Massively Parallel Computing on Silicon: SIMD Implementations

V.M. Brea
Univ. of Santiago de Compostela
Spain
GOAL

- Give an overview on the state-of-the-art of Digital on-chip CMOS SIMD Solutions, mainly Visual Processors, through papers found in the literature.
Outline

- Parallel Computing
- SIMD Computing
- Digital Solutions
- Conclusions
Outline

- Parallel Computing
- SIMD Computing
- Digital Solutions
- Conclusions
Trends in High Performance Computing
IEEE Circuits & Devices Magazine - January/February 2006

- Supercomputers - evolution
Trends in High Performance Computing
IEEE Circuits & Devices Magazine - January/February 2006

- **Supercomputers - Evolution**
  - Late 70’s- Early 80’s - Vector Systems
  - 80’s - Symmetric Multiprocessors (SMPs) - memory sharing
  - Late 80’s - Distributed memory computer system - overcoming the hw scalability limitations of shared memory
  - 90’s - Massively Parallel Processors (MPP) - 256 to 10000 processors, scalable (distributed memory)
  - Today - off-the-shelf components, clusters of PC’s or workstations
  - Today and tomorrow - Grid computing
Outline

- Parallel Computing
- SIMD Computing
- Digital Solutions
- Conclusions
A System for Evaluating Performance and Cost of SIMD Array Designs

J. of Parallel and Distributed Computing
60, 217-246, 2000

- Goal: efficient evaluation of SIMD arrays with respect to complex applications while accounting for operating frequency and chip area

- Remark: The first 10% of the design cycle determines nearly 80% of a system’s cost ... many alternatives in SIMD realization
A System for Evaluating Performance and Cost of SIMD Array Designs
J. of Parallel and Distributed Computing 60, 217-246, 2000

- Design Alternatives in SIMD Machines (I)
  - Asymmetric- Control and SIMD array
  - PE Design- CPU, datapath: 1-, 8-, 32-bits …
  - PE Memory Hierarchy- on- and/or off-chip memory, cache (Propagating templates)
  - PE Nearest Neighbor Communication- NEWS network
A System for Evaluating Performance and Cost of SIMD Array Designs
J. of Parallel and Distributed Computing
60, 217-246, 2000

- Design Alternatives in SIMD Machines (II)
  - General Communication - additional inter-PE communication networks
  - Feedback from array to controller – global OR mechanisms, count of corresponding PEs, flags …
  - Array Instruction Issue-
    - Speed at which instructions can be issued
    - The latency of the instructions from the issuer to the PEs
    - The skew in instruction distribution to the various PEs in the array
  - Mapping data to processing elements - one-to-one correspondence, virtual PEs (VPEs)
A System for Evaluating Performance and Cost of SIMD Array Designs
J. of Parallel and Distributed Computing 60, 217-246, 2000

- Variable Parameters at PE level:
  - PE datapath from 1 to 64 bits
  - FP units, possibly shared among PEs
  - ALU complexity, multiplier and/or divider
  - Nearest-neighbor communication network parameters
  - Number of internal buses and parts in the register file
  - Local indexing
  - Per PE caching
  - Simple pipelined, PE datapaths
A System for Evaluating Performance and Cost of SIMD Array Designs
J. of Parallel and Distributed Computing 60, 217-246, 2000

- Variable Parameters at Array level:
  - Size- 256 to 16 million PEs
  - Load/store mechanism
  - Communication networks- broadcast …
  - Feedback, including OR and count
Outline

- Parallel Computing
- SIMD Computing
- Digital Solutions
- Conclusions
Digital Implementation of Cellular Sensor-Computers

- Contributions
  - Bit-sliced digital implementation method of the sensor-computer. For 0.18 um and below, digital is a viable alternative to analog technology
  - Photocells and A/D converters shared among several Pes (4) without performance degradation
  - Note: The chip, named as XENON, has not been tested …
Digital Implementation of Cellular Sensor-Computers

- The Processing Element
  - Look-up-Table (LUT) for morphology, hit-and-miss, and whatever bitwise logic operations
  - Arithmetical datapath- full-adder for addition/subtraction-based arithmetic
  - Independent path; both can work in parallel
The Processing Element

- 64 bits for each pixel: 7 pieces of 8 bit variables, LAM, plus 8 Boolean variables with random access
- Neighborhood interconnections: directly from the memory, or through the crossbar switch
- Global buses for Global interconnection, and Data condition flags
Digital Implementation of Cellular Sensor-Computers


Figure 1. Cell architecture in the digital implementation.
Digital Implementation of Cellular Sensor-Computers

Figure 2. The processing block.
Digital Implementation of Cellular Sensor-Computers


Figure 3. The storage units and the local area interconnection.
Digital Implementation of Cellular Sensor-Computers

- Sensor Signal AD Conversion
  - Single-slope-type AD conversion- the slope is provided by a global analog signal
  - The digital staircase figure is calculated with the processor itself, particularly with the arithmetic part
  - Need of an S/H
Digital Implementation of Cellular Sensor-Computers

Figure 7. Sensor control model of the integrated sensor–processor architecture.
Digital Implementation of Cellular Sensor-Computers

- Sensing Schemes
  - Local integration time control feature
  - Feedback local intensity information to the sensing medium to increase the integration time in the dark regions, and decrease it in the bright regions
  - As a consequence, realistic compressed dynamic range images
Digital Implementation of Cellular Sensor-Computers

- Computational Throughput
  - Higher for bit-serial arithmetic than for bit parallel
  - Straightforward pipeline structures
  - High speed logic
Digital Implementation of Cellular Sensor-Computers

Chip Data- 0.18 um

- Clock- 100MHz
- Cell size- 33x33 um
- Processor kernel size- 33000 \( l^2 \), being \( l \) the feature size
- Sensor- 5x5 um
- AD and accompanying circuitry 3000 \( l^2 \)
- Note: there are no measurements!!!
A Digital Vision Chip Specialized for High-Speed Target Tracking

- Goal
  - High-speed target tracking including multitarget tracking with collision and separation
A Digital Vision Chip Specialized for High-Speed Target Tracking

Features

- Global feature extraction- I/O bottleneck, scalar features extracted from the array
- Need of high speed global operations
- The chip calculates moments (global information)
- Serial communication through nearest-neighbor connection, and global operations performed with bit serial cumulative adders at the end of rows and columns
- Sensor output binarized- time-controled
- Datapath- bit-serial
A Digital Vision Chip Specialized for High-Speed Target Tracking
A Digital Vision Chip Specialized for High-Speed Target Tracking

- Tracking
  - Local operations- logic and arithmetic
  - Global operations- extraction of moments
A Digital Vision Chip Specialized for High-Speed Target Tracking  

B. Tracking

Target tracking using the self-windowing algorithm [12] is handled by hardware. Self-windowing is an algorithm to search only inside the object and its one-pixel neighbors. The algorithm uses the property of high speed vision that the change between the images of two successive frames is very small. Details of this algorithm are shown below.

We define the input image at time $k$ as $f_k(x, y)$, the tracking image as $g_k(x, y)$, and the window image as $W_k(x, y)$:

1) Initial determination of the object

$$g_0(x, y) = \text{(the image of the object at the time 0)}.$$  

2) Making the window

$$W_{k+1}(x, y) = g_k(x, y) \cup g_k(x + 1, y) \cup g_k(x - 1, y) \cup g_k(x, y + 1) \cup g_k(x, y - 1).$$  

3) Extracting the target

$$g_{k+1}(x, y) = W_{k+1}(x, y) \cap f_{k+1}(x, y).$$
A Digital Vision Chip Specialized for High-Speed Target Tracking

<table>
<thead>
<tr>
<th>Specifications of the Second Prototype Chip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
</tr>
<tr>
<td>Chip area</td>
</tr>
<tr>
<td>Number of pixels</td>
</tr>
<tr>
<td>Pixel area</td>
</tr>
<tr>
<td>Photodiode area</td>
</tr>
<tr>
<td>Max Clock Freq.</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Power Consumption</td>
</tr>
</tbody>
</table>
A Dynamically Reconfigurable SI MD Processor for a Vision Chip

Goal

- To overcome the poor performance in global operations shown by conventional SI MD image processors
- Contribution- SI MD vision chip that reconfigures its hdw dynamically by chaining processing elements
A Dynamically Reconfigurable SIMD Processor for a Vision Chip

- Features
  - Early visual processing- edge detection, smoothing …
  - Global feature calculation- calculate moments and output them as scalar values
  - Scalar feature values such as summations at high speed
  - Fast communication between distant PEs
  - Grain size of the PE and network structure dynamically reconfigurable
  - Key to high speed- n PEs chained,
    - The ALU behaves as one n-bit ALU, memory capacity multiplied by n
    - Number of instructions reduced
A Dynamically Reconfigurable SIMD Processor for a Vision Chip

Fig. 2. Structure of the PE.

Fig. 3. Examples of PE chains.
Near-Sensor Image Processing: A New Paradigm

- Goal - contributions
  - Physical properties of the image sensor itself is utilized to do part of the signal processing task
  - Analog temporal behavior of photodiodes combined with thresholding amplifiers
  - Non-linear operations like median filter or some other tasks like convolution
  - Adaptivity to different light levels
  - Moments and shape factors
The Photodiode Sensor

- Inversely biased diode
- Generated photocurrent used to discharge a small capacitor that has been initially charged to a nominal level
  - $U = U_0 - kI_t$
- Two ways to proceed
  - 1st- exposure time constant, change the threshold level until it is equal to the photodiode voltage (CCD)
  - 2nd- keep threshold constant and measure the time (NSIP)
Near-Sensor Image Processing: A New Paradigm

Fig. 2. Photodiode element and sense amplifier.
Near-Sensor Image Processing: A New Paradigm

- The Processing Element
  - Based on the traditional bit-serial SIMD architecture
  - Extensions to handle image operations of global nature
  - PE register-accumulator oriented
  - Multilevel, gray-scale operations performed bitwise using the registers for intermediate storage (bit-serial operations)
  - Neighbor communication- NEWS
  - Global status value- COUNT
Near-Sensor Image Processing: A New Paradigm

Fig. 4. One NSIP SPE.
Some operations: Location of the highest intensity pixel- solution sense the array of photodiodes continuously after precharge. The first photodiode to pass the threshold level will correspond to the location of highest intensity.

Fig. 5. Photodiode array when brightest intensity reaches $U_{ref}$. 
Some operations- Median filter- for each neighborhood of three, the second pixel to cross the threshold is the median pixel.
Near-Sensor Image Processing: A New Paradigm  

- Global feature extraction


Some other solutions

- SIMD solutions
  - SRAM or any other memory-based chip (still to read something on this issue)
  - CAM-based (still to read something on this issue)
  - FPGA solutions
FPGA-Based Real-Time Optical-Flow System

Goal- contributions

- Pipelined optical-flow processing system - virtual motion sensor
- Conventional system - conventional camera acting as a front-end supporting by an FPGA processing device
- FPGA programmability permits easy change of parameters to adapt to different conditions like speed, environment or light intensity
- Stand-alone- 30 Hz with 320x240 pixels
- Hdw- Stand-alone or PCI board hdw accelerator
  - RC1000-PP, Celoxica- Virtex 2000E-6 Xilinx FPGA
  - RC 200-PP, Celoxica- XC2V2000-4 FPGA
Conclusions

- Early visual processing
  - SIMD solutions with CMOS on-chip implementations for local and global operations
    - Digital
    - Analog
  - FPGA-based solutions

Things to explore: SRAM and memory-based approaches along with how our techniques could be adopted in conventional SIMD architectures (not CNNs)

Contributions on SIMD: At the first stage of the design cycle, where the ideas have significant impact at layout level. Look at new paradigms, techniques, algorithms …