A New Distributed DSP Architecture Based on the Intel IXS for Wireless Client and Infrastructure

Ernest Tsui
Communications and Interconnect Technology Lab
Corporate Technology Group

Kumar Ganapathy
Edge Access Division
Intel Communications Group

Contributors:
Hooman Honary, Rich Nicholls, Tony Chun, and Lee Snyder

HOT CHIPS 14
Session 7: Digital Signal Processors

20 August 2002

Outline

– Vision for Wireless Networks - ubiquitous
– Anticipated Issues – plethora of “standards”
– Future Wireless Requirements – “soft” with intelligence to increase capacity
– Architectural Objectives – flexible and low power
– How will we go about it? – distributed at the “right granularity”
– Distributed Architectural Summary – based on power, size, and wireless protocols we can derive a “good” (near optimal?) distributed architecture
– Comparison to other Wireless DSP research – flexible but within 2x of Berkeley Research Wireless Center’s Pleiades Arch.
– Summary – infrastructure architecture is near-optimum in granularity and power
– Next Steps – client architecture next

Vision for Wireless Networks

- Ubiquitous Internet Connections for all Mobile Client Devices
  - Handhelds, PDAs, Tablet PCs, and Laptops
  - Always-on
- New Paradigm for Wireless Basestations
  - Proliferation of basestations due to lack of spectrum
  - Agility across Multiple Bands
  - Multi-Network (WLAN, WWAN)

intel

Anticipated Future Issues

Wireless Protocol Plethora

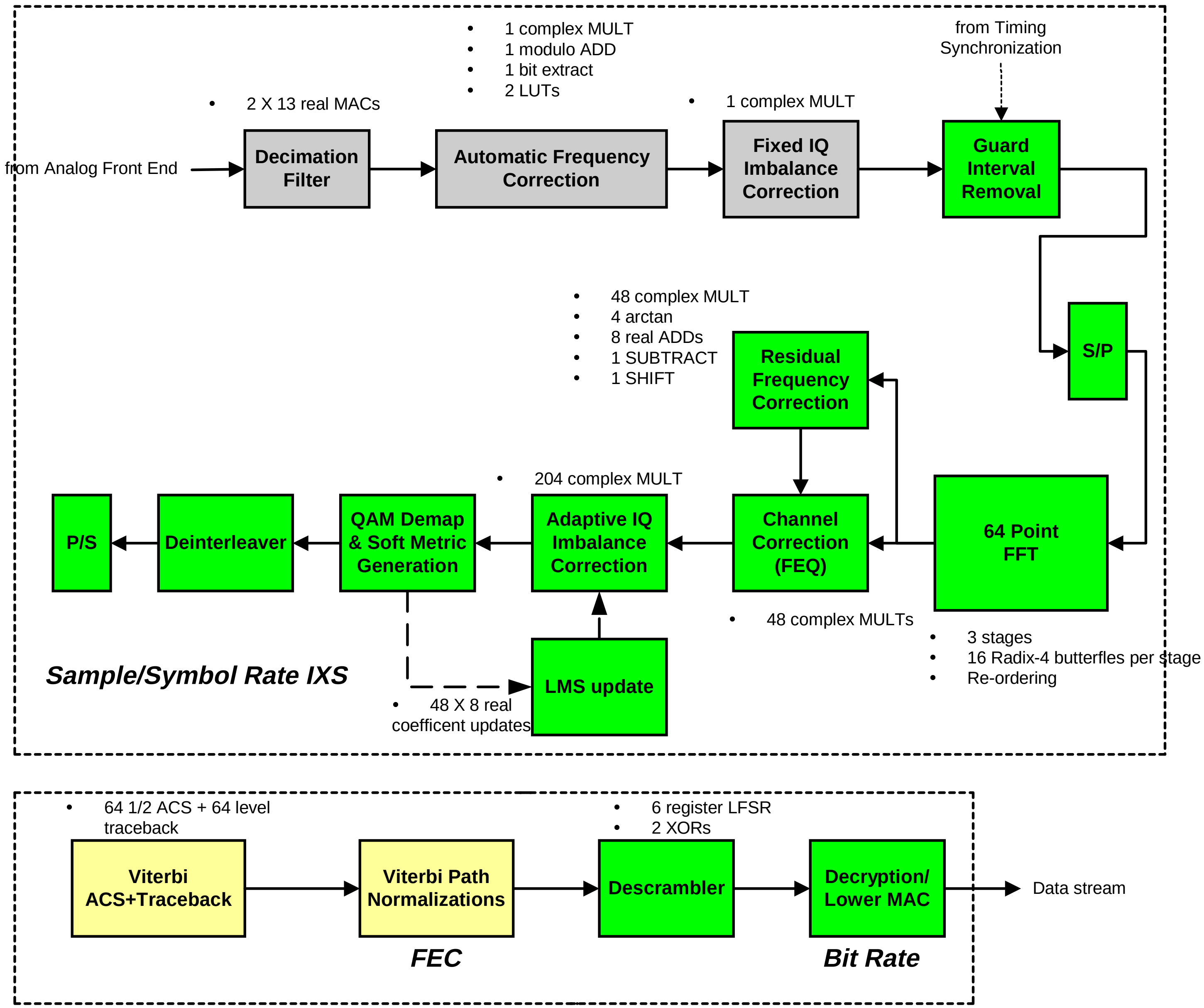
- PAN, WLAN, and WAN
  - PAN: Bluetooth (UWB, Wireless USB2)
  - WLAN (4 protocols): 802.11b/a (11g, Hiperlan II)
  - WAN (9 protocols):
    - 2G: IS-95, GSM
    - 2.5G: GPRS/EGPRS, cdma2000
    - 3G: WCDMA (FDD, TDD, SC), CDMA 1xE DV

intel

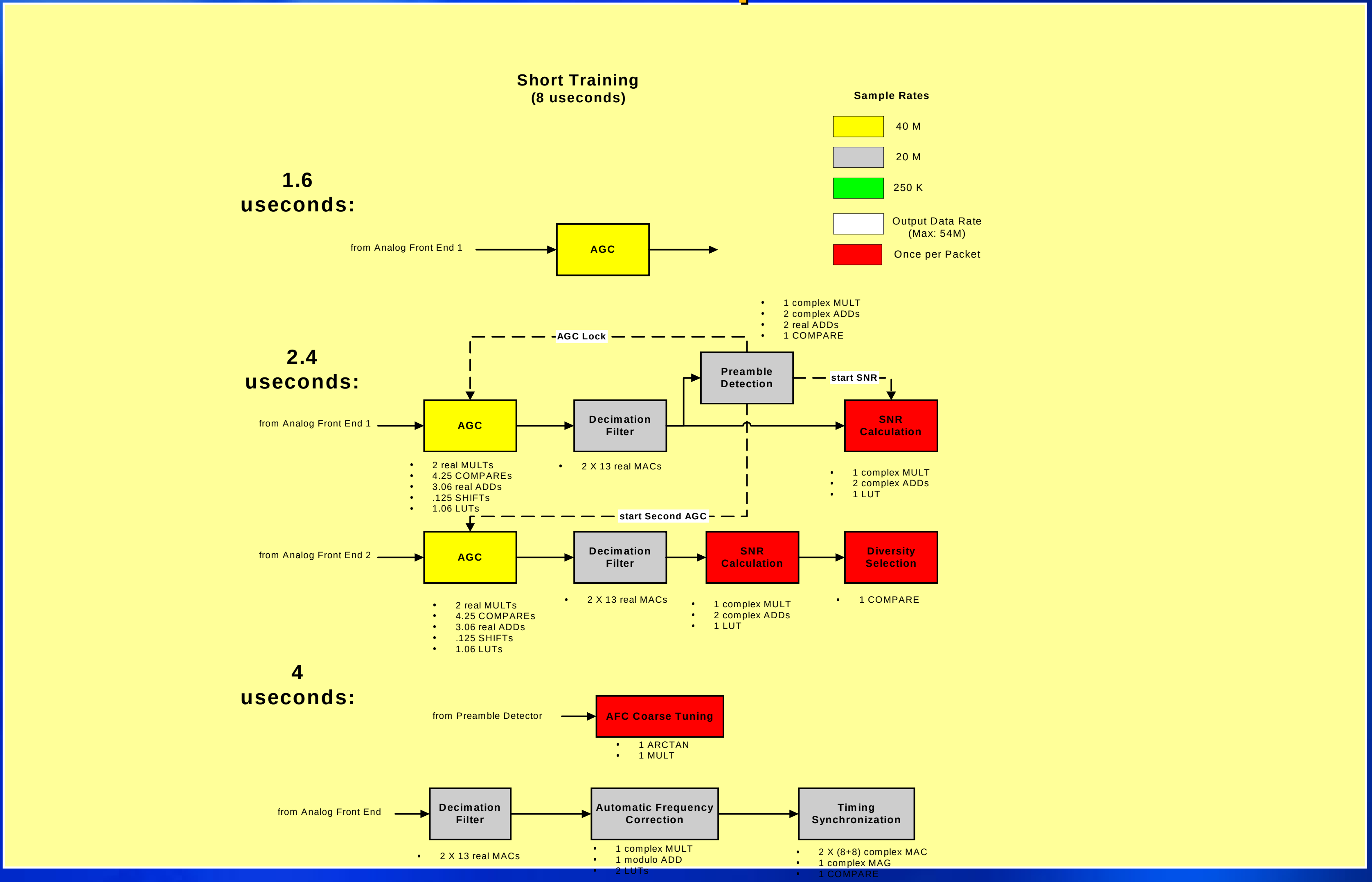
Wireless Requirements Summary

- Soft Radios at Basestations (deployed initially)
  - Low Power (<1 W) but highly flexible
  - Large no. of channels per core
  - Scaleable
- Reconfigurable Client Radios (deployed later)
  - Seamless Client Roaming
    - Two Concurrent Wireless Protocols
    - Selected 802.11a and WCDMA as the most intensive protocols
  - Variable User Environments require “adaptive” resource allocation
  - Adaptive to Broadband AFE distortions
  - Very Low Power (<< 1W)
    - Digital Baseband is < 10% of total PHY pwr
  - Reconfigurable to allow Si Re-use
  - Scaleable

802.11a Signal Processing Flow Example



802.11a Initial Acquisition Flow

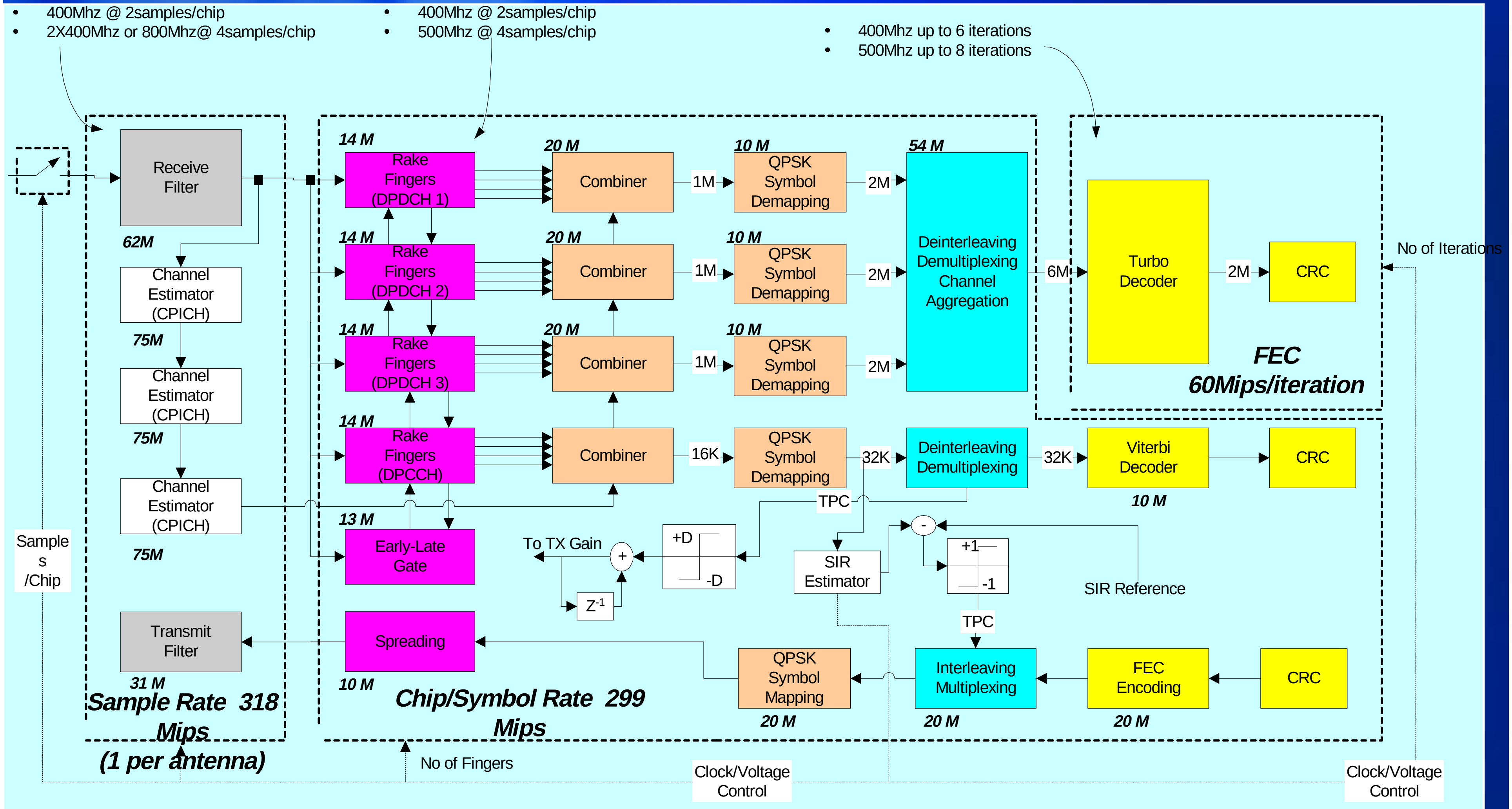


Hooman Honary

For new Packet Communications schemes – significant processing goes on during very short intervals of the preambles

intel

WCDMA Signal Processing Flow

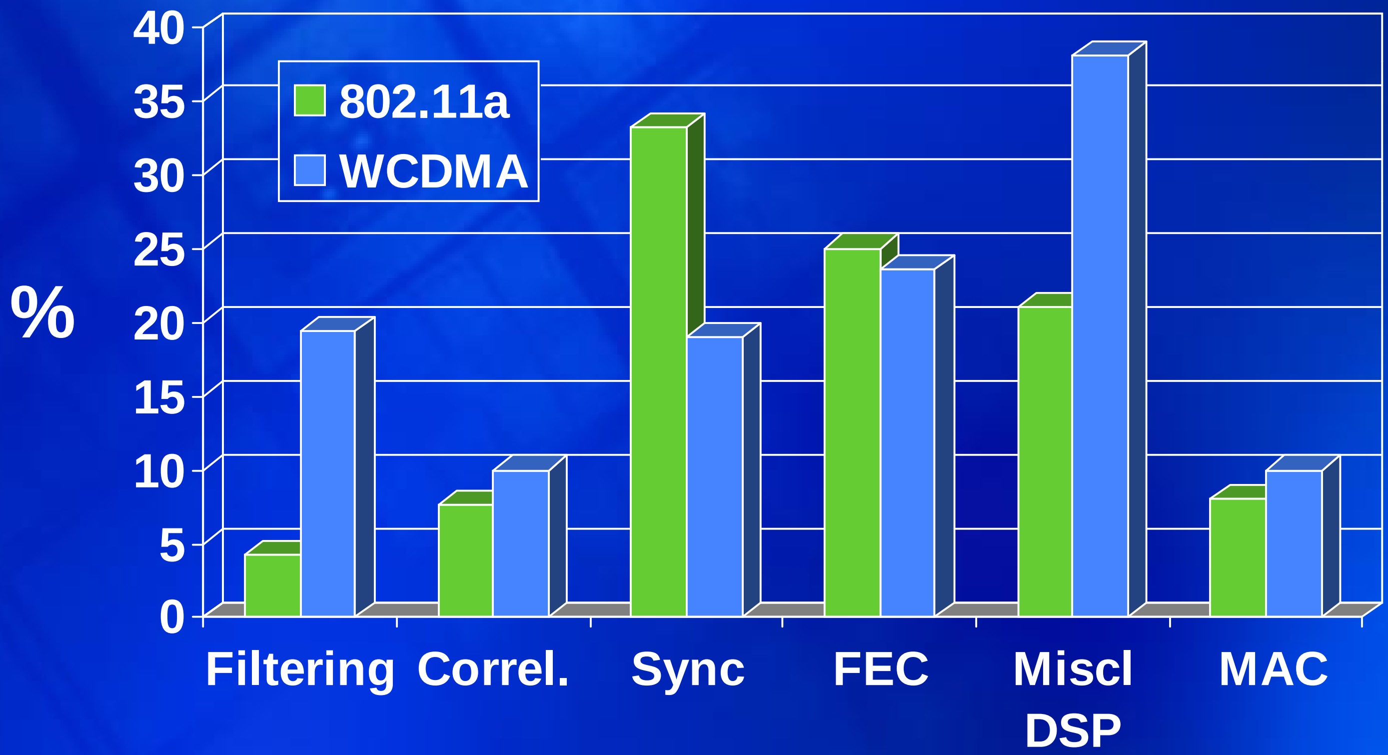


Tony Chun and Hooman Honary

The high data rates in 3G result in multi-code, -antenna, and -despreader (finger) processing requirements

intel

Computational Mix for Wireless Protocols

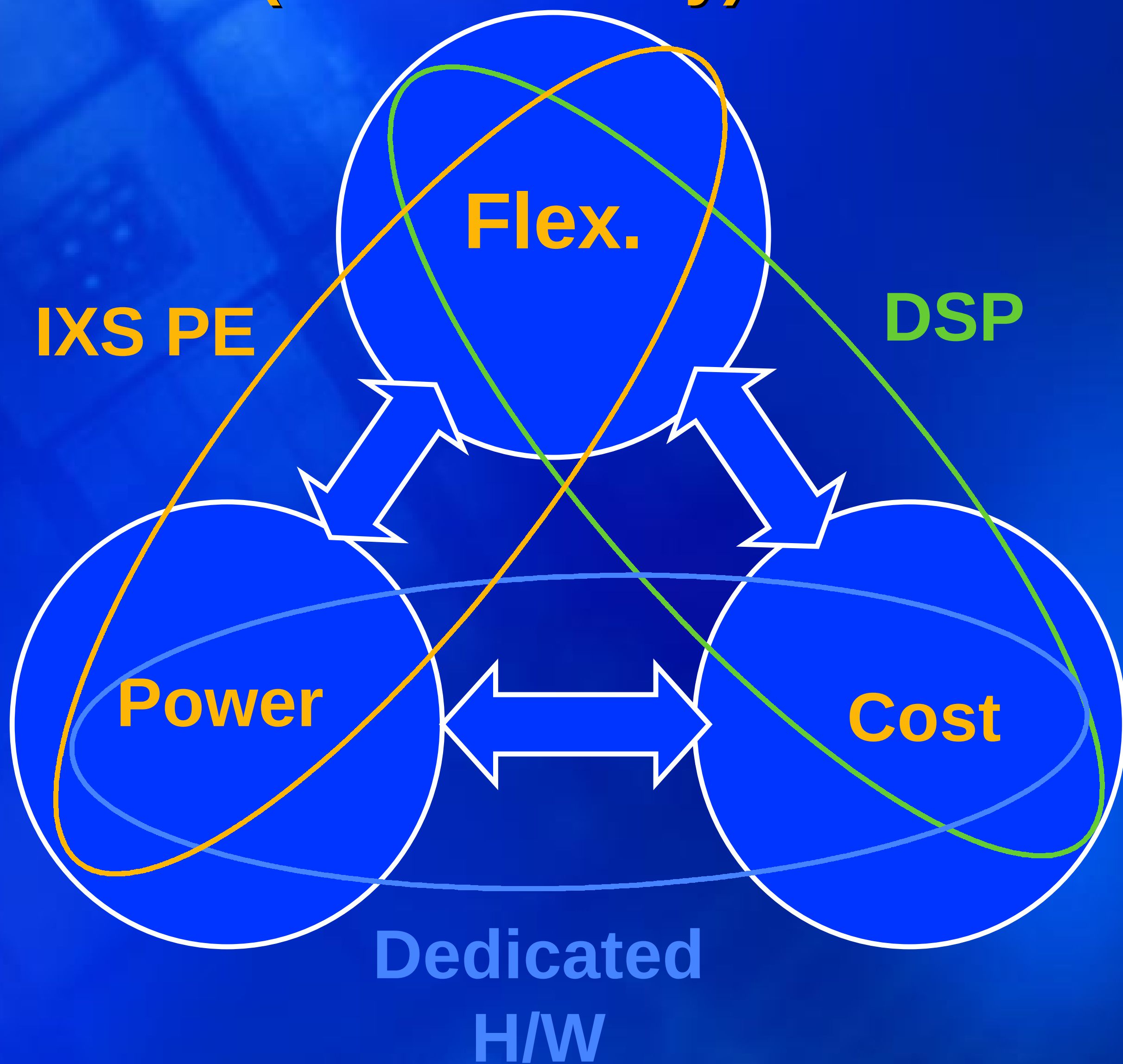


- Miscl. DSP ~ 10 separate signal processing threads

How will we go about it?

intel

Flexibility, Power, and Cost Trades (Pick two only)



Present Status of Soft Radios

- Prior Infrastructure Approaches
  - DSP + ASIC
    - Inflexible ASIC and Costly DSP
  - DSP + Closely Coupled Accelerators
    - Increased Power and Costly DSP
  - Reconfigurable
    - Hard to Program
    - Costly
    - High Power
    - Granularity problem has not been completely solved
- Need Evolved Architecture

Architectural Objectives

- Client:
  - 2-3x Power/Size of Dedicated Hardware for the most intensive protocol as a goal
  - Related to no. of protocols possibly in the client device
- Basestation:
  - 5-10x Power/Size of Dedicated Hardware for the most intensive protocol as a goal
  - Related to no. of protocols possibly in the infrastructure device

General Architectural Issues

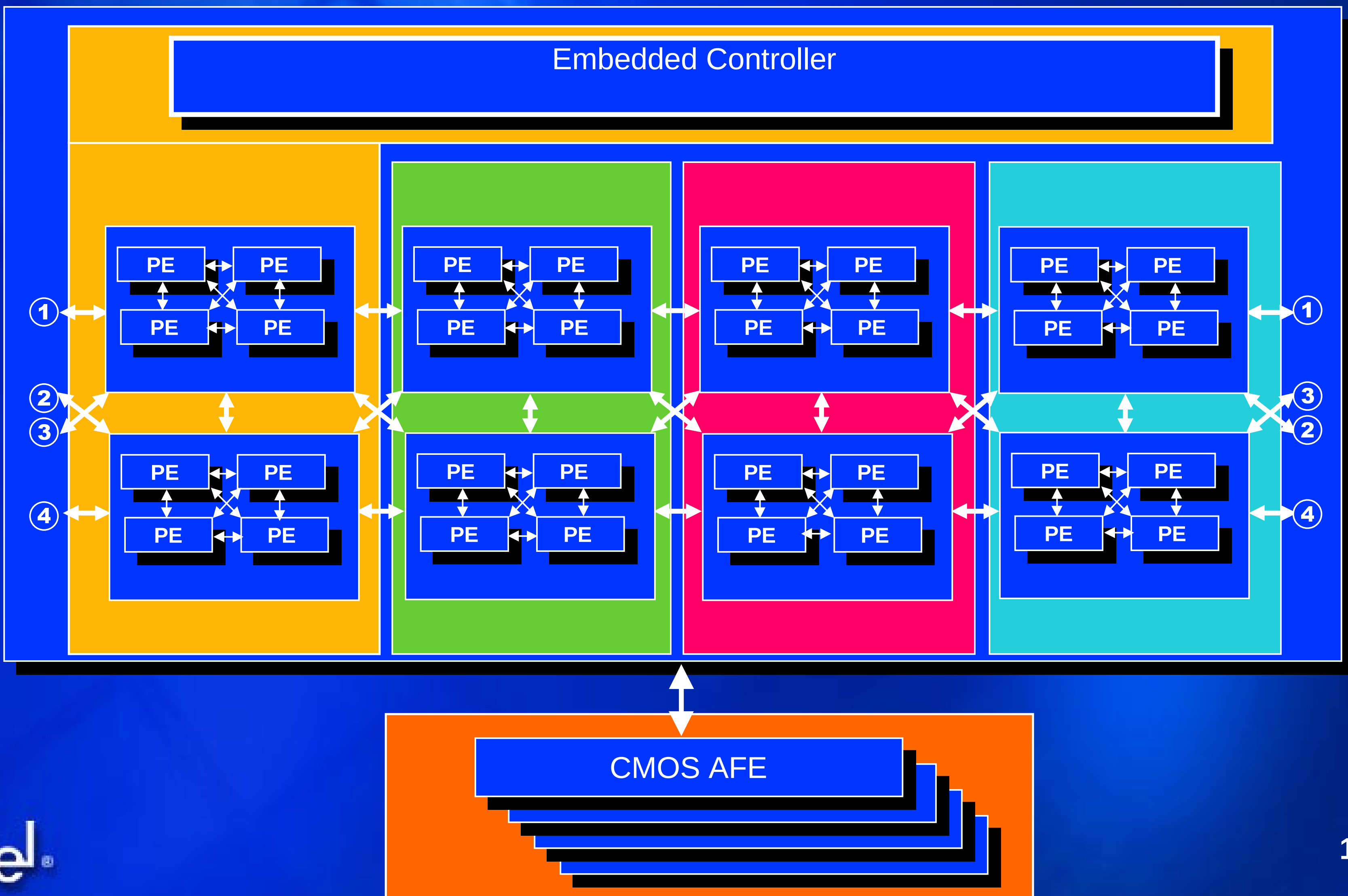
- Low power requires a **highly distributed** architecture
  - Low voltage helps quadratically lower power
  - Low clock frequency linearly lowers power
  - Large size penalties associated with distributed elements must be avoided
- What is the low power **interconnect** strategy?
- How do we simplify the distributed processor **programming** problem?

intel

Architecture Approach

- **Investigate Homogeneous Processing Elements (PE)**
  - Easy to Scale and to Program for Basestations
  - Heterogeneous better for Client
- **Interconnect with Nearest Neighbor Mesh**
  - Eliminates High Speed (and power) buses [J. Rabaey, Silicon Architectures for Wireless, Hotchips 2001 Tutorial]
  - PHY connections are 95% nearest neighbor
- **Number of Distributed Processing Elements**
  - Driven by:
    - Computational Load
    - Size and Power Constraints
    - Feature parameters, e.g., Average Load Capacitance, Vdd, etc.
- **Type of Element**
  - General Purpose DSP combined with:
  - Acceleration of "Standard Operations" with the right granularity
- **S/W programming via High Level Language**
  - Explicitly indicates parallelism and connections

System Architecture



intel

Does a Good (near optimal) PE Solution Exist?

intel.

Macro-architectural Constraints

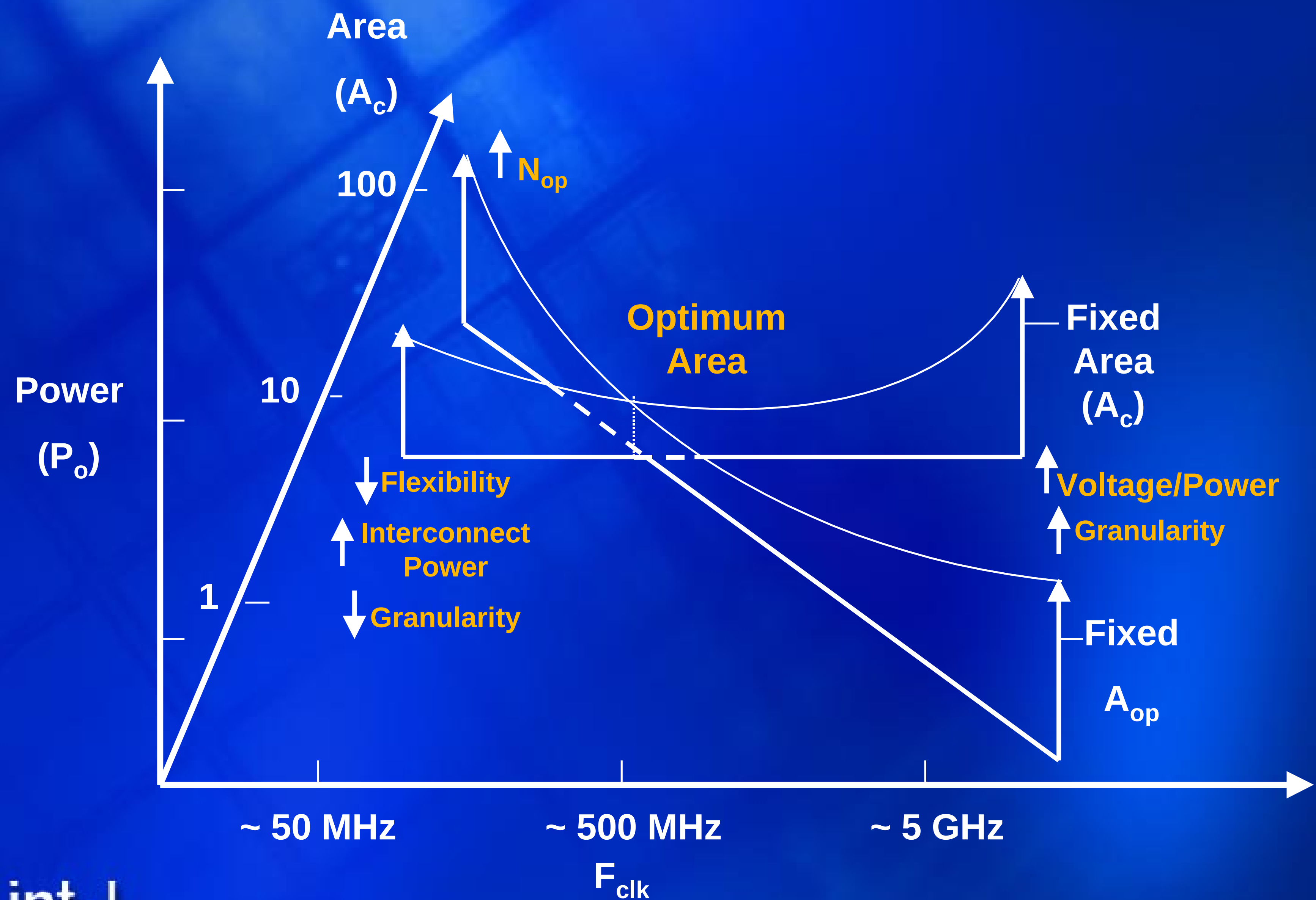
- **First, must meet Power, Size, and Computational Load constraints**
  - Computational Load = R_c (ops/sec.)
    - N_{op} = No. of parallel significant operations (multiplies, etc.) in one cycle [R. Brodersen, ISSCC'02]
    - F_{clk} = Clock frequency
    - $N_{op} \times F_{clk} > R_c$
  - Power Constraint = P_o (mW)
    - Power (dynamic, leakage (P_{leak}), short circuit (Psc)) < Po
  - Size Constraint = A_c (mm2)
    - $N_{op} \times A_{op} < A_c$
    - A_{op} = Average area of a significant computational unit (e.g., multiplier-memory-address-decoder, etc.) (mm2)
    - **Aop ~ Granularity Factor**
  - Constraints on F_{clk}
    - $R_c / N_{op} < F_{clk}$
    - $R_c \times A_{op}/A_c < F_{clk}$

Clock Rate Bounds

- F_{clk} **is upper bounded by power constraints**
  - $a \times C_{sw} \times Vdd^2 \times F_{clk} + P_{leak} < P_o/(b \times A_c)$
    - where P_{leak} is the average pwr leakage density in mW/mm2
    - C_{sw} is the average switching (load) capacitance per mm2
    - 'a' is the activity factor
    - 'b' is the average active area (incl. Datapath, cache, cache memory bus, etc. and excl. L2 memory, etc.)
    - 'b' varies from ~ 10% for microprocessors to ~ 80% for dedicated hardware and also is a function of clock gating strategies
- F_{clk} **is lower bounded by computational and area constraints**
  - $\mathbf{R_c \times A_{op}/A_c < F_{clk} < (P_o/(b \times A_c) - P_{leak})/ (a \times C_{sw} \times Vdd^2)}$
  - Key Issues:
    - Find the F_{clk} that meets upper and lower bounds
    - Derive the A_{op} and N_{op}

intel

General Power, Area, F_{clk} Trends



Reconfigurable Power Trend Summary

- **There is an optimum F_{clk} for a fixed A_{op}**
  - (Recall that A_{op} is the fundamental processing size)
  - The optimum meets Size and Computational requirements and minimizes power for the above
  - Higher F_{clk} increases power and lower F_{clk} increases area and interconnect power
- **Is there a similar optimum as A_{op} is Varied?**
  - As A_{op} decreases – interconnect Power increases exponentially
    - Simpler elements must be connected in a more complex manner to retain flexibility
  - As A_{op} increases - the voltage requirement (and Power) increases
    - More complex element requires time-multiplexing
- **Thus, is there a globally “good” design?**
  - Conjecture:
    - Determine the Minimum Aop (for the flexibility desired) and find the optimum F_{clk}

Example of "Good" Architecture Parameters in the optimum area

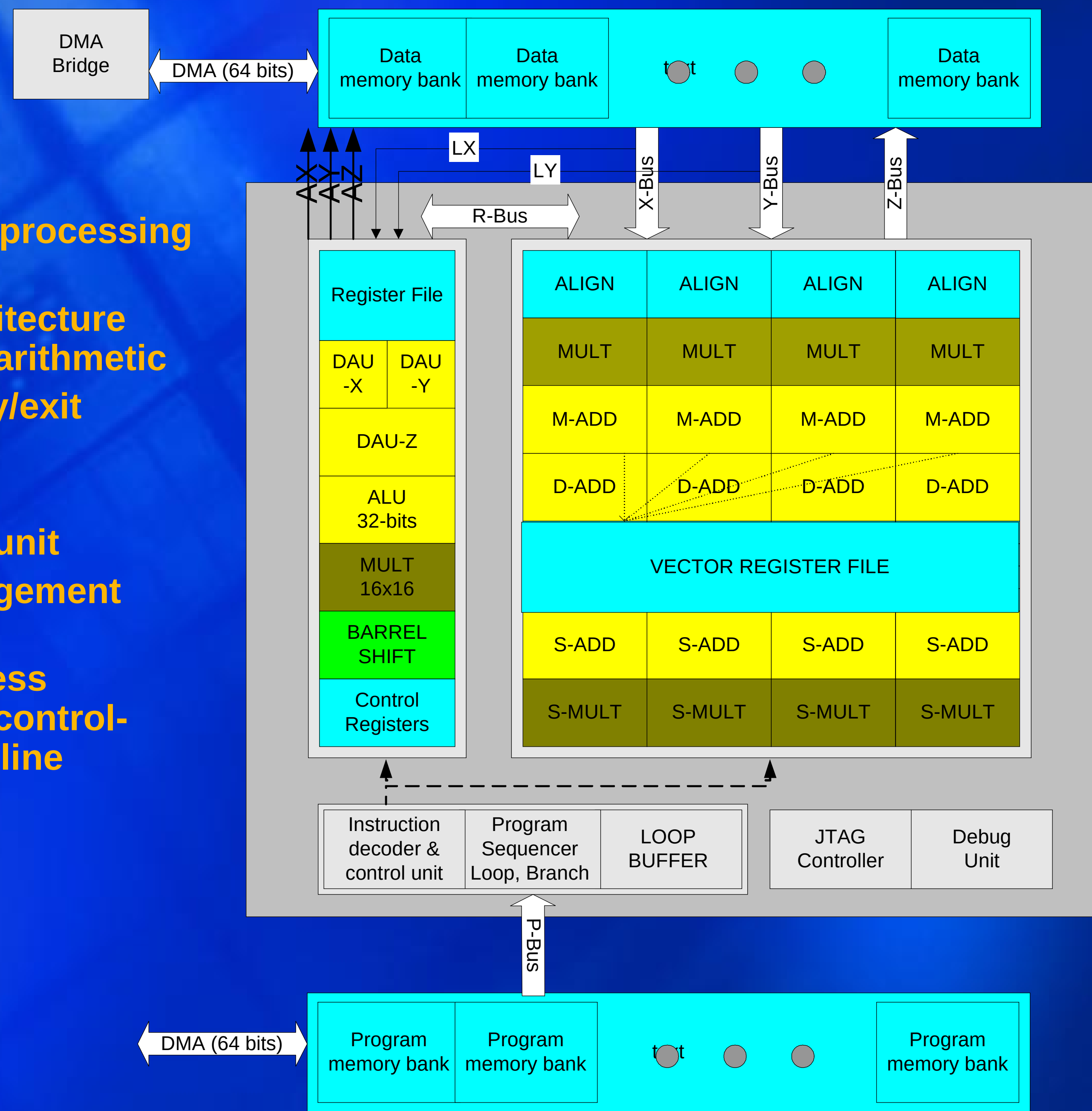
- N_{op} (No. of parallel Significant operations), for 90 nm:
  - N_{op} ~ 50
  - A_{op} ~ 0.6 mm^2
    - Is this an optimum Granularity A_{op}??
  - F_{clk} ~ 400 MHz
  - P_o ~ 750 mW
  - R_c ~ 20 GOPs

intel

Key Computing Element IXS Core

IXS core

- Efficient Vector processing architecture
- Octal-MAC architecture with 8/16/32-bit arithmetic
- Quick loop entry/exit mechanisms
- Loop buffer
- Data alignment unit
- Resource management engine
- Integrated address generation and control-processing pipeline

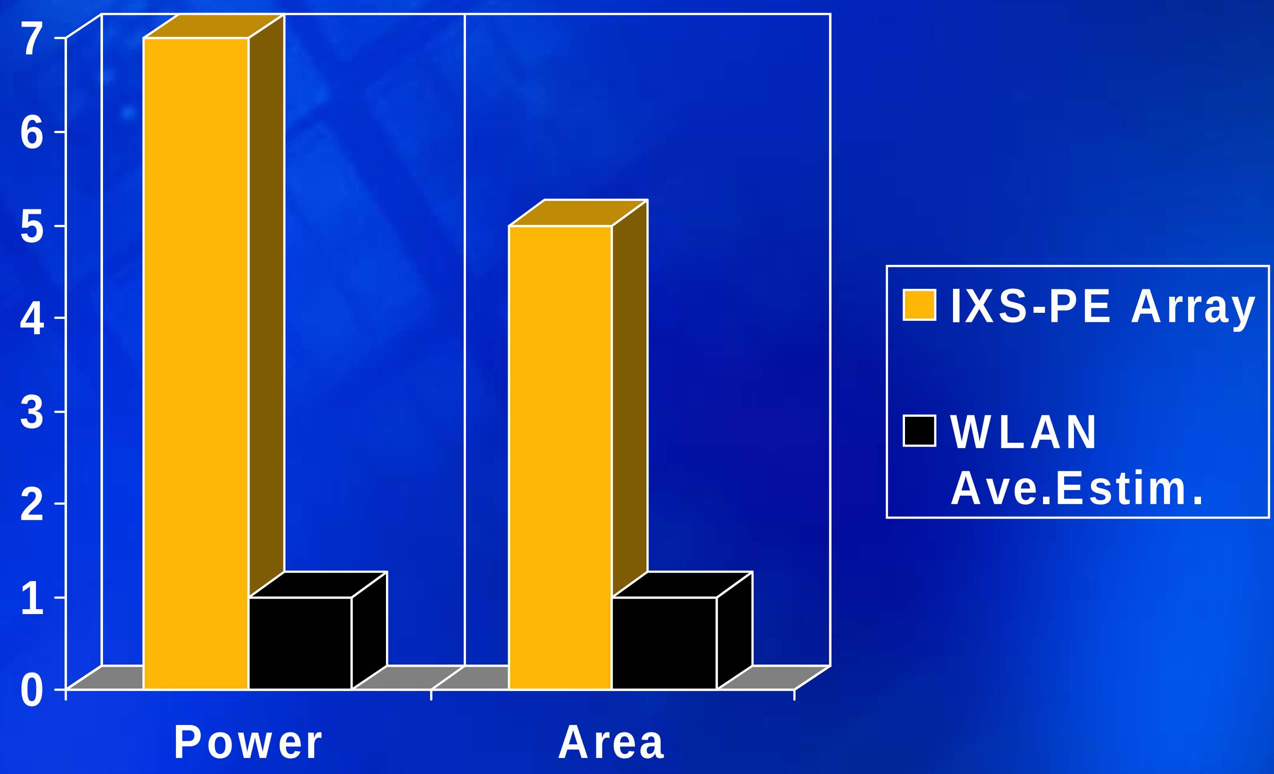


Architecture Summary

- IXS Processor Octal MAC units
  - RISC-tightly coupled
  - Acceleration H/W
    - Viterbi/Turbo
    - Correlation, De-spreading, etc.
    - Filter
  - Parameters within the N_{op} Range (50)
    - 5 PEs x 9 MACs = 45 MACs
    - 32 – 8 bit adders per PE
- Mesh-Connected to Surrounding Processors (5 PEs total)
- Do we have the optimal A_{op}?
  - Lower A_{op} will start to increase interconnect Power

How does the IXS PE Compare against Dedicated Hardware?

Power and Area Efficiency of IXS PE vs Dedicated H/W for WLAN Benchmark
Still 5-7x Dedicated H/W

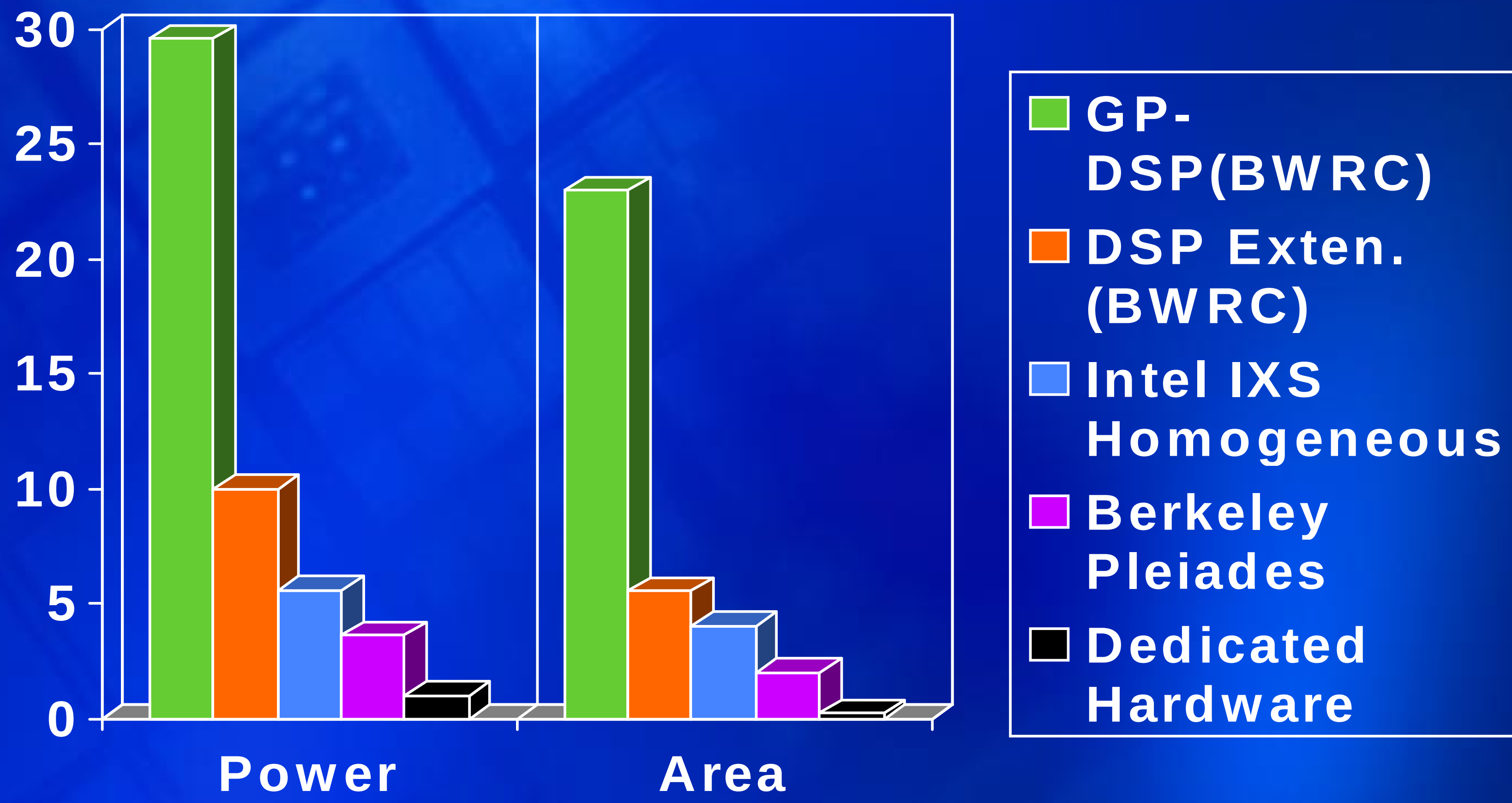


Baseband PHY and lower MAC estimates

(all scaled to 90 nm)

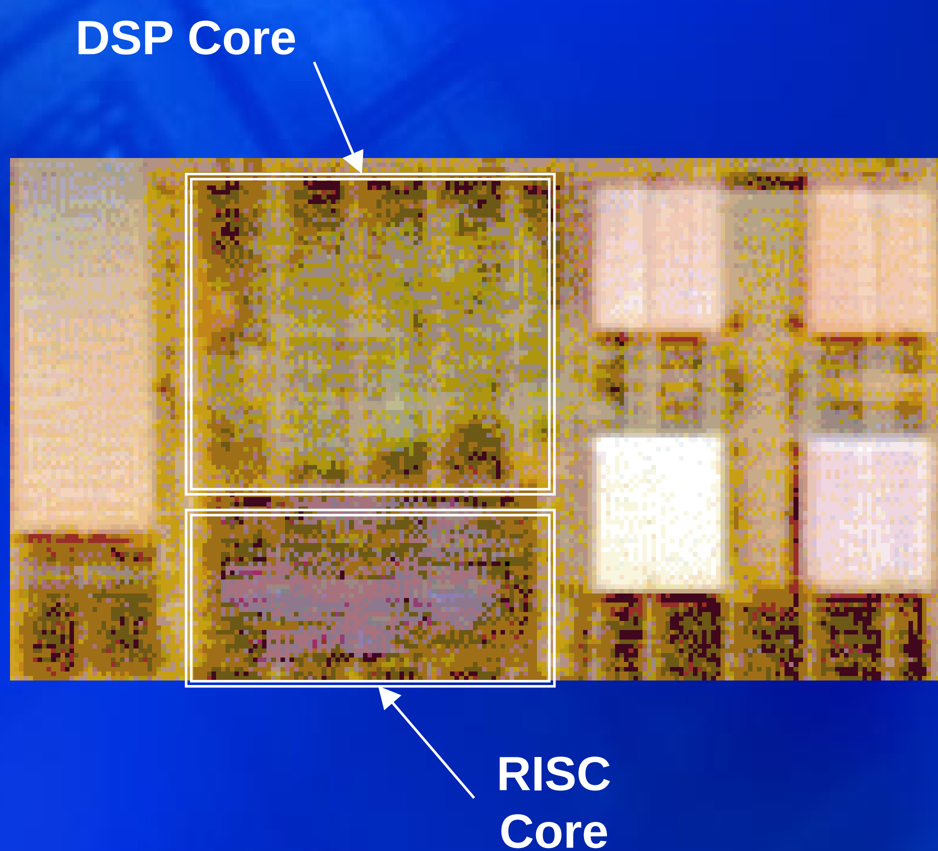
How do we compare against other Reconfigurable Approaches?

How Does our Architecture Compare?
Multi-User Detector Benchmark



BWRC and Lee Snyder

Die Photo



Summary

- Homogeneous Mesh-Connected Array of IXS Processing Elements for Infrastructure
  - Low power/size (5-7x dedicated h/w)
  - Flexibility where it's needed
  - Scaleability
  - For given size/power and feature size constraints a "good" solution can be found
  - Key Processing element
    - Minimum Memory
    - "Maximum-Datapath" Units
- Next Steps:
  - "What is the optimal A_{op} Size?"
  - "What is the right Arch. for the Client?"

intel