

Datapath Design for a VLIW Video Signal Processor



Jason Fritts

Princeton University

Department of Electrical Engineering

Co-Authors: Andrew Wolfe, Santanu Dutta, Edil S. T. Fernandes

Our Research

**Princeton is designing an ultra-high performance,
single-chip, VLIW Video Signal Processor.**

Architecture

Compiler

VLSI Design

Benchmarking

- High-level language programmability is the key to innovative applications
- Programmable VSPs are just becoming practical

2

Digital Video Applications

- **Recent explosion of multimedia industry**
 - video compression and teleconferencing
 - Web
- **Microprocessor video support**
 - MMX
- **Video is critical for a wide range of applications**
 - entertainment
 - communication
 - education
 - image understanding, robotics, surveillance

Most current activity focused on compression standards

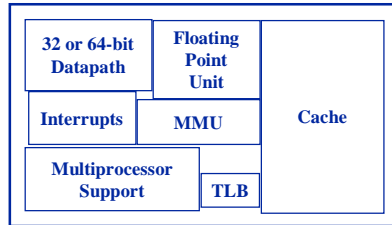
3

The Raw Numbers - Why VSP is hard

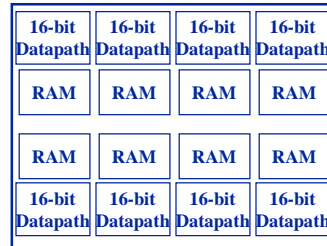
- **Limited multimedia today**
 - Today's "multimedia grade" digital video is a toy
 - 128x96 pixel, 10fps - compressed
 - playback only - requires ~3.6MIPS
- **Major applications will expect broadcast grade video**
 - 720x480 pixel, 30fps
 - decompression - requires $\sim 10^8$ ops/s
 - compression - requires $\sim 1 \times 10^{10}$ ops/s
 - image understanding - requires $\sim 10^{11}$ - 10^{12} ops/s
- **HDTV-grade video will be important.**
 - 4 to 5 times more data than current broadcast grade.

4

Why not Microprocessors?



Microprocessor

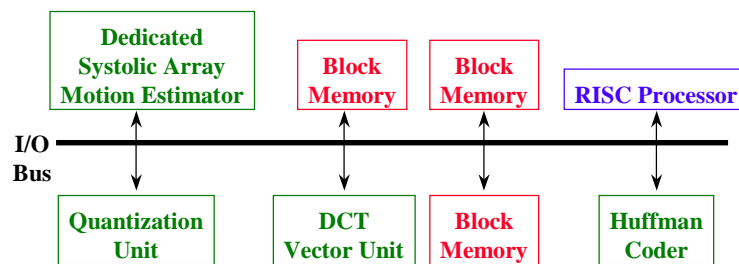


VSP

- **Too much logic for supporting:**
 - operating systems
 - large data types
 - memory hierarchy
- **Microprocessors are too unpredictable**
- **VSP uses extra transistors for fast ALUs and fast RAM**

5

Today's VSPs



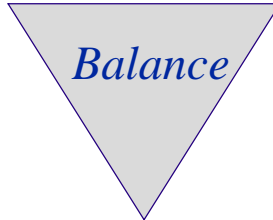
- **Dedicated, fixed-function VSPs - example: encoder**
- **Special purpose, custom-designed logic**
- **Hand programmed via microcode**

6

Programmable VSP Challenge

Throughput

- fast clock speed
- high parallelism
- high utilization



Storage

- large on-chip memory
- large register file
- efficient memory I/O

Programmability

- high connectivity
- regular arrangement
- optimizing compiler

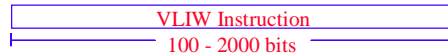
7

Project Goals

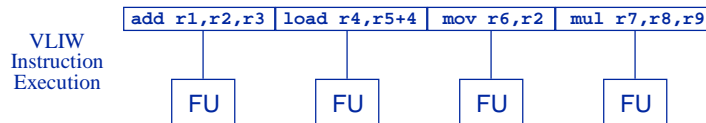
- **High-performance single-chip VSP**
 - >20 GOPS peak performance
- **Buildable by end of '90s**
 - estimate .25 μ rules, 500MHz+ clocks, 400mm² die capability
- **Efficiently HLL programmable**
 - parallization performed by compiler
- **Flexible**
 - high performance on a wide range of real applications
- **Meet real-time constraints**

8

VLIW Architecture



- **Statically scheduled Load-Store architecture**
- **Multiple functional units**
- **Wide instructions specify many independent operations**
- **Fine-grain parallelism within each instruction**
 - processor concurrently executes all operations in instruction



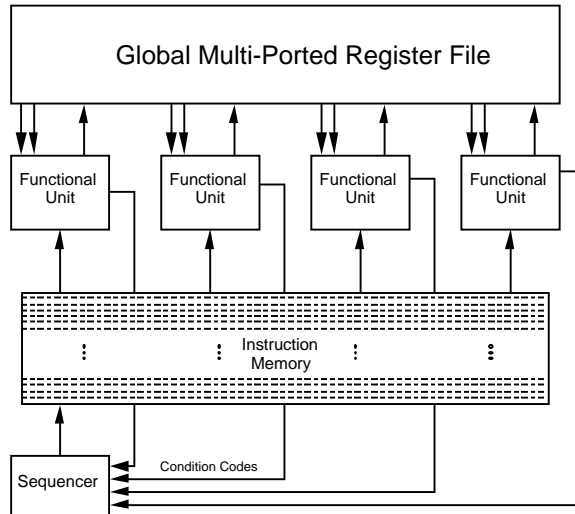
9

VLIW Advantages

- **Statically scheduled**
 - no run-time scheduling or arbitration
 - fast, simple architecture
- **Compiler technology is relatively mature**
 - locates and extracts parallelism from the program
 - schedules operations to exploit available parallel resources
- **Global compiler optimization**
 - globally search for opportunities for parallelism
 - restructure the program for greater parallelism
- **Transistors are dedicated to fast ALUs & memories**
- **Early indications of performance advantages**
 - flexible programming paradigm
 - fast cycle time
 - dense VLSI layout

10

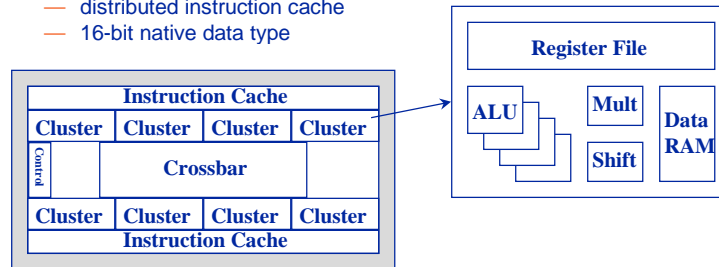
Canonical VLIW Model



11

Princeton VLIW VSP

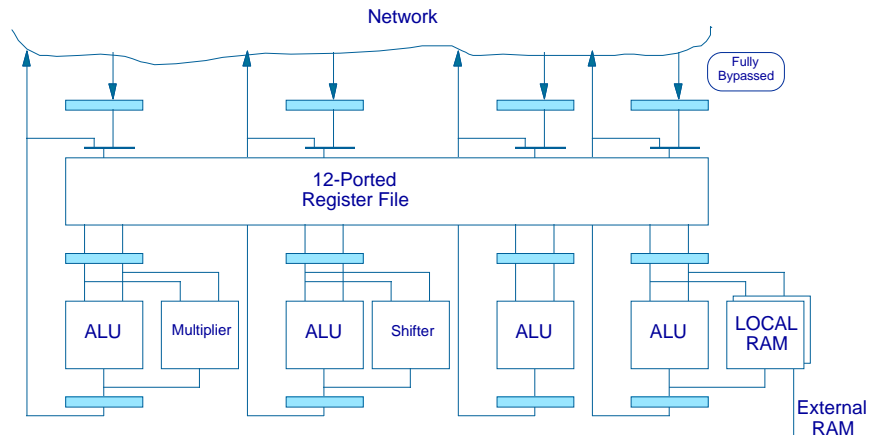
- many (>32) operations/cycle
- clusters of:
 - multiported register file
 - arithmetic units
 - local memories (double buffered)
- low latency crossbar between clusters
- distributed instruction cache
- 16-bit native data type



Single Long Instruction

12

Cluster Architecture



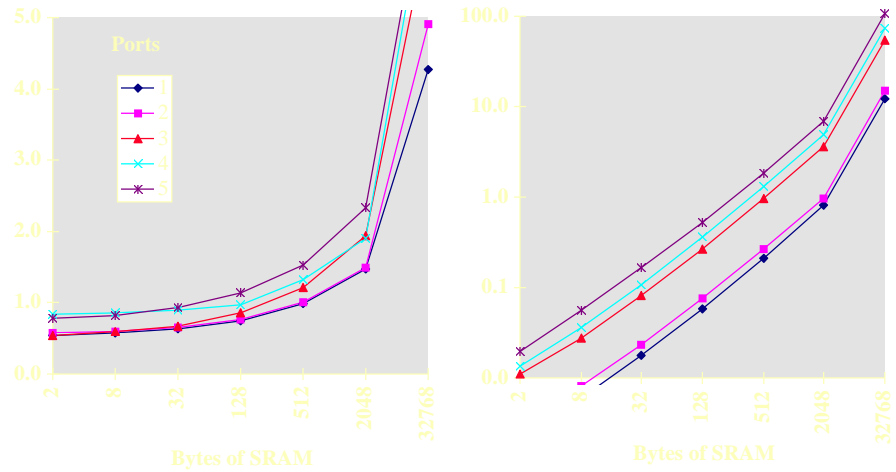
13

VLSI Modeling Methods

- **Perform transistor-level simulation of key modules:**
 - memories
 - register files
 - interconnect networks
 - functional units
- **VLSI simulation results indicate:**
 - cycle time tradeoffs
 - area tradeoffs
- **Target technology is .25 μ CMOS.**
 - Layout and simulation tools from Lucent Tech.
 - Tested process parameters.
 - Simulations at 3.0V, 25C
 - 2 levels of metal used for module layout
 - 2-3 remaining for inter-module connections
 - conservative design rules
 - Static CMOS

14

Memory Performance



15

Initial Cluster Size

- **Preliminary Model (I4C8S4)**
 - 12-ported register file - 128 registers @ 3 mm²
 - 4 ALUs @ 0.4 mm² each
 - 8-bit multiplier @ 1 mm²
 - shifter @ 0.5 mm²
 - 32K Local RAM @ 6.5 mm²- 13 mm²
 - 4-stage pipeline w/ full bypassing @ 0.4 mm²
- **Approx. 20-22 mm²/cluster**
- **8 clusters + 32x32 crossbar = ~180mm² datapath**
- **650 MHz operation possible**
- **32 operations/cycle -> 20.8GOPS peak**

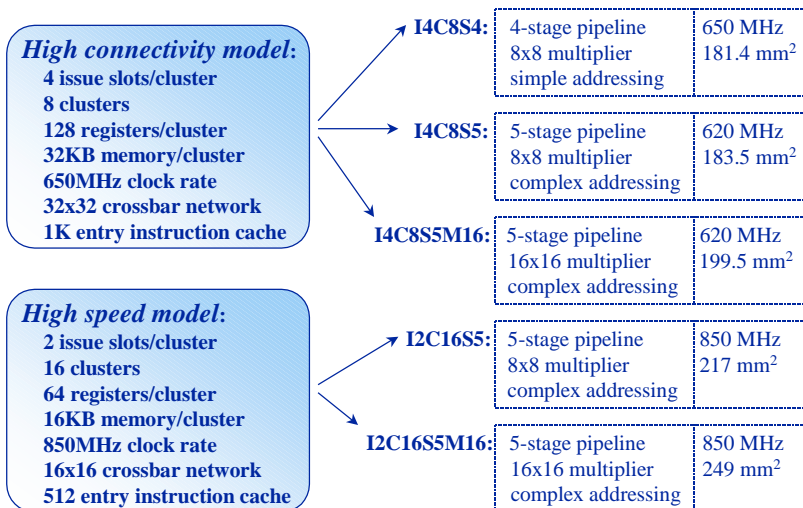
16

Primary Design Alternatives

- **Connectivity vs. Speed:**
 - higher connectivity generally leads to more effective compilers
 - higher speed may provide higher throughput
- **Simple vs. Complex Addressing Modes:**
 - complex addressing modes eliminate many address calculations
 - lengthens pipeline
 - creates load-use hazard
- **8x8 vs. 16x16 Multiplier:**
 - 8x8 multiplier requires numerous calculations for larger multiplies
 - 16x16 multiplier lengthens pipeline
 - requires more area
 - creates multiply-use hazard

17

VLIW VSP Candidate Architectures



18

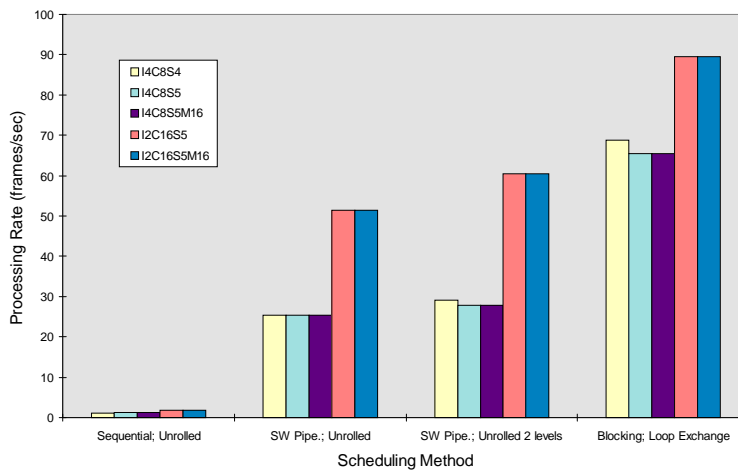
Hand-Scheduling Methods

- **VSP Kernels and their characteristics:**
 - Motion Estimation: **Full Search** and **3-Step Search**
 - considerable memory usage
 - DCT: **Traditional 2D DCT** and **Row-Column DCT**
 - heavy on multiplies and loads
 - Color Space Conv.: **RGB to 4:2:0 YCrCb**
 - boundary conditions
 - Lossless coding: **Variable Bit Rate encoder**
 - long dependency chains and limited parallelism
- **Compilation Techniques**
 - loop unrolling
 - list scheduling
 - software pipelining
 - blocking
 - SIMD

19

Simulation Data

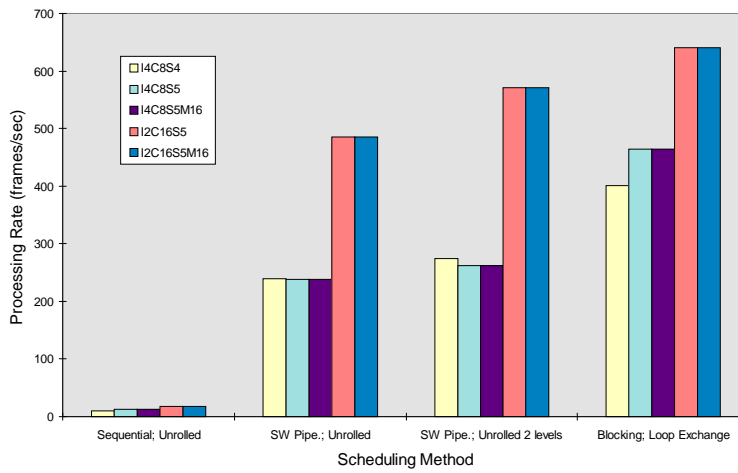
Full Search Motion Estimation



20

Simulation Data

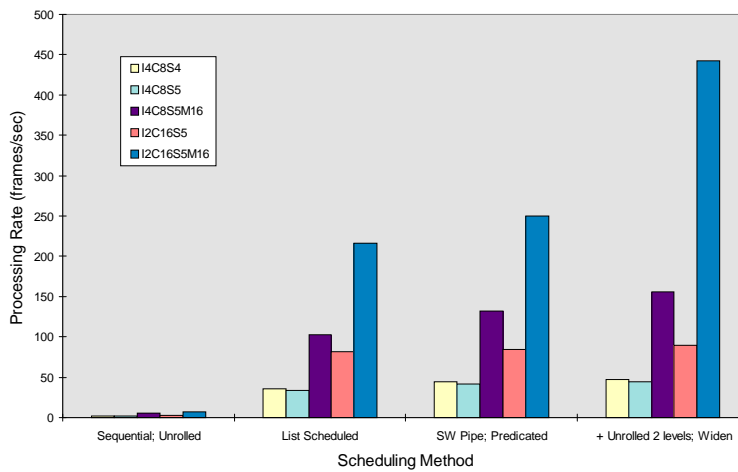
3-Step Search Motion Estimation



21

Simulation Data

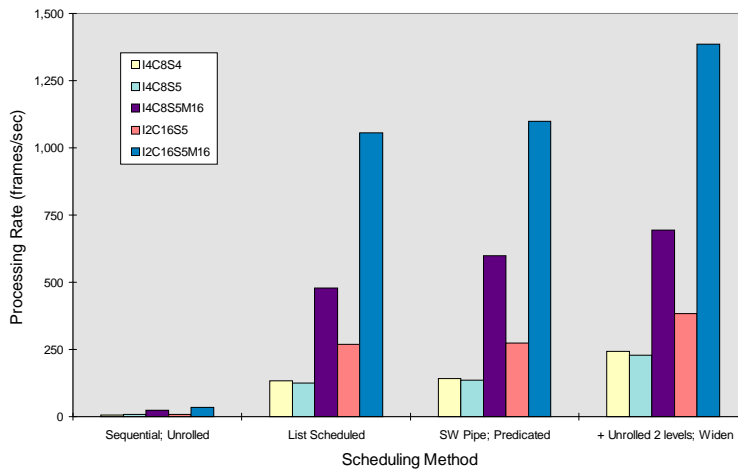
Traditional 2D DCT



22

Simulation Data

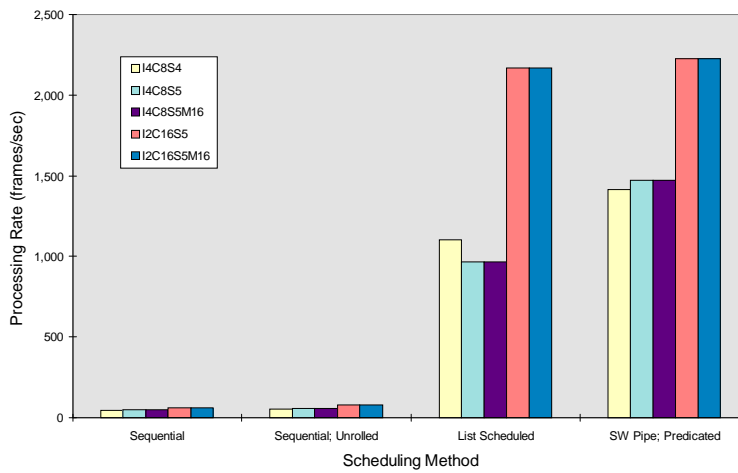
Row-Column DCT



23

Simulation Data

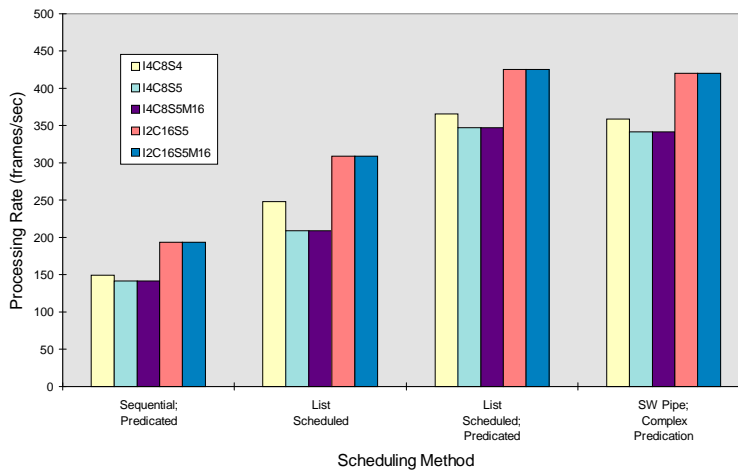
RGB to 4:2:0 YCrCb Color Space Conversion



24

Simulation Data

Variable Bit Rate Encoder



25

Scheduling Results

- **Raw results:**
 - CCIR601 full-motion search using 33%-44% of CPU time.
 - CCIR601 3-step search using 4.7%-7.5% of CPU time.
 - basic scheduling techniques produce speedups of 1.5x to 30.2x
 - aggressive scheduling can get another 2-3x
- **All candidate architectures were relatively balanced**
 - no model was particularly deficient in a specific resource
- **Design criteria results:**
 - complex addressing improved performance, but only minimally
 - 16x16 multiplier models were considerably faster on DCT
 - high-speed/low-connectivity model was typically much faster
- **Compilation Strategies**
 - loop unrolling particularly useful
 - dominant parallel method: SIMD across clusters/software pipelining within cluster

26

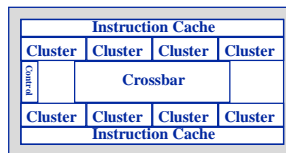
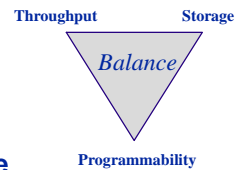
Continuing Research

- **VLSI development**
 - continued refinement and prototyping
 - arithmetic units, global structures, dynamic logic
- **Compiler development**
 - IMPACT-based C compiler
 - simple back-end optimizers initially
 - partitioning algorithm
- **Architectural simulation**
 - more/bigger examples
 - compiled code
- **Design tool development**
 - design exploration tool
 - functional simulator

27

Conclusions

- **Programmable Video Signal Processors**
 - key enabling technology for digital video
 - challenging demands
- **VLIW architectures are feasible for VSPs**
 - performance >20Gops peak seams feasible
 - achieved well-balanced design
- **Benefits of distributed cluster architecture**
 - large high-speed memory
 - large register file
 - only small sacrifice in connectivity



28