Explaining the paradox. How are these tremendous speedups by up to 4 orders of magnitude and the enormous power savings by about one order of magnitude possible despite of the really bad technology parameters of FPGAs? The impact of the paradox shift is the solution of this riddle. It is the paradigm shift, which brings completely different optimization mechanisms, which are so tremendously more effective and capable to override all these bad parameters. It is the main advantage that the machine paradigm is not instruction-stream-based and avoids the von Neumann bottleneck and related problems by three categories of new features: (1) highly parallel distributed memory organization with auto-sequencing memory banks (ASM), and, (2) no caches, and, (3) stalling-proof massive pipeline. Another reason is the fact, that the development of important parameters of von Neumann processors (fig. 3) are slowing down or stopping (like d. g. the growth of the clock frequency), or even negative growth rates, like the computational density (the computational effect per transistor). We should not forget the increase of power dissipation of microprocessors, so that even faster future models would need liquid cooling.

Proven by a multimedia application example. It has been demonstrated [92], that all algorithms needed for a world HDTV set (frame rate conversions, noise and artefacts removal, ...
contrast improvements, image data format standards conversion, adaptation to a wide variety of screen sizes, media server network functions and many more algorithms) can be successfully operated with 8 memory banks on board of a coarse grain array (rDP chip needing a clock frequency of only 250 MHz.

Minimizing the number of memory cycles. The most important means of speedup is the enormous reduction of the number of memory cycles needed with reconfigurable solutions because of the growing processor communication bandwidth gap sometimes called memory wall [100]. After switching on the supply power, the downloading of configware code into the hRAM, which reminds to booting, is a kind of super „instruction fetch“ before run time. FPGAs or, the Kress-Kung machine, respectively, has no „instruction fetch“ at run time. By avoiding the memory wall this massively contributes to the speedup obtained by software to configware migration (fig 1 b).

Algorithmic cleverness. RC is also the case for highly effective algorithm transformations. For instance, in genome analysis a datapath width of 64 bits is an overkill, causing an immense wasting of resources, because here 2 bit path width is an optimum. The space in this paper is too short to mention all occasions to draw an enormous payoff from fine grain parallelism. Often the many enormous speedups having been published are the result of algorithmic cleverness of the implementer. By the way: this kind of algorithmic cleverness is usually not taught at academic CS and CE departments.

4. WHAT MEANS RECONFIGURABLE COMPUTING

It may be called the second paradox of Reconfigurable Computing, that despite of its enormous pervasiveness, most professionals inside computer Science and related areas do not really understand its issues. To support configware engineering projects often a hardware expert is hired who may be good implementer, but is not a good translator. From a traditional CS perspective most people do not understand the key issues of this paradigm shift, or, do not even recognize at all, that RC is paradigm shift. A good approach of explanation is to compare the mid set of the Software area vs. the one of the configware field. An dominant obstacle for understanding is also the lack of a common accepted terminology, which massively causes confusion.

Software Engineering vs. Configware Engineering. In total we have 3 different kinds of programming sources (fig 4). The dual-paradigm model can be illustrated by contrasting via Nick Tredenick’s model of computer history (fig 4a vs. fig 4c). With the classical software processor only the algorithm is variable, whereas the resources are fixed (hardwired), so that only one type of program source is needed: software (fig 4a), from which the compiler generates software machine code to be downloaded into the processor RAM - the instruction schedule for the software processor (fig 4b). For the Kress-Kung machine paradigm, however, not only the algorithm, but also the resources are programmable, so that we need two different kinds of programming sources (fig 6): Configware and Flowware (fig 4c).

1) Configware [7] deserves structural programming of the resources by the „mapper“ using placement and routing or similar mapping methods (for instance by simulated annealing [50] [51] [52] [53] [54]) (fig 4d).

2) Flowware [55] deserves programming of the data streams by the „data scheduler“ (fig 4d), which generates the flowware code needed for downloading into the generic address generators (GAG) within the ASM auto-sequencing memory banks (fig 4d)

These two different fundamental machine principles, von Neumann software machine vs. the Kress-Kung machine, the configware machine, are contrasted by fig. 7.

Flowware Languages are easy to implement [55] [56]. A comparison with software programming languages is interesting [6]. Flowware language primitives for control instructions like jumps and loops can be simply adopted from classical software languages, however, for being used for manipulation of data addresses instead of instruction addresses. Flowware languages are more powerful than software languages and permitting parallel loops by using several data counters used simultaneously, such flowware language primitives are more powerful than these software primitives. Not handling instruction streams, flowware languages are much more simple (because at run time there is only “data fetch”, however, no “instruction fetch”).

Terminology. Since the basic paradigm is not instruction-stream-based, necessarily the term „Configware“ should be used for program sources, instead of the term „Software“, which would be confusing (fig 8). The term „software“ must be unconditionally restricted to traditional sources of classical instruction-stream-based computing (which is reasoned in fig 4). In fact this paradigm relies on data streams, however, not on instruction streams.

Equivocality of the term „data stream“. In computing and related areas there is a babylonian confusion around the term „stream“, „stream-based“ or „data stream“. There is an urgent need to establish a standards committee to work on terminology. For the area of reconfigurable computing the best suitable definition of „data stream“ has been established around the year 1980 by the systolic array scene [57] [58], where data streams enter and leave a datapath array being a pipe network (illustrated by fig 9). In fact, there a set of data streams is a data schedule specifying, which data item has to enter or leave which port of the array at which point of time.

The tail is wagging the dog. Because if their memory-cycle-hungry instruction-stream-driven sequential mode of operation microprocessors usually need much more powerful accelerators [44]: the tail is wagging the dog. The instruction-stream-based-only fundamental mind set (vN-only paradigm) as a common model often is still a kind of monopoly inside the qualification background of CS graduates. The real model practiced now is not the von Neumann paradigm (vN) handed down from the mainframe age. In fact, during the PC age it has been replaced by a symbiosis of the vN host and the non-vN (i.e. non-instruction-stream-based) accelerators. Meanwhile we have arrived at the kind of post-PC morphware age with a third basic model, where the accelerator has become programmable (reconfigurable). Useful for application development are Co-Compilers (fig 12), automatically partitioning from the programming source into software and configware [60]. The methodology is known from academic co-compilers [60] [61], easy to implement since most of their fundamentals have been published decades ago [63]. There is a number of trend indications pointing toward an auxiliary clerk role of the CPU for running old software and taking care of compatibility issues. „FPGA main processor vs. FPGA co-processor“ asks the CEO of Natalltech [65]: Is it time for vN to retire? The RAMP project, for instance proposes to run the operating system on FPGAs [66]. In fact, in some embedded systems, the CPU has this role already.
The Dichotomy of Machine Paradigms is rocking the foundation walls of Computer Science. Because of the lack of a common terminology this duality of paradigms is difficult to understand for people with a traditional CS background. A taxonomy of platform categories and their programming sources, quasi of a terminology floor plan, should help to catch the key issues (fig 8). The Kress-Kung machine is the data-stream-based counterpart of the instruction-stream-based von Neumann paradigm. The Kress-Kung machine does not have a program counter (fig 7 b), and, its processing unit is not a CPU (fig 7 a). Instead, it is only a DPU (Data Path Unit): without an instruction sequencer (fig 7 b).

The enabling technology of the Kress-Kung machine has one or mostly several data counters as part of the Generic Adress Generators (GAG) [67] [69] [70] within data memory banks called ASM (Auto-sequencing Memory, see fig 7 b). ASMs send and/or receive data streams having been programmed from Flowware sources [68] (fig 9). An ASM is the generalization of the DMA circuit (Direct Memory Access) [71] [72] for executing block transfers without needing to be controlled by instruction streams inside. ASMs, based on the use of distributed memory architectures [76] are very powerful architectural resources, supporting the optimization of the data storage schemes for minimizing the number of memory cycles [70]. The MoM Kress-Kung machine based on such generic address generators has been published in 1990 [73] [74]. The use of data counters replacing the program counter has first been published in 1987 [75].

Hardwired versions of the Kress-Kung machine. We may distinguish 2 classes of Kress-Kung machines (fig 8): programmable ones (morphware: reconfigurable Kress-Kung machine, needing 2 types of programming sources (see next paragraph and fig. 4 c/d): Configware for structural programming, and Flowware, for data scheduling. However, also hardwired Kress-Kung machines can be implemented for instance, (the BEE project [77]), where the configuration is been frozen and cast into hardware before fabrication. The lack of reconfigurability after fabrication by not using FPGAs of such hardwired Kress-Kung machines substantially improves the computational density (fig 1 a) for much higher speedup factors and might make sense for special purpose or domain-specific applications. Since after fabrication a reconfiguration is impossible, only one programming source is needed: Flowware.

Dynamically reconfigurable architectures and their environments illustrate the specific flavor of Configware Engineering being able to rapidly shift back and force between run time mode of operation and configuration mode. Even several separate macros can be resident in the same FPGA. Even more complex is the situation when within partially reconfigurable FPGAs some modules are in run time mode, whereas at the same time other modules are in the configuration phase, so that a FPGA could reconfigure itself. Some macros can be removed at the same time, when other macros are active by being in the run time mode. Configware operating systems are managing such scenarios [78] [79].

New educational approaches needed. Although configure engineering is a discipline of its own, fundamentally different from software engineering, and, a configure industry is already existing and growing, it is too often ignored by our curricula. Modern FPGAs as COTS (commodities off the shelf) have all 3 paradigms on board of the same VLSI chip: hardwired accelerators, microprocessors (and memory banks), and FPGAs, and we need both, software and configure, to program the same chip. To cope with the clash of cultures we need interdisciplinary curricula merging all these different backgrounds in a systematic way. We need innovative lectures and lab courses supporting the integration of reconfigurable computing into progressive curricula.

5. RECONFIGURABLE SUPERCOMPUTING

The penetration of reconfigurable platforms like FPGAs in the supercomputing community is demonstrated by the number of hits of Google [8] in reply to FPGA, in combination with high performance computing (81,200), or supercomputing (65,500). The pervasiveness of FPGAs for many typical supercomputing application areas is shown by the number of Google hits on FPGA in combination with, for instance, the keywords oil and gas (710,000 times), physics (508,000), defense (287,000), and weather (118,000) etc. [1]. In the supercomputing application on oil and gas for the migration onto FPGAs a speedup factor of 17 has been reported [87] [88], together with an enormous reduction of the electricity bill s a side effect, since with FPGAs you can save energy [89]. A geophysicist reports, that with 7 US-Cents pro kWh more than 10,000 US-Dollars per year on the electricity bill could be saved - for each 19 inch module with 64 processors: a yearly saving of half a million US-Dollars on the electricity bill s a side effect, since with FPGAs you can save energy [89]. A geophysicist reports, that with 7 US-Cents pro kWh more than 10,000 US-Dollars per year on the electricity bill could be saved - for each 19 inch module with 64 processors: a yearly saving of half a million US-Dollars on the electricity bill s a side effect, since with FPGAs you can save energy [89]. A geophysicist reports, that with 7 US-Cents pro kWh more than 10,000 US-Dollars per year on the electricity bill could be saved - for each 19 inch module with 64 processors: a yearly saving of half a million US-Dollars on the electricity bill s a side effect, since with FPGAs you can save energy [89]. A geophysicist reports, that with 7 US-Cents pro kWh more than 10,000 US-Dollars per year on the electricity bill could be saved - for each 19 inch module with 64 processors: a yearly saving of half a million US-Dollars on the electricity bill s a side effect, since with FPGAs you can save energy [89]. A geophysicist reports, that with 7 US-Cents pro kWh more than 10,000 US-Dollars per year on the electricity bill could be saved - for each 19 inch module with 64 processors: a yearly saving of half a million US-Dollars on the electricity bill s a side effect, since with FPGAs you can save energy.

Platform FPGAs. So far this paper has pointed toward RC using FPGAs. Originally simple FPGAs have been general purpose devices, since flipflops and gates (LUTs) are general purpose elements. But for completeness it should be mentioned, that modern FPGAs, which are often called

<table>
<thead>
<tr>
<th>#</th>
<th>FPGA</th>
<th>rDPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>terminology</td>
<td>field-programmable gate array</td>
</tr>
<tr>
<td>2</td>
<td>reconfiguration granularity</td>
<td>fine-grained</td>
</tr>
<tr>
<td>3</td>
<td>data path width</td>
<td>~ 1 bit e.g. ~ 32 bits</td>
</tr>
<tr>
<td>4</td>
<td>physical level of basic reconfigurable units (RU)</td>
<td>gate level</td>
</tr>
<tr>
<td>5</td>
<td>typical RU examples</td>
<td>LUT (look-up table), determines the logic function of the RU (and, or, not, etc. or flip-flop) ALU-like, floating point, special functions, etc.</td>
</tr>
<tr>
<td>6</td>
<td>configuration time</td>
<td>milliseconds</td>
</tr>
<tr>
<td>7</td>
<td>clock cycle time</td>
<td>~ 0.5 GHz</td>
</tr>
<tr>
<td>8</td>
<td>typical effective integration density compared to Gordon Moore curve</td>
<td>reduced by a factor of more than 10,000 (fig. 1a)</td>
</tr>
</tbody>
</table>

Fig. 5. fine-grained vs. coarse-grained reconfigurability.
platform FPGAs, include other modules embedded in their reconfigurable interconnect fabrics, like, for instance, one or several microprocessors cores (Power PC, ARM, or others) multiple memory banks, multipliers, floating point units, etc., and fast I/O interfaces. Such platform FPGAs are not fully general purpose. The particular collection of on-chip extras usually targets a particular user market segment and makes the platform more or less domain-specific.

**rDPA (reconfigurable Data Path Array).** Distinguishing reconfigurable devices has several dimensions. A second dimension makes us compare fine-grained vs. coarse-grained reconfigurability (fig. 5). FPGAs with Es of ~1 bit datapath width are fine-grained reconfigurable. However, coarse-grained reconfigurable architectures (fig 10) [44], rDPAs (reconfigurable Data Path Array) [90] [91], have path widths like e. g. 32 bits [92] [97]., like, for instance, the KressArray [50] [51] [98], a generalization of the systolic array [57] [58], which is supported by an architecture „Design Space Xplorer“.

<table>
<thead>
<tr>
<th>Program Source</th>
<th>Machine Paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>(no „program“)</td>
<td>(none)</td>
</tr>
<tr>
<td>Flowware[6]</td>
<td>von Neumann (V)</td>
</tr>
<tr>
<td>Software + Configware + Flowware</td>
<td>dual paradigm: vN + AM</td>
</tr>
</tbody>
</table>

**rDPUs (reconfigurable Data Path Units)** [50] [51] [92], like, for instance, the KressArray [50] [51] [98], a generalization of the systolic array [57] [58], which is supported by an architecture „Design Space Xplorer“.

**Conforming to the mind set of CS.** Other advantages are the ease of compilation for rDPs, and configuration times by 2 - 3 orders of magnitudes faster. Because rDPs and their rDPUs are objects of a higher abstraction level, where we find ALUs, registers, and memory banks, these resources come very close to the mind set of CS experts, where the abstraction level of FPGAs better fit to the background of hardware people. Another advantage over FPGAs is the ease of co-compilation. This is a very important advantage, because the state of the art in application development tools for FPGAs [99] is insufficient. The slogan system level design „The business press doesn’t understand EDA, because EDA is too esoteric“ [93] could be generalized into: „The CS community doesn’t understand EDA, because EDA is too esoteric“. ESL [19] offered the enticing hope of specifying a system in an implementation-neutral language, pushing a button, and out would emerge the full, hardware [ configure ] / software co-design. But the dream remained elusive [94] The CS community should take the responsibility instead of waiting for the EDA industry to follow. With the personal supercomputer (chapter 6), based on an on-chip rDPA [19] accelerator, and supported by software / configure co-compilers as demonstrated would de-couple scientific high performance computing from the problems of the EDA industry.

**The communication complexity** in classical supercomputing is growing disproportionally because of bottlenecks which are typical for classical supercomputing, mainly determined by the memory communication gap [100]. Bus systems and other switching subsystems tend toward a high memory-cycle-hungry overhead [101]: implications of the von Neumann bottleneck Data transport at run time is a dominant problem, whereas there is no lack of affordable CPU resources. A typical mind set is shown by the standard of MPI (Message Passing Interface), based on Tony Hoare’s model of Communicating Sequential Processes model (CSP) [102] [103] [104] for implementing the exchange of messages between processors in shared memory architectures. The throughput scalability of a particular application often substantially misses the peak performance, which the platform seems to offer [105]. Amdahl's law explains just one of several reasons of inefficiency [106]. However, Kress-Kung machines do not have a von Neumann bottleneck. Fig. 13 summarizes the most important sources of speed-up by memory cycle saving from software to configure migration.

**Moving the stool - not the grand piano.** Due to the instruction-stream-based mind set with classical supercomputing the data are moved between CPUs and memory by buses and/or switch boxes at run time. This memory-cycle-hungry method reminds to moving the grand piano to the pianist’s stool (fig. 11). However, the data-stream-based mind set of Reconfigurable Computing follows the inverse approach: moving the stool and not the grand piano. Primarily the data are not transferred from memory directly to/from all rDPUs [19]. But the locality of an operation is placed to the right place, where the data stream within a pipe network comes by anyway. These routes through the pipe networks are optimized and decided at compile time, are configured before run time, and remain unchanged all the time during run time. Intermediate results are not stored in memory, but always

---

**Fig. 8. Contemporary Terminologie fo the dual paradigm computing age (compare fig 4 and 7).**

---

**Fig. 7. Fundamental computing machine principles: a) von Neumann, b) Kress-Kung machine (reconfigurable or fixed).**

---

piped through from one rDPU to its neighbor. The only memory cycles are needed for transfers between the rDPAs [19] external ports and the ASMs connected to it (fig 7 b and 9). Since most applications have drastically less rDPUs than data items, there is much less to be moved. Avoiding instruction fetch at run time is another speedup aspect. MPI is not needed in such an environment.

**Implementation of Flowware.** In a pure bred Reconfigurable Computing system the only form of communication via memory modules is by data streams (fig 9) between the rDPA(s) and the distributed ASM data memory banks, which store only initial operands and final results, however, no intermediate results nor other messages. For optimum parallelism the number of ASM banks [76] should fit to the number of DPA ports. Over compilation from flowware sources the data streams are implemented (fig 4) by programming the **Generic Address Generators (GAG)** [67] [69] [70] [107] within the ASMs (fig 7 b).

**Storage schemes** are not restricted to vectors and matrixes, since GAGs support a wide variety of analytical transformations for a rich supply of storage schemes with minimum memory cycle requirements. GAGs usually do not need memory cycles for run time address computation. GAGs by 4 orders of magnitude, its much better communication and computational biology make CS more fascinating, not only for students.

**Reconfigurable Supercomputing** has been commercialized by Cray offering a 19 inch module XD1 with 6 Xilinx FPGAs Virtex-4 [109] and sgi (Silicon Graphics) offering a 19 inch module XD1 with 6 Xilinx FPGAs Virtex-4 [109] by Cray. Networks of personal supercomputers (NOPS) are near to reach unbelievable very high performance horizons.

Within this context a two-dimensional address space opens up an unexpected wealth of efficient and easily GAG-transformable storage schemes [67] [76]. An useful side effect of a 2-D addressing is an excellent versatile visualization support.

Reconfigurable Supercomputing is near. Because of the ease of co-compilation, with the computing density better than that of FPGAs by 4 orders of magnitude, its much better communication bandwidth and other advantages (compare fig. 5), the rDPA [19] is the case for the desktop personal supercomputer.

**Clearing out the microprocessor chip.** Within 8080-like microprocessors the CPU takes only a small percentage of the chip area. Most of the area is lost for caches to cope with the memory wall. Caches are useful for highly frequented instruction loops where, however, the acceleration factors are limited and strongly depend on the type of application. But for the Kress-Kung machine caches are useless, since in data loops the values usually do not repeat. With this paradigm other such much more powerful mechanisms are available to minimize memory cycles. This is the case for a rDPA as a co-processor, replacing the cache and other stuff not needed any more on the microprocessor chip.

**The Personal Supercomputer.** PCs with a powerful programmable accelerator board on board of the processor chip and using a co-compiler (fig 12) are the key of the personal supercomputer (PS). Forerunners of such a PS have been published years ago, for instance, related to n-body simulation [43]. Astrophysicists have complained, that even the most powerful available supercomputer enabled the simulation of star clusters only up to a size of 100. GRAPE, a PC extension board, however supported sizes up to about 1000 [116] [118], but without changing the algorithm, for example the goal of Prof. Rainer Spurzem of the more than 300 years old Astronomisches Rechen-Institut (University of Heidelberg) [117], together with Prof. Reinhard Manner (University of Mannheim): a reconfigurable
accelerator AHA-GRAPE [119] supporting much more than just n-body simulation. However, by orders of magnitude more powerful is the use of coarse-grained morphware (fig 10). In fact, the PS is near: not only for the desktop of individual users, but also as a component, not only for networks of reconfigurable computers (NORCs [120] [121]), but particularly for networks of personal supercomputers (NOPS), new supercomputing centers reaching hitherto unbelievable very high performance horizons, as well as for grids of personal supercomputers (GOPS).

The Technology is available. Coarse-grained reconfigurable datapath arrays, rDPAs, with up to more than a hundred DPUs on a single microchip are available already today [92]. Since caches do not make sense for data, such a rDPA could easily be placed onto the intel-compatible processor chip: the PS chip would be ready. Also software / configware co-compiler (fig 12) for the PS have been implemented in academia [60] [61] [122] [123], for instance accepting C language sources for the coarse-grained reconfigurable KressArray [50] [51] [98] [125]. Implementing such a co-compiler is no problem and parts of the methodology are decades old [126]. The personal supercomputer (PS) is near. Only an investor is missing for a commercial co-compiler [127].

7. CURRICULUM RECOMMENDATIONS

The Productivity Crisis. Rapidly growing complexity and pervasiveness of RC-based multi-paradigm devices leads to a productivity crisis of major proportions. On the other hand RC is an efficient approach to cope with the accelerating VLSI design crisis. While the economic importance of RC and its FPGAs is widely acknowledged, but the strategic dimension of RC has not been appreciated until recently, academia has failed to pay sufficient attention to the education of a community of high-quality system designers and configware programmers using such platforms. This has motivated a recent but ever growing interest in the question of educating specialists in this domain and this has also been recognized as a particularly difficult problem.

Data Transport:

Algorithmic Cleverness is missing. To-day experts with different backgrounds and diverging points of view are needed, not only for test and verification of modern designs, experts with different backgrounds and diverging points of view needed, if possible at all, which is expensive and substantially delays the product introduction. Although the economic necessity of RC and FPGAs has been widely recognized, the academic domain mainly missed the education of a sufficiently large share of highly qualified system designers and configware developers. Configware engineering and the programming of morphware requires much more computer science skills, rather than tricks from the culture of a particular application domain. A typical problem is the lack of algorithmic cleverness needed for software to configware migration. A new taxonomy of algorithms and architectures is needed, which extends the notion of algorithm beyond the time domain.

The harmful Monopoly of the von Neumann Paradigm. Our growing configware industry is still mainly ignored by our curricula - mainly, but not only, by our CS curricula. Commodity of the shelf (COTS) FPGAs have all two paradigms together and with several memory banks on board of the same chip. To master the collision of cultures we need transdisci-
Unified foundations needed. Meanwhile it has become evident that many fundamental problems are directly going across many application domains. We need to counter the current trend, where specialization is the target of education systems. We need to go toward interdisciplinary CS-related curricula for unifying the foundations of the discipline since it has become evident that fundamental problems are shared across several different application domains. We need a transdisciplinary approach toward hardware/configware/software co-design, not only in practice, but even more urgently for curricula in Electrical Engineering, Computer Engineering, Computer Science, and Information Technology.

Reconfigurable Computing Education. Although the target areas of all these consortia are the main application domains of reconfigurable resources, FPGAs are hardly mentioned in their recommendations. Our answer to this one-eyed viewpoint is our Reconfigurable Computing Education initiative also including all areas of supercomputing by founding a new workshop series: The 1st International Workshop on Reconfigurable Computing Education (RE education 2006) [137] on March 1, 2006, at Karlsruhe, in conjunction with the IEEE Computer Society Annual Symposium on VLSI (ISVLSI) on March 2 – 3, 2006 [138]. I would support founding an IEEE Computer Society task force, as well as of a GI / ITG Fachgruppe on Reconfigurable Computing.

8. CONCLUSIONS

Compared to the instruction-stream-based von Neumann paradigm, FPGAs and coarse-grained reconfigurable platforms for Reconfigurable Computing offer - in addition to drastic savings on the electric energy budget - speedup factors by several orders of magnitude. Their programming is RAM-based too, which in practice leads to a dual paradigm methodology by using Configware Engineering as the counterpart of Software Engineering. The 2nd RAM-based paradigm avoids most of the often serious communication bottlenecks coming along with concurrent instruction streams. The personal supercomputer is near, not only for the desktop, but also for a new road map to large scale supercomputing of up to now unthinkable highest performance dimensions. Our academic education system should accept this fascinating challenge, especially with new curricula in CS and CE for providing an integrating dual paradigm mind set to bridge the gap and to cure severe qualification deficiencies of our graduates. We need a unification in dealing with problems, which are shared across many different application domains. Interdisciplinary must become transdisciplinary.

9. LITERATURE


[16] K. Gaj, T. El-Ghazawi: Cryptographic Applications; RSSI Reconfigurable Systems Summer Institute, July 11-13, 2005, Urbana-Champaign, IL, USA URL: [17]


[19] [ASM = Auto-Sequencing Memory; DSP = Digital Signal Processing; EDA = Electronics Design Automation; ESL = Electronic System-Level Design; FIR = Finite Impulse Response; FPGA = Field-Programmable Gate Array; MAC = Multiply and Accumulate; PU = Processing Unit rDPA = reconfigurable Data Path Array; rDPU = reconfigurable Data Path Unit; rER = reconfigurable Element


[22] Y. Gu, et al.: FPGA Acceleration of Molecular Dynamics Computations; PCCM 2004 URL: [23]


[26] H. Singpiel, C. Jacobi: Exploring the benefits of FPGA-processing technology for genome analysis at Acconovis; ISCC 2003, June 2003, Heidelberg, Germany URL: [27]


[28] N. N. (Starbridge): Smith-Waterman pattern matching; National Cancer Institute, 2004


[34] http://helios.informatik.uni-kl.de/RCeduction/

[35] R. Porter: Creation on FPGAs for Feature Extraction; Ph.d. thesis; Queensland Un. of Technology Brisbane, Australia, 2001: [36]


[37] F. Chatalwala: Starbridge Solutions to Supercomputing Problems; RSSI Reconfigurable Systems Summer Institute, July 11-13, 2005, Urbana-Champaign, IL, USA


[39] M. Kuulusa: DSP Processor Based Wireless System Design; Tampere Univ. of Technology, Publ. No. 296; URL: [40]


[46] W. Nebel et al.: PISA, a CAD Package and Special Hardware for Pixel-oriented Layout Analysis; ICCAD 1984


[49] Rick Kornfeld (personal communication)


[57] N. Petkov: Systolic Parallel Processing; North-Holland; 1992


[65] A. Cantle: Is it time for von Neumann and Harard to Retire?; RSSI Reconfigurable Systems Summer Institute, July 11-13, 2005, Urbana-Champaign, IL, USA

[66] http://bwrc.eecs.berkeley.edu/Research/RAMP/


[74] Invited reprint of [73] in Future Generation Computer Systems 7 91/92, p. 181-198, North Holland


[77] C. Chang et al: The Biggascal Emulation Engine (Bee); summer retreat 2001, UC Berkeley


[82] N. Ambrosiano: Los Alamos and Surrey Satellite contract for Cibola flight experiment platform; Los Alamos National Lab, Los Alamos, NM, March 10, 2004 URL: [83]


[84] M. Gokhale et al.: Dynamic Reconfiguration for Management of Radiation-Induced Faults in FPGAs; Proc. IPDPS 2004


[87] N. N.: R. Associates Joins Nallatech's Growing Channel Partner Program; Companies are Using FPGAs to Reduce Costs and Increase Performance in Seismic Processing for Oil and Gas; Wireless World, May 2005
