A 167-processor Computational Array for Highly-Efficient DSP and Embedded Application Processing

Dean Truong, Wayne Cheng, Tinoosh Mohsenin, Zhiyi Yu, Toney Jacobson, Gouri Landge, Michael Meeuwsen, Christine Watnik, Paul Mejia, Anh Tran, Jeremy Webb, Eric Work, Zhibin Xiao and Bevan Baas

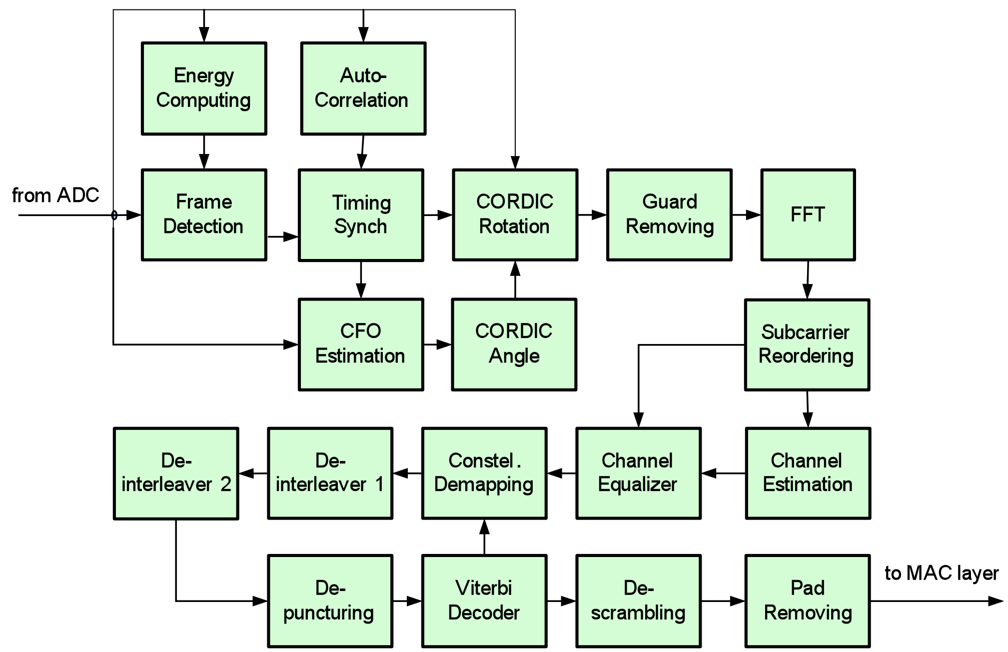
VLSI Computation Lab
University of California, Davis

Outline

- **Goals and Key Ideas**
- The Second Generation AsAP
  - Processors and Shared Memories
  - On-chip Communication
  - Dynamic Voltage & Clock Frequency
- Analysis and Summary

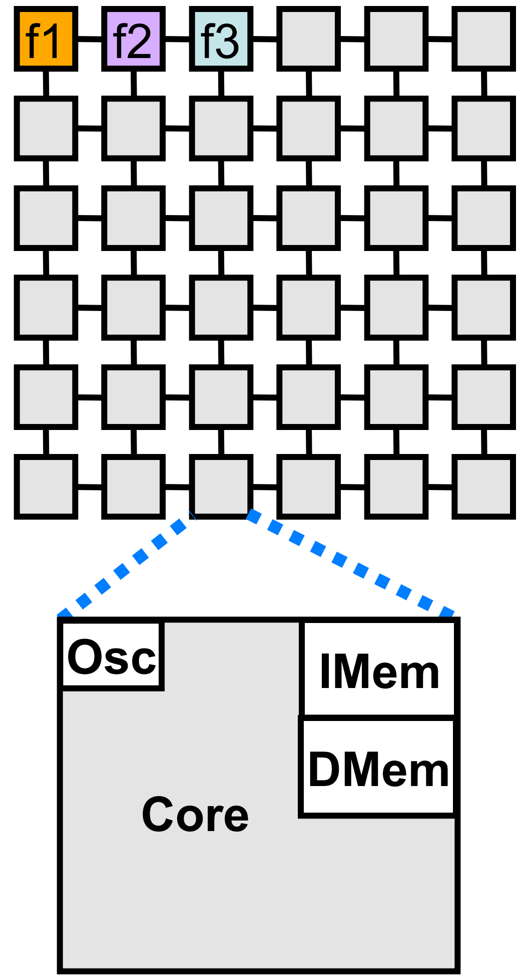
Project Goals

- Fully programmable and reconfig. architecture
- High energy efficiency and performance
- Exploit task-level parallelism in:
  - Digital Signal Processing
  - Multimedia
- Example: 802.11a Wi-Fi baseband receiver



Asynchronous **A**rray of Simple **P**rocessors (AsAP)

- Key Ideas:
  - Programmable, small, and simple fine-grained cores
  - Small local memories sufficient for DSP kernels
  - Globally Asynchronous and Locally Synchronous (GALS) clocking
    - Independent clock frequencies on every core
    - Local oscillator halts when processor is idle



Asynchronous **A**rray of Simple **P**rocessors (AsAP)

- Key Ideas, con't:
  - 2D mesh, circuit-switched network architecture
    - High throughput of one word per clock cycle
    - Low area overhead
    - Easily scalable array
  - Increased tolerance to process variations
- 36-processor fully-functional chip, 0.18 µm, 610 MHz @ 2.0 V, 0.66 mm^2 per processor [HotChips 06, ISSCC 06, TVLSI 07, JSSC 08,…]

Outline

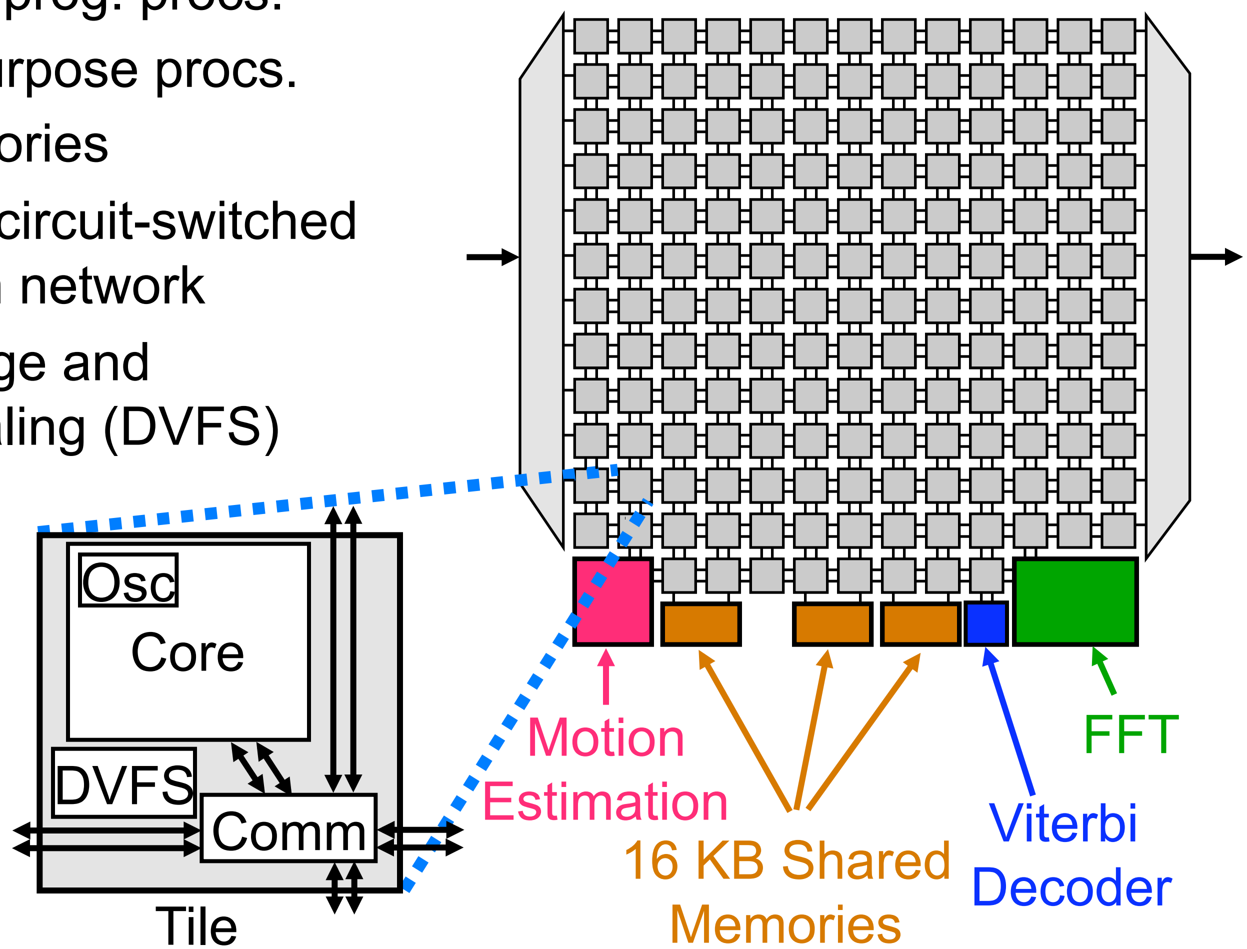
- Goals and Key Ideas
- **The Second Generation AsAP**
  - **Processors and Shared Memories**
  - **On-chip Communication**
  - **Dynamic Voltage & Clock Frequency**
- Analysis and Summary

New Challenges Addressed

1. Reduction in the power dissipation of
   - Lightly-loaded processors (lowering *Vdd*)
   - Unused processors (leakage)
2. Achieving very high efficiencies and speed on common demanding tasks such as FFTs, video motion estimation, and Viterbi decoding
3. Larger, area efficient on-chip memories
4. Efficient, low overhead communication between distant processors

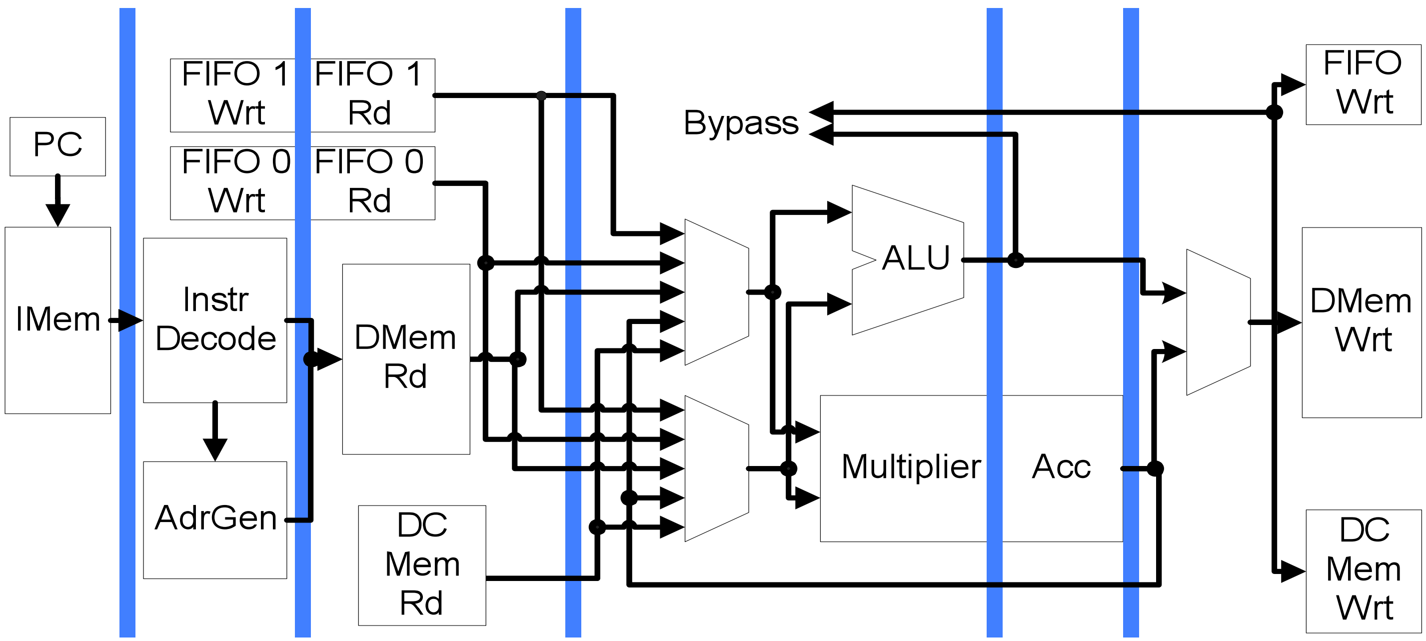
167-processor Computational Platform

- Key features
  - 164 Enhanced prog. procs.
  - 3 Dedicated-purpose procs.
  - 3 Shared memories
  - Long-distance circuit-switched communication network
  - Dynamic Voltage and Frequency Scaling (DVFS)



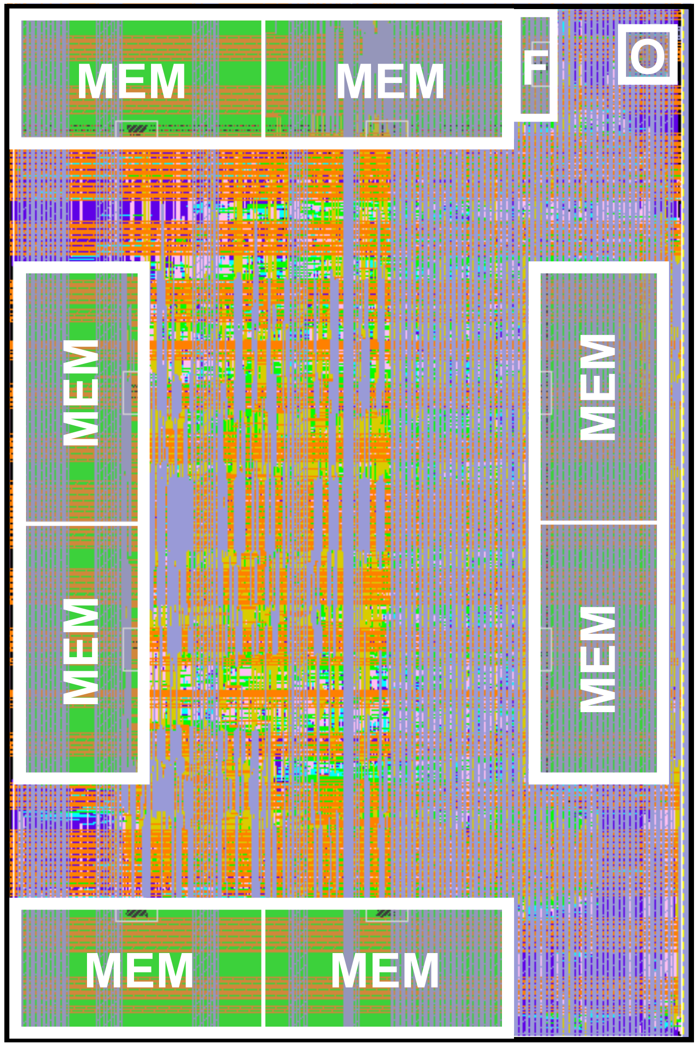
Homogenous Processors

- In-order, single-issue, 6-stage processors
  - 16-bit datapath with MAC and 40-bit accumulator
  - 128x16-bit data memory
  - 128x35-bit instruction memory
  - Two 64x16-bit FIFOs for inter-processor communication
  - Over 60 basic instructions and features geared for DSP and multimedia workloads



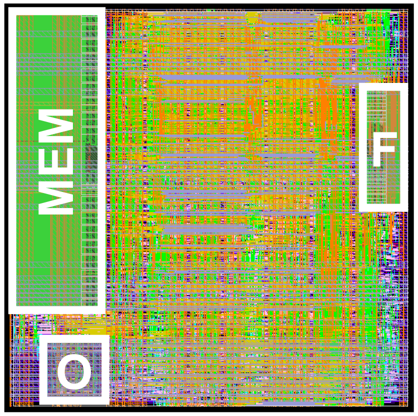
Fast Fourier Transform (FFT)

- Uses
  - OFDM modulation
  - Spectral analysis, synthesis
- Runtime configurable from 16-pt to 4096-pt transforms, FFT and IFFT
- 1.01 mm^2
- Preliminary measurements functional at 866 MHz, 34.97 mW @ 1.3 V
  - 681 M complex Sample/s with 1024-pt complex FFTs



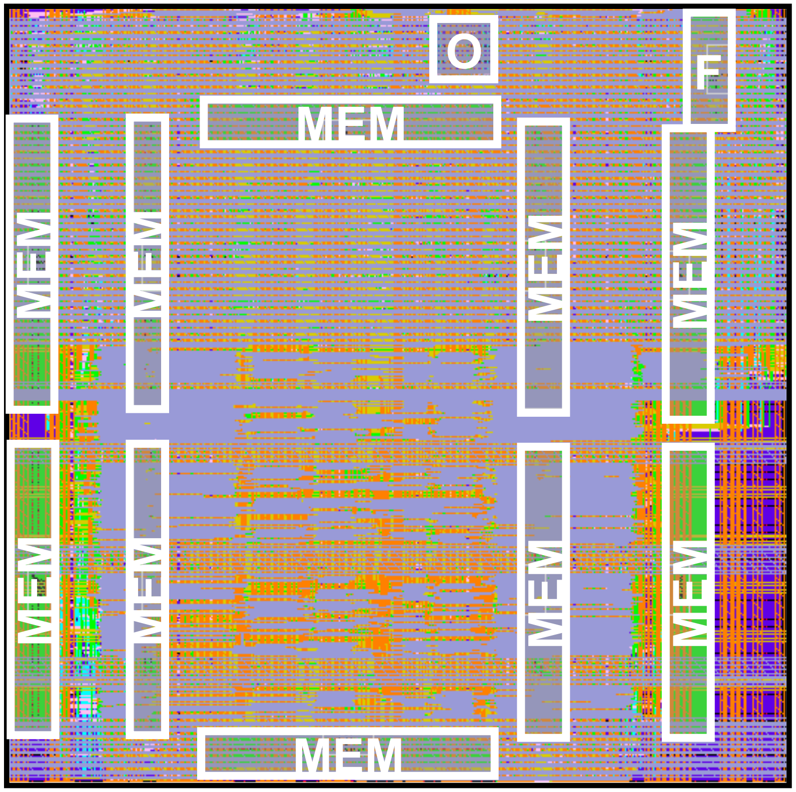
Viterbi Decoder

- Uses
  - Fundamental communications function (wired, wireless, etc.)
  - Storage apps; e.g., hard drives
- Decodes configurable codes up to constraint length 10 with up to 32 different rates
- 0.17 mm^2
- Preliminary measurements functional at 894 MHz, 17.55 mW @ 1.3 V
  - 82 Mbps at rate=1/2



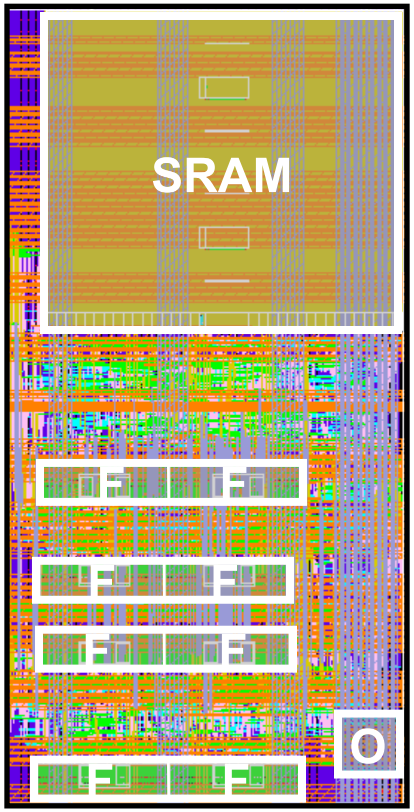
Motion Estimation for Video Encoding

- Uses
  - H.264, MPEG-2, etc. encoders
- Supports a number of fixed and programmable search patterns including all H.264 specified block sizes within a 48x48 search range
- 0.67 mm^2
- Preliminary measurements functional at 938 MHz, 196.17 mW @ 1.3 V
  - 15 billion SADs/sec
  - Supports 1080p HDTV @ 30fps



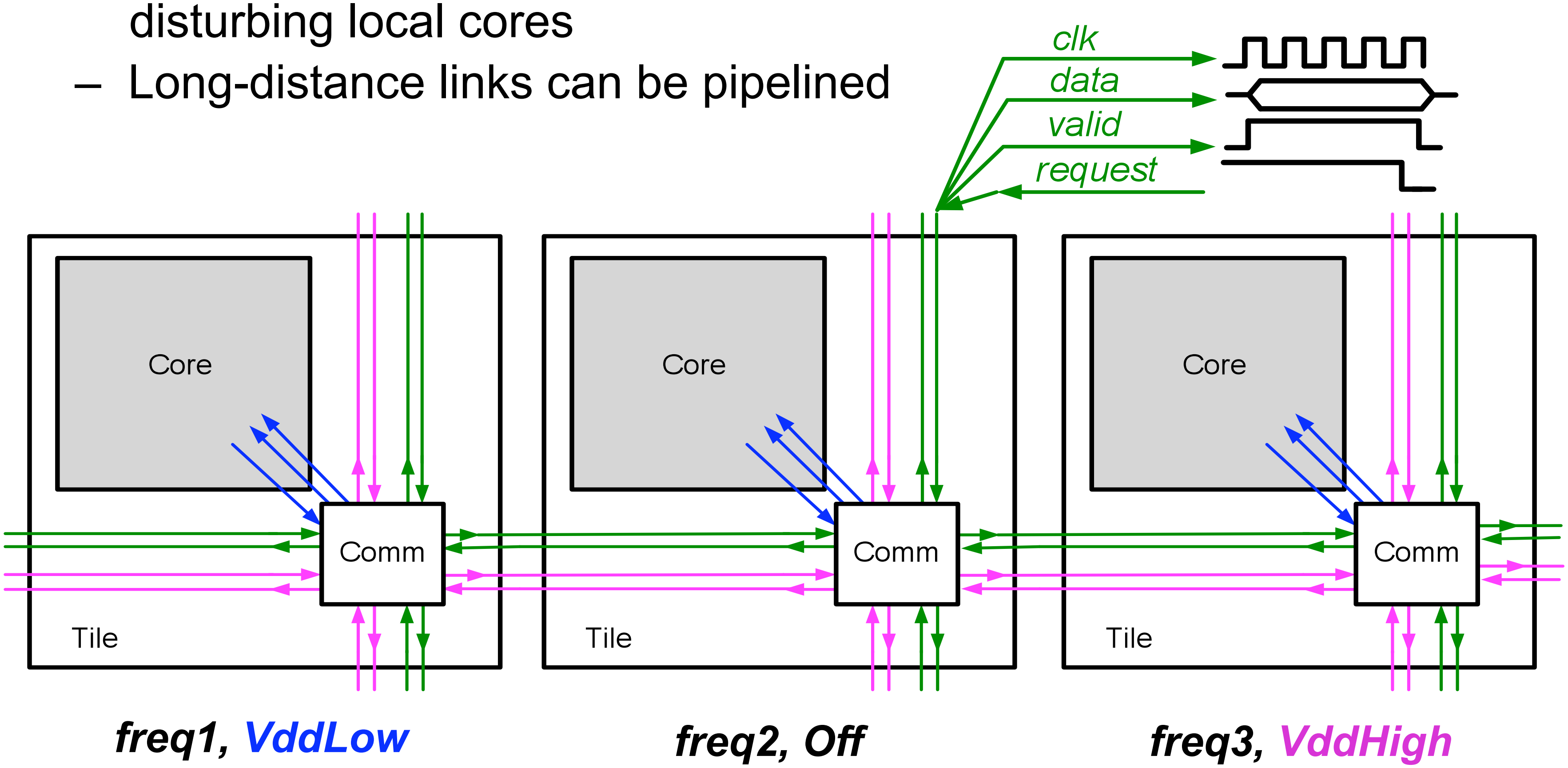
Shared Memories

- Ports for up to four processors (two connected in this chip) to directly connect to the memory block
  - Port priority
  - Port request arbitration
  - Programmable address generation supporting multiple addressing modes
  - Uses a 16 KByte single-ported SRAM
  - One read or write per cycle
- 0.34 mm^2
- Preliminary measurements functional at 1.3 GHz, 4.55 mW @ 1.3 V
  - 20.8 Gbps peak throughput



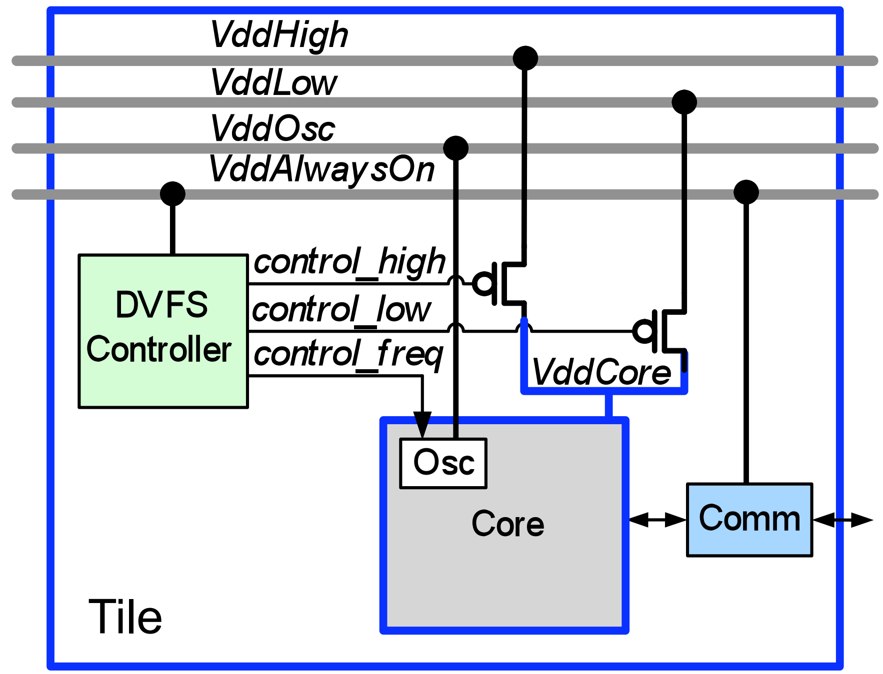
Inter-Processor Communication

- Circuit-switched source-synchronous comm.
  - 8 software controlled outputs and 2 configurable circuit-switched inputs (out of 8 total possible inputs)
  - Long-distance communication can occur across tiles without disturbing local cores
  - Long-distance links can be pipelined



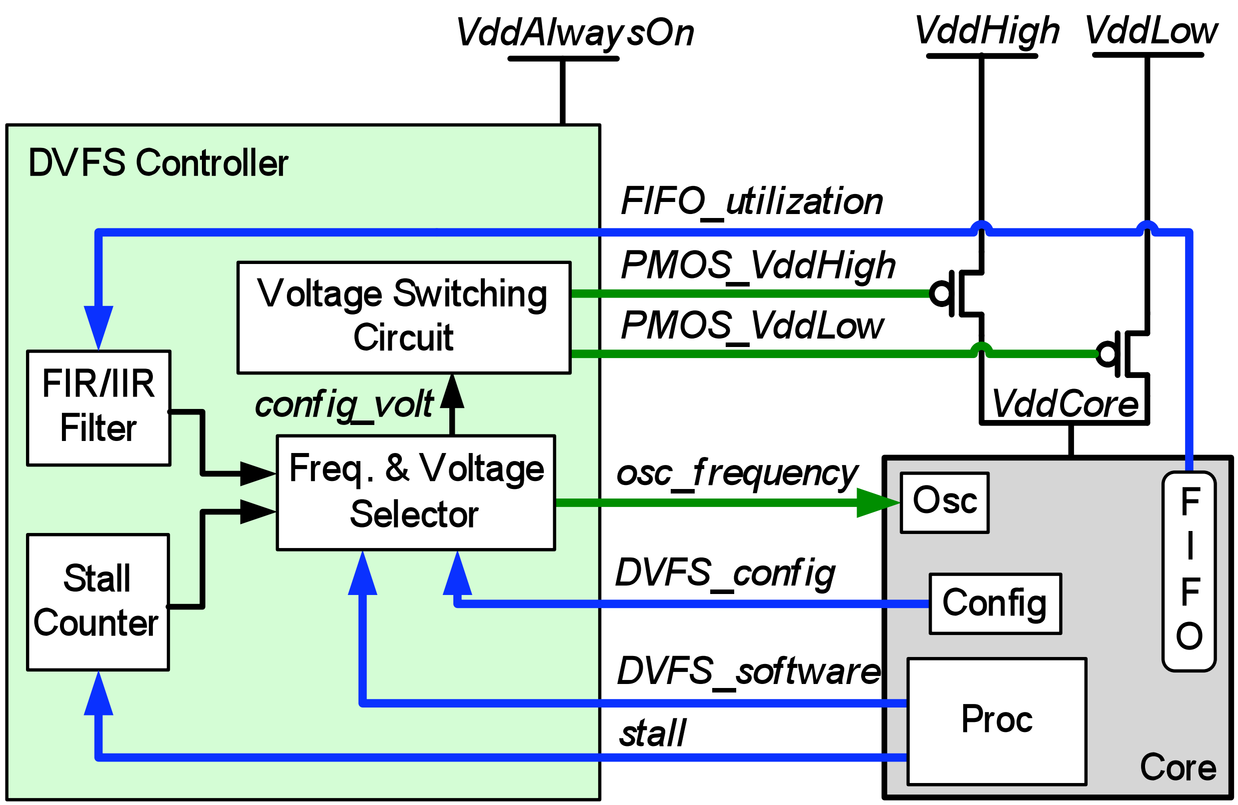
Per-Processor Dynamic Voltage & Clock Frequency

- Each processor tile contains a core that operates at:
  - A fully-independent clock frequency
    - Any frequency below maximum
    - Halts, restarts, and changes arbitrarily
  - Dynamically-changeable supply voltage
    - *VddHigh* or *VddLow*
    - Disconnected for leakage reduction
    - Each power gate comprises 48 individually-controllable parallel transistors
- *VddAlwaysOn* powers DVFS and inter-processor comm.



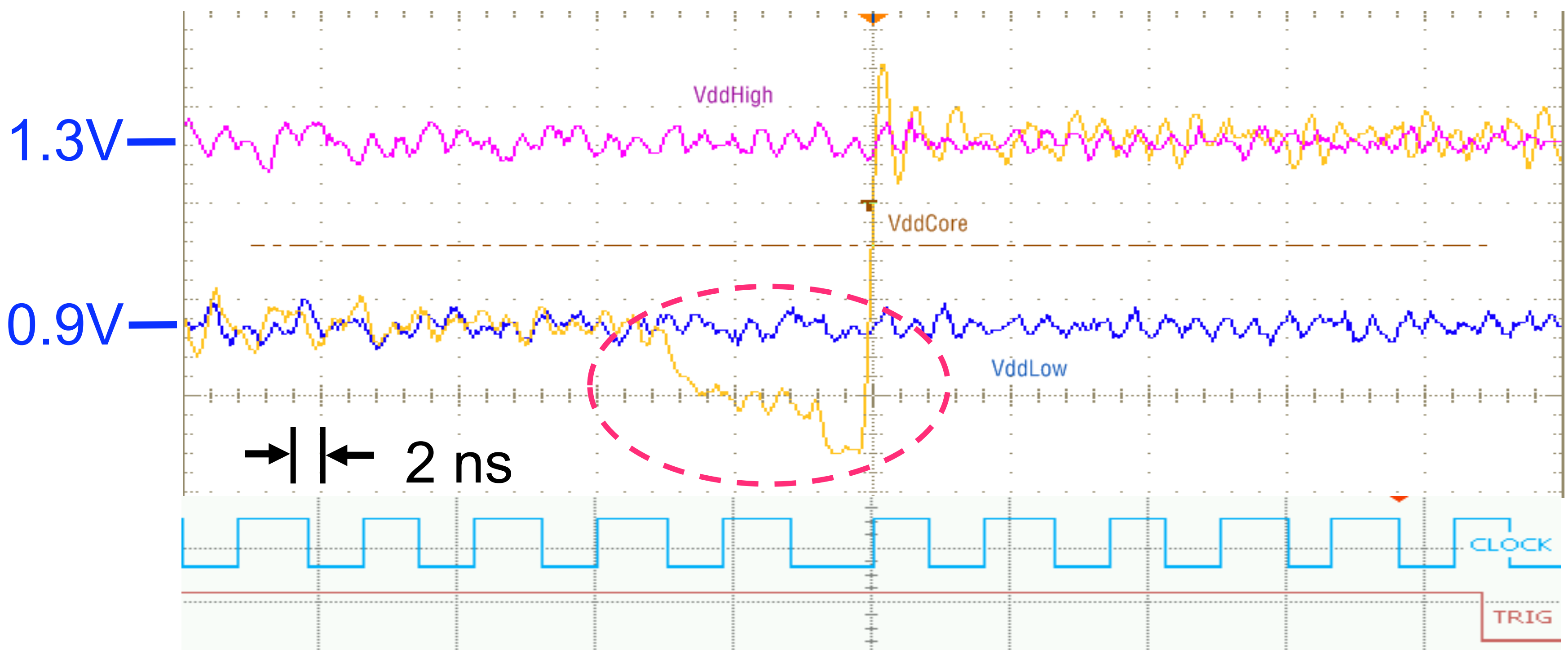
Dynamic Voltage & Frequency Controller

- Voltage and frequency are set by:
  - Static configuration
  - Software
  - Hardware (controller)
    - FIFO "fullness"
    - Processor "stalling frequency"



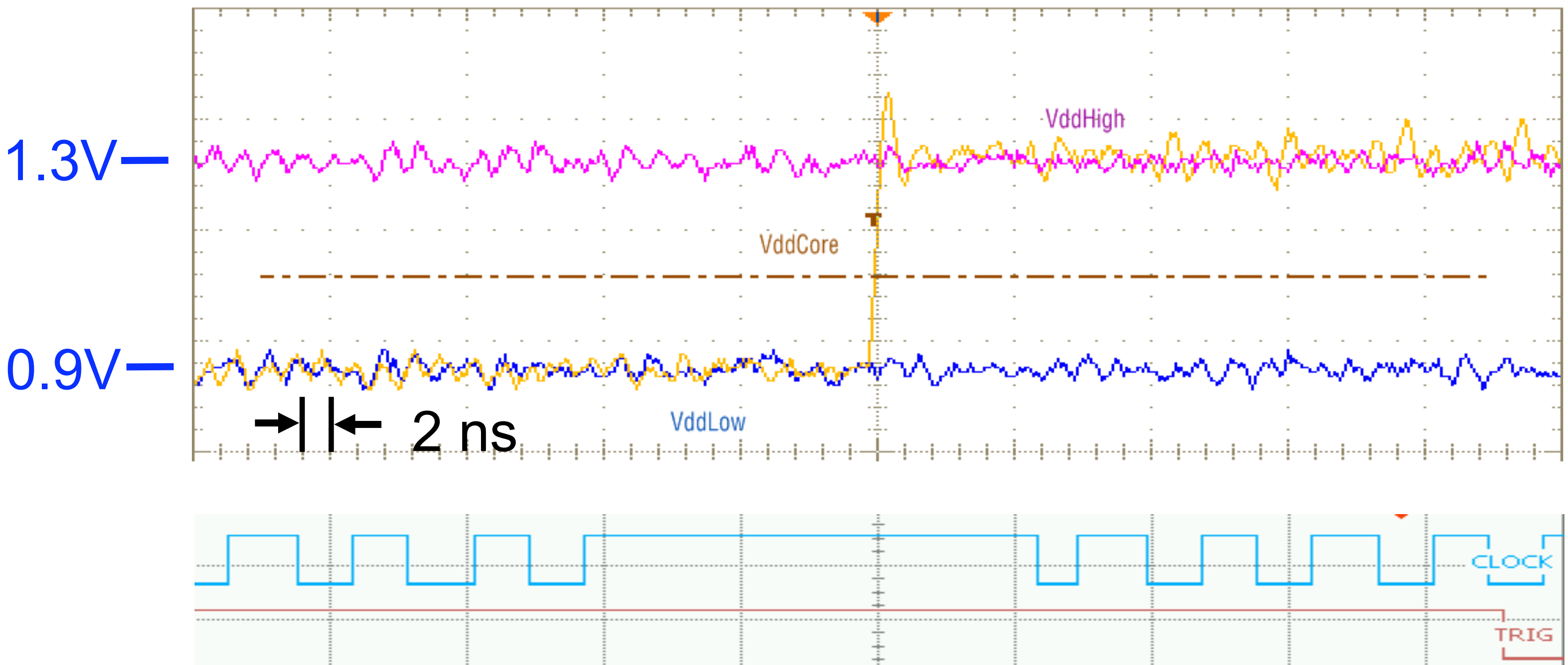
Measured Supply Voltage Switching

- Slow switching results in negligible power grid noise
- Early ***VddCore*** disconnect from ***VddLow*** with oscillator running results in a momentary ***VddCore*** voltage droop (circled below)



Measured Supply Voltage Switching

- Oscillator halting while ***VddCore*** disconnects from ***VddLow*** and connects to ***VddHigh*** results in a negligible voltage droop due to leakage



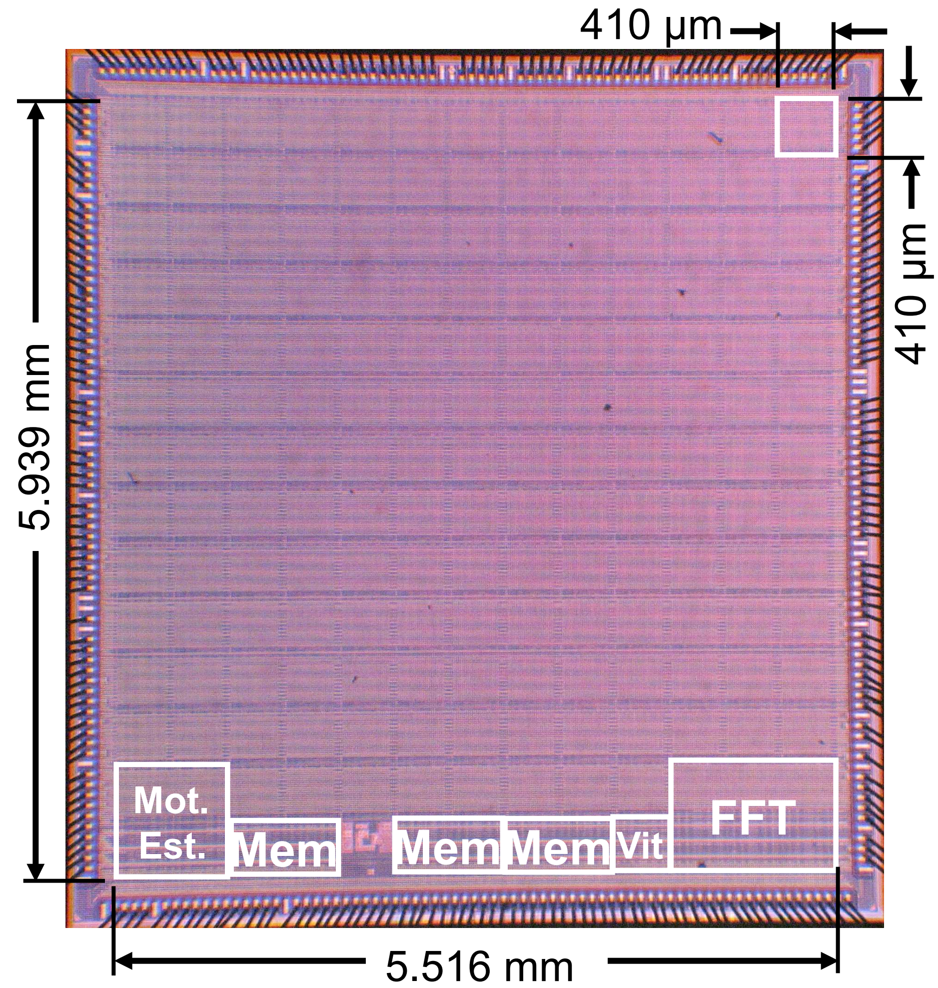
Outline

- Goals and Key Ideas
- The Second Generation AsAP
  - Processors and Shared Memories
  - On-chip Communication
  - Dynamic Voltage & Clock Frequency
- **Analysis and Summary**

Die Micrograph and Key Data

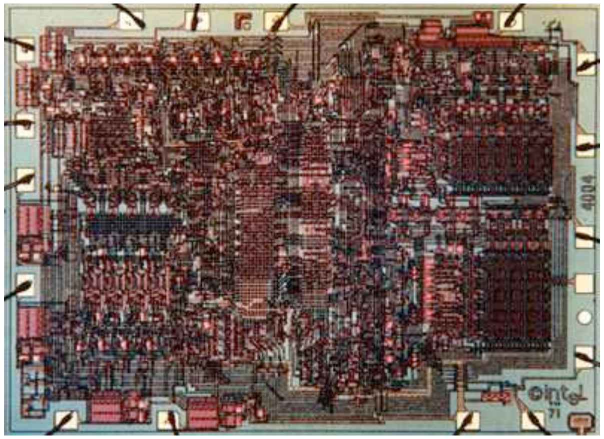
Single Tile	
Transistors	325,000
Area	0.17 mm^2
CMOS Tech.	65 nm ST Microelectronics low-leakage
Max. frequency	1.19 GHz @ 1.3 V
Power (100% active)	59 mW @ 1.19 GHz, 1.3 V
	608 μW @ 66 MHz, 0.675 V
App. power (802.11a rx)	16 mW @ 590 MHz, 1.3 V

55 million transistors, 39.4 mm^2



New Parallel Processing Paradigm

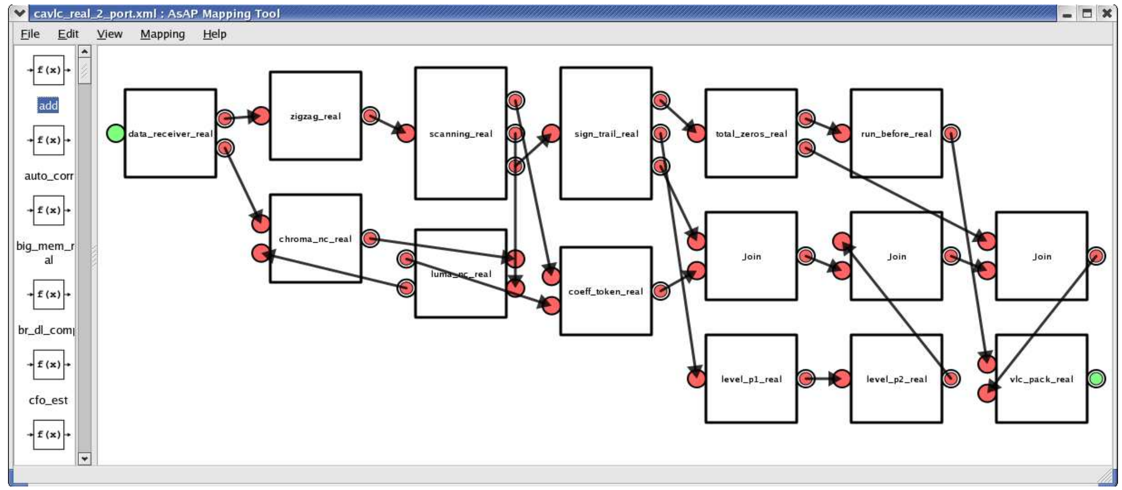
- Intel 4004 4-bit CPU, 1971
  – Utilized 2300 transistors
- The presented chip would have 2300 *processors* in 19.8mm x 19.8mm

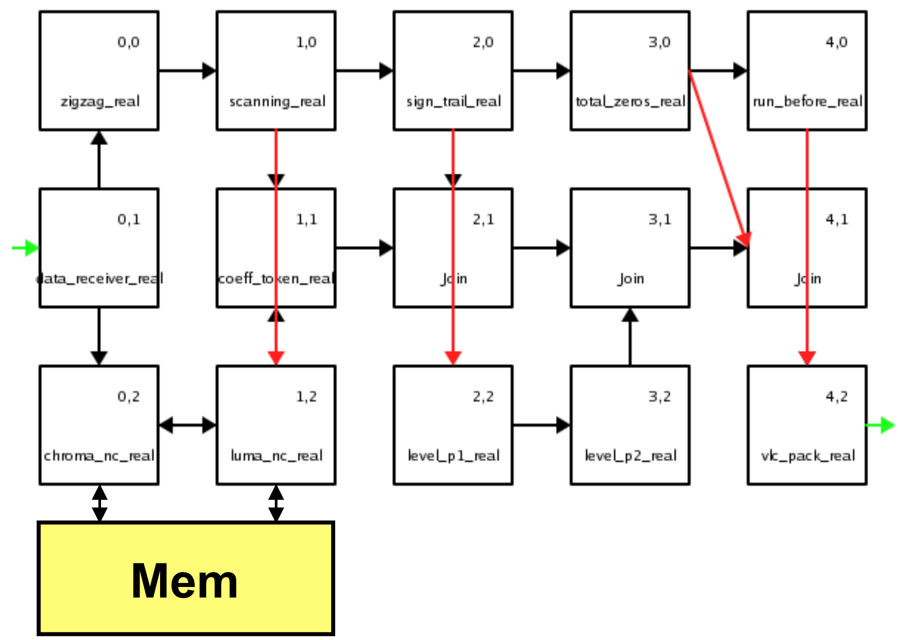


- New parallel processing paradigm
  – Enabled by numerous efficient processors
  – Focus on simplified programming and access to large data sets
  – Much less focus on load balancing or “wasting” processors for things like memories or routing data

H.264 CAVLC Encoder

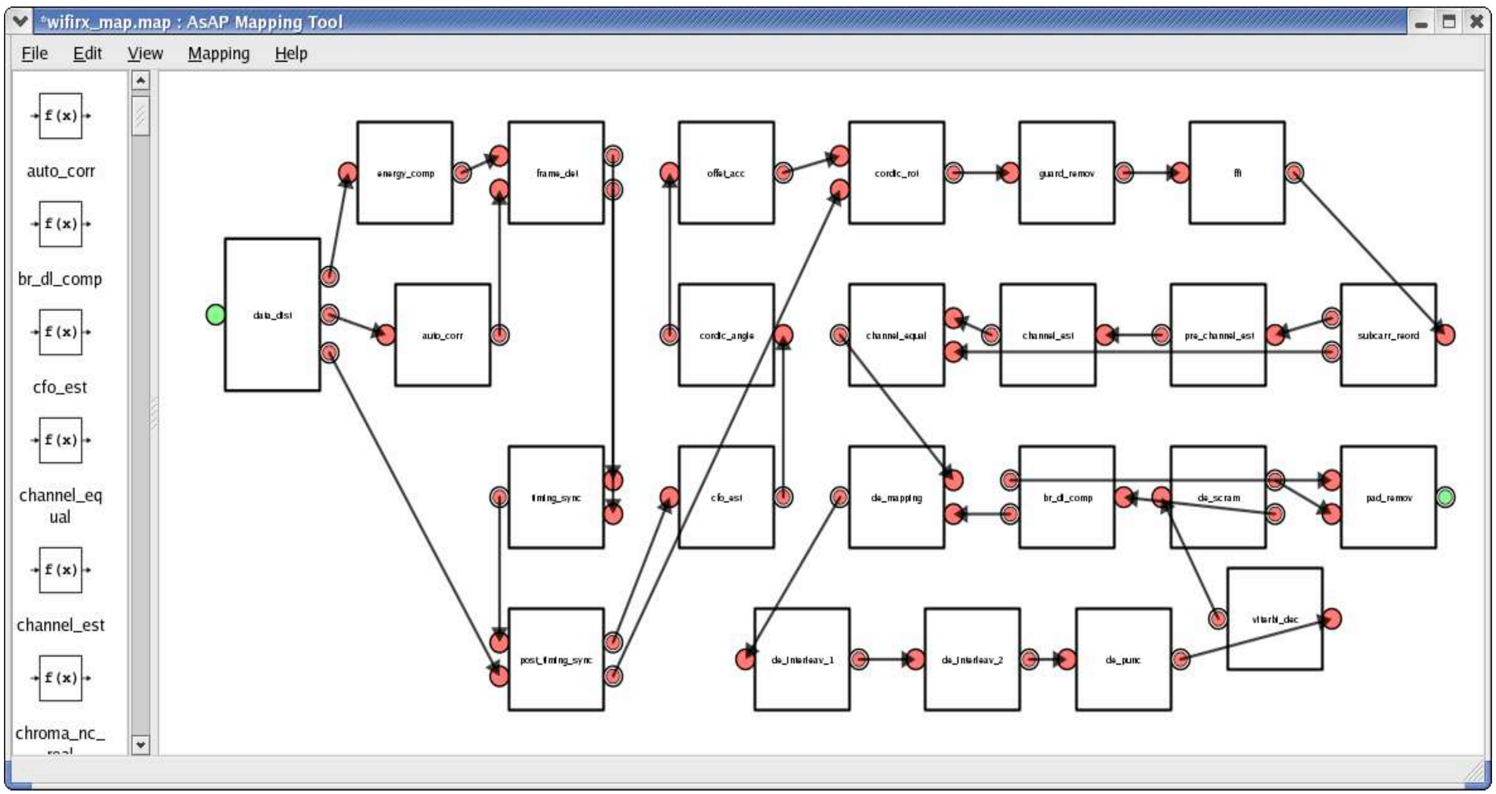
- Context-adaptive variable length coding (CAVLC) used in H.264 baseline encoder
- 15 processors with one shared memory
- 30fps 720p HDTV @ 1.07GHz
- ~1.0-6.15 times the throughput of TI C62x and ADSP BF561 (scaled to 65 nm, 1.3 V)





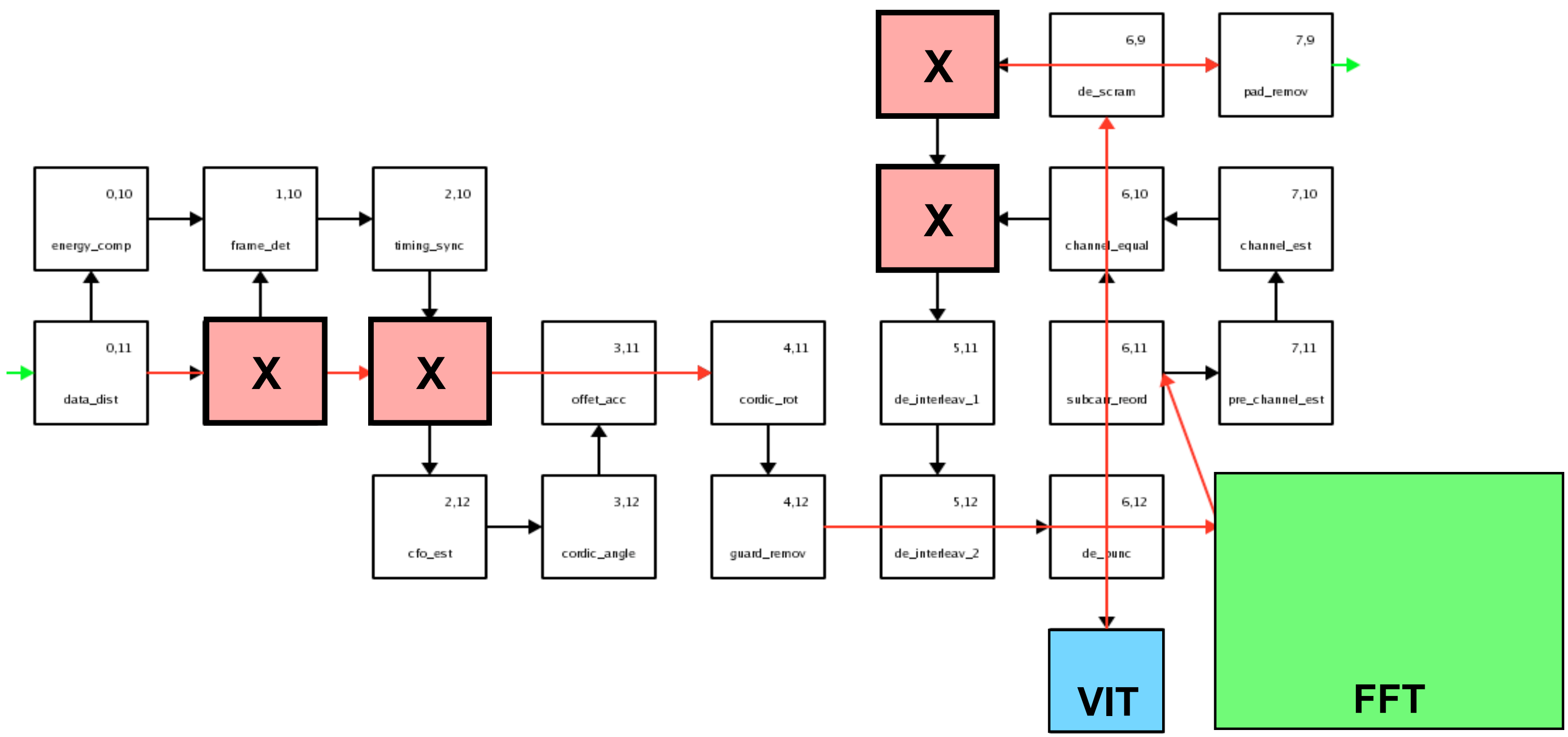
Complete 802.11a Baseband Receiver

- 22 processors plus Viterbi and FFT accelerators
- Includes: frame detection and synchronization, carrier-frequency offset estimation and compensation, channel equalization



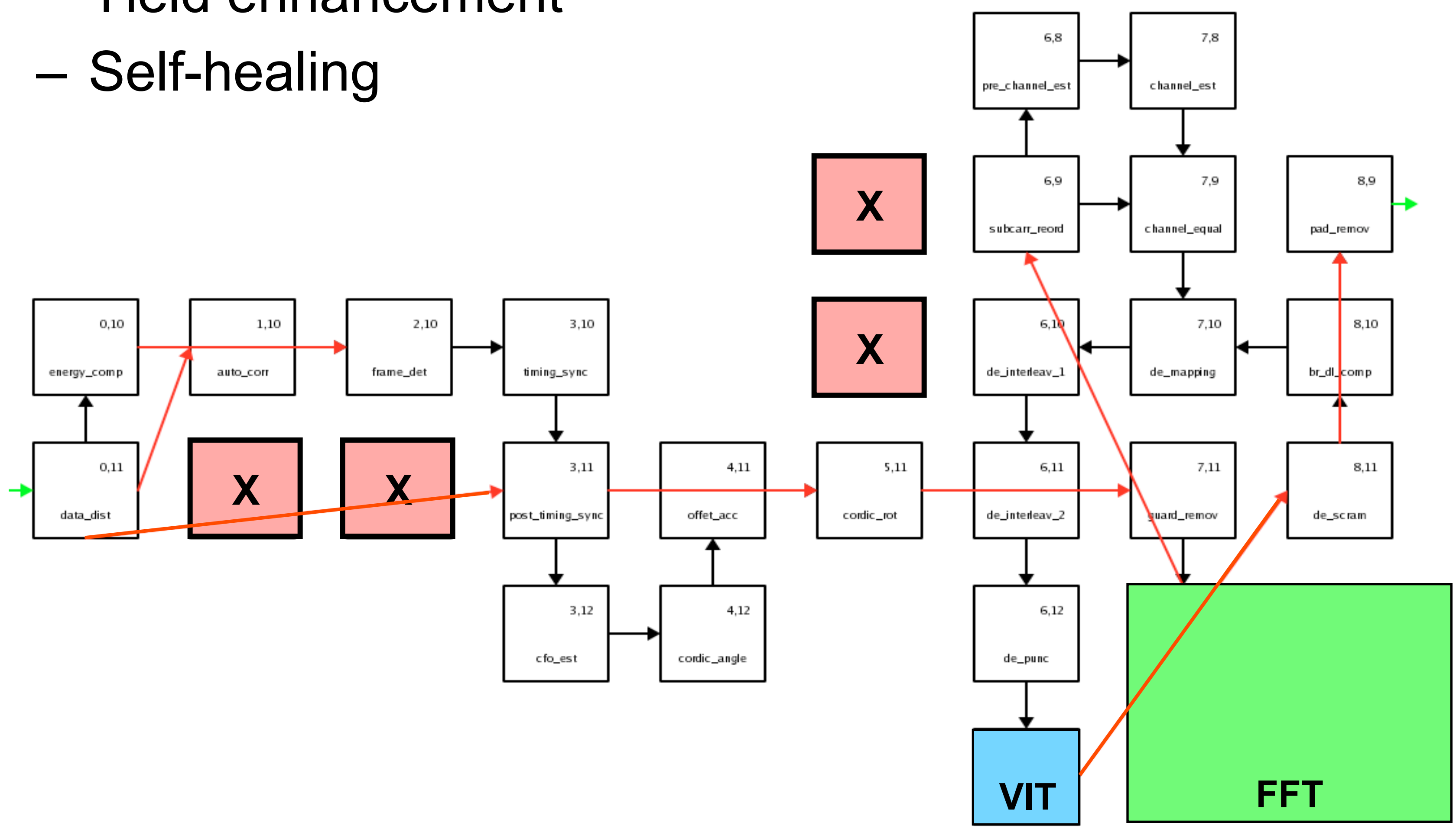
Complete 802.11a Baseband Receiver

- 54 Mbps throughput, 342 mW @ 590 MHz, 1.3 V
- 23x faster than TI C62x, 5x faster than strongARM, 2x faster than SODA (all scaled to 65 nm @ 1.3 V)



Complete 802.11a Baseband Receiver

- Re-mapped graph avoids bad processors
  - Yield enhancement
  - Self-healing



Summary

- All processors and shared memories contain fully independent clock oscillators
- 164 homogenous processors
  - 1.2 GHz, 59 mW, 100% active @ 1.3 V
  - 608 µW, 100% active @ 66 MHz, 0.675 V
- Three 16 KB shared memories
- Three dedicated-purpose processors
- Long-distance circuit-switched communication increases mapping efficiency with low overhead
- DVFS nets a 48% reduction in energy for JPEG application with an 8% performance loss

Acknowledgements

- ST Microelectronics
- NSF Grant 430090 and CAREER award 546907
- Intel
- SRC GRC Grant 1598 and CSR Grant 1659
- Intellasys
- UC Micro
- SEM
- J.-P. Schoellkopf, K. Torki, S. Dumont, Y.-P. Cheng, R. Krishnamurthy and M. Anders