SRC/NSF Workshop on Virtual Immersion

versus

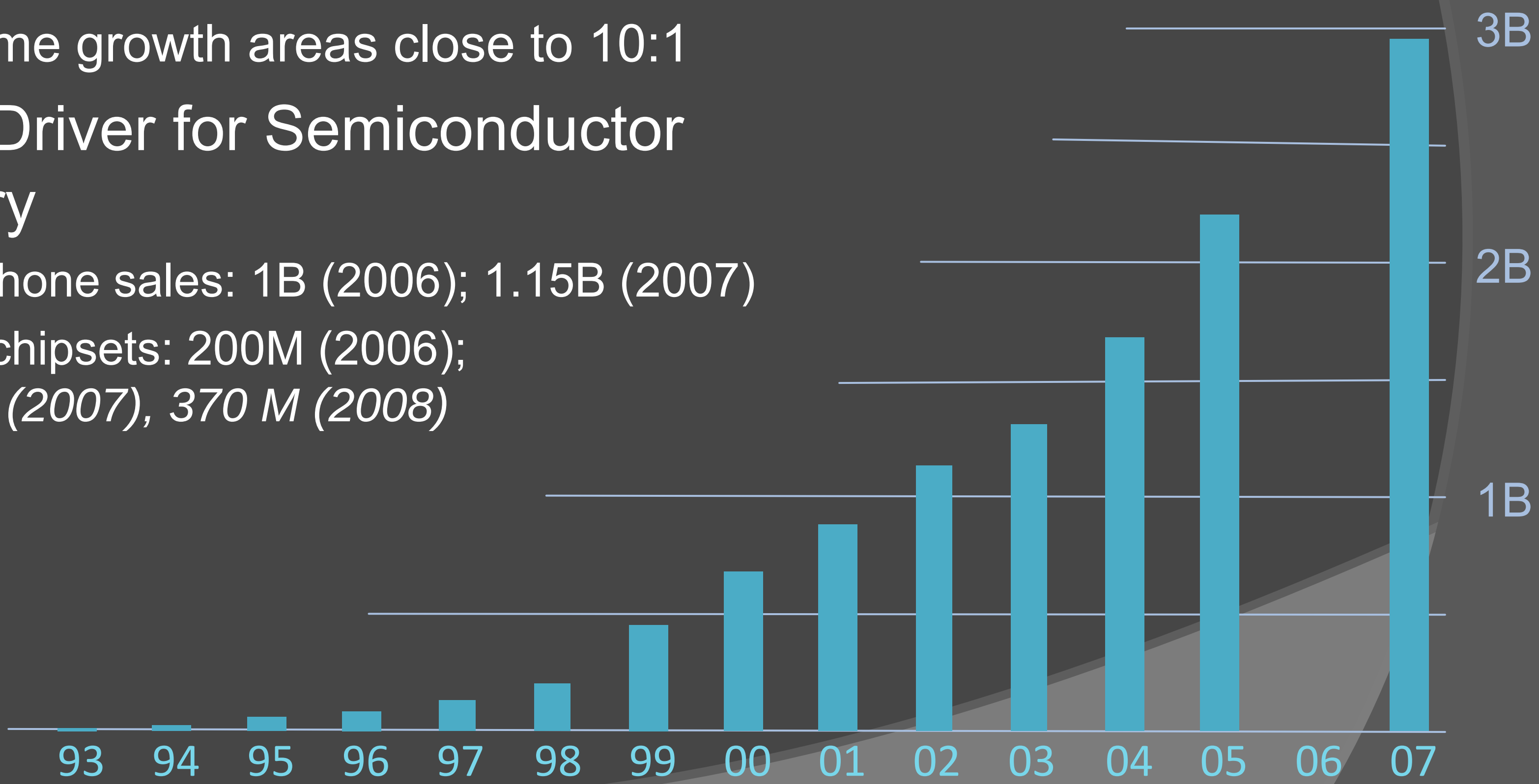
?

Jan M. Rabaey, Donald O. Pederson Distinguished Professor

Director Gigascale Systems Research Center (GSRC)
Scientific Co-Director Berkeley Wireless Research Center (BWRC)
University of California at Berkeley

The Era of True Mobility is Here

- Wireless subscribers expected to top 3 Billion in 2008! (40% penetration)
- Mobile devices outnumber PCs 5:1
  - In some growth areas close to 10:1
- Major Driver for Semiconductor Industry
  - Cell phone sales: 1B (2006); 1.15B (2007)
  - WIFI chipsets: 200M (2006); *280M (2007), 370 M (2008)*



Worldwide wireless subscribers

Exponentials Bound to Continue

EE Times: Latest News

Wireless is everywhere; ignore it at your peril

Bolaji Ojo

Page 1 of 2

EE Times

(01/07/2008 9:00 AM EST)

PRINT THIS STORY
SEND AS EMAIL
REPRINTS

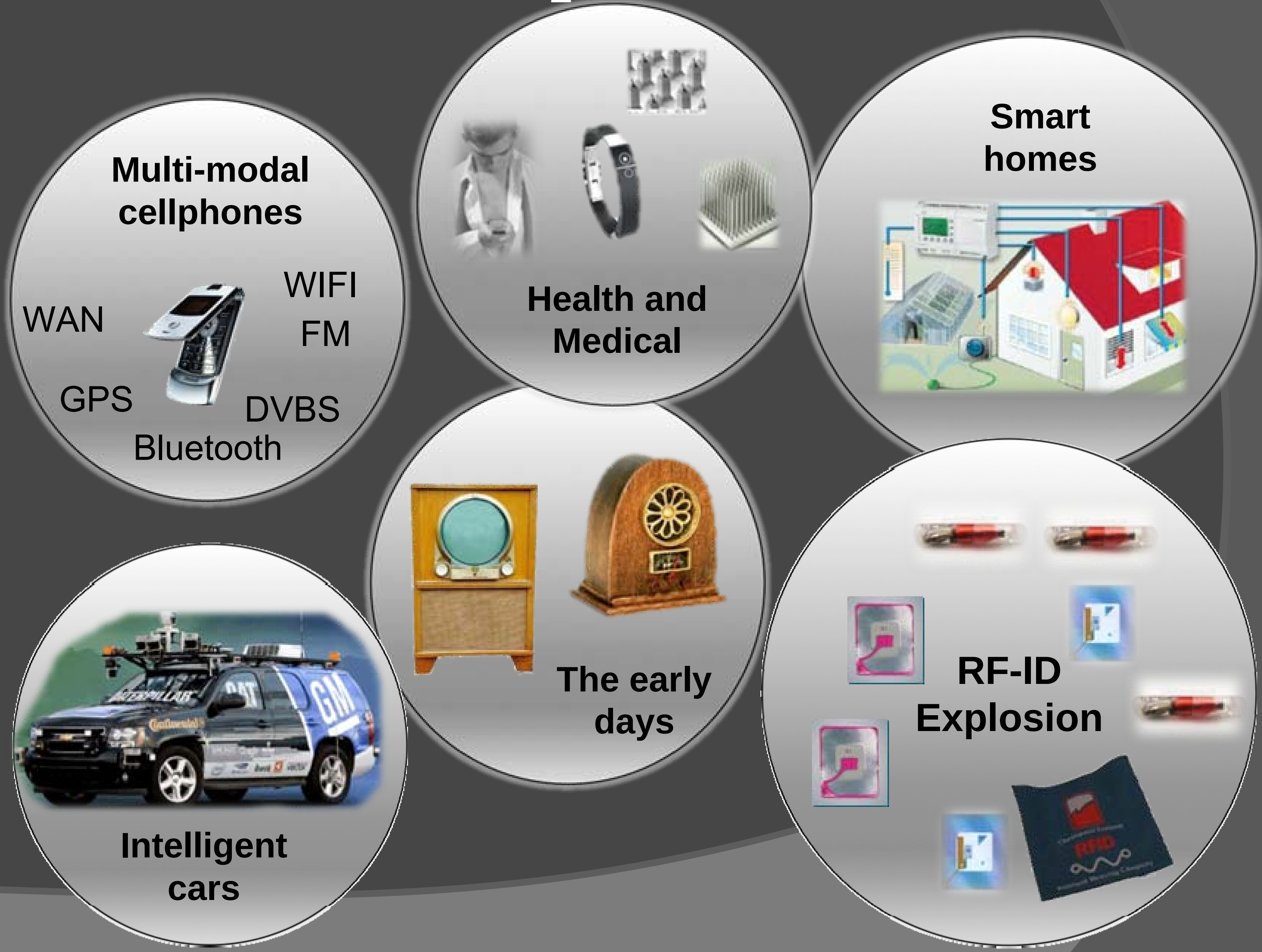
The search is over for the next killer app. It is wireless, it is all around you, and it will leave no sector of the global economy untouched.

EE Times, January 07, 2008

- 5 Billion people to be connected by 2015 (Source: NSN)
- The emergence of Web2.0
  - The "always connected" community network
- 7 trillion wireless devices serving 7 billion people in 2017 (Source: WWRF)
  - 1000 wireless devices per person?

[Courtesy: Niko Kiukkonen, Nokia]

A 1000 Radios per Person?



Information-Technology in Turmoil

From Batch

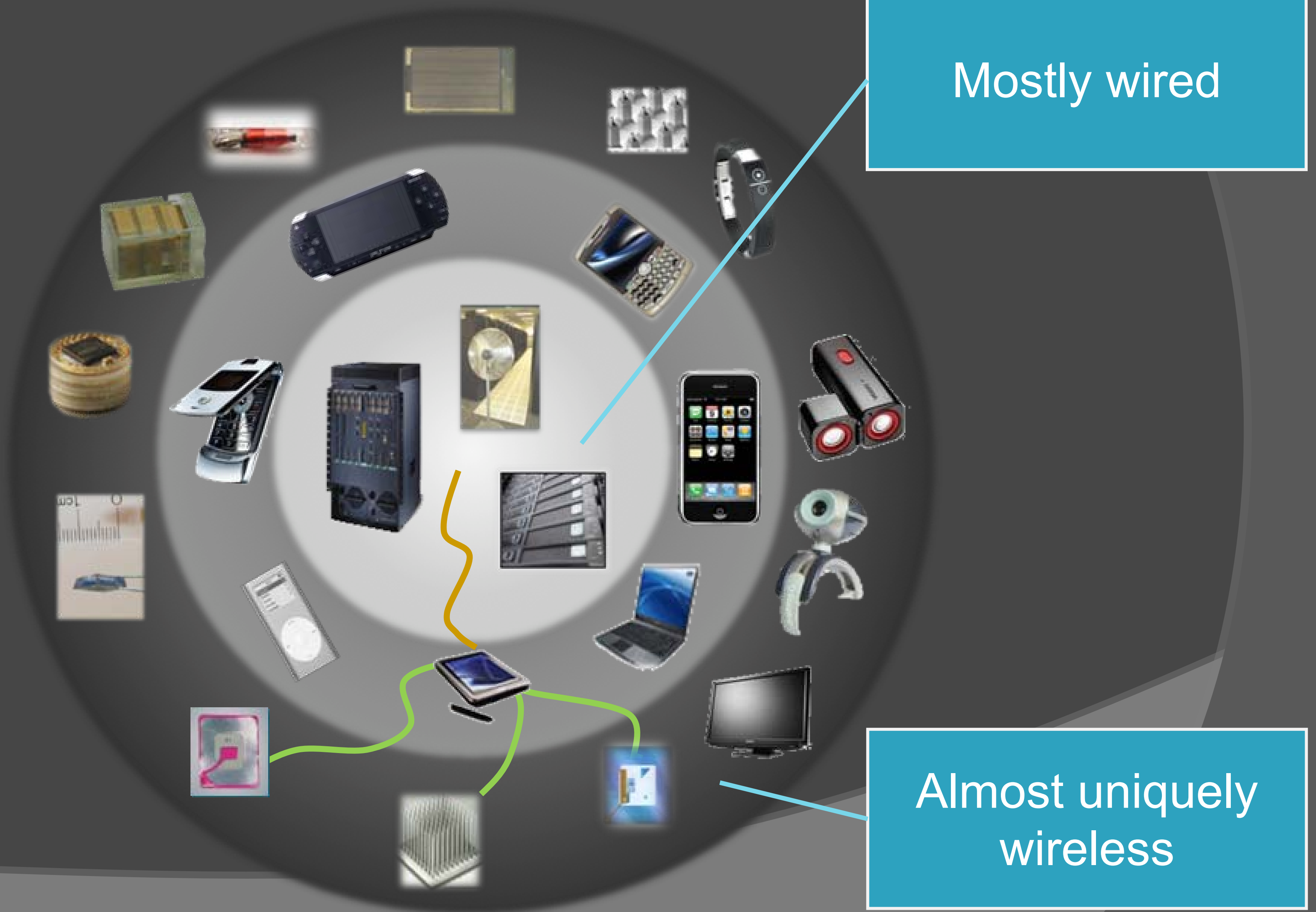
Over Interactive

To Immersion

The "Immersive" IT Scene

The Technology Gradient: Computation

The Technology Gradient: Communication



The Birth of "Societal IT Systems (SiS)"

"A complex collection of sensors, controllers, compute nodes, and actuators that work together to improve our daily lives"

- The Emerging Service Models
  - Intelligent data access and extraction
  - Immersion-based work and play
  - Environmental control, energy management and safety in "high-performance" homes
  - Automotive and avionic safety and control
  - Management of metropolitan traffic flows
  - Distributed health monitoring
  - Power distribution with decentralized energy generation

SiS Wireless– The Challenges

- Reliable universal coverage at all times!?
  - 7 trillion radios will quickly run out of spectrum …
  - Wireless is notoriously unreliable
    - Fading, interference, blocking
  - Mobility requires dynamic reconfiguration
  - Heterogeneity causes incompatibilities
    - Large number of standards to co-exist
    - Devices vary in form-factor, size and energy source

TOP STORY

Updated Mon, 14 Jan 2008 01:03:01 PM EST

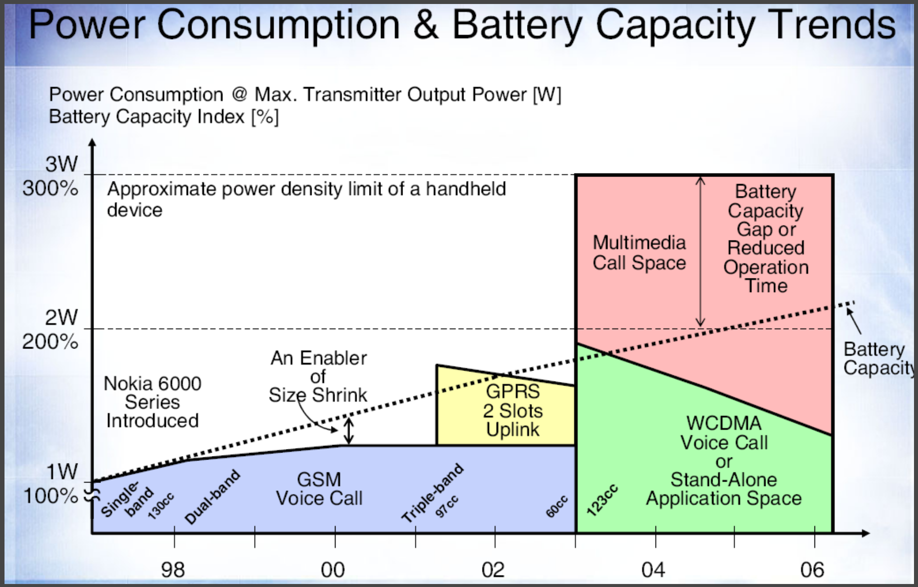
CE's wireless Babel: Connectivity strategies are all over the map

Now that consumer electronics companies are delivering a full suite of product to the digital living room, they are working out how to connect them.

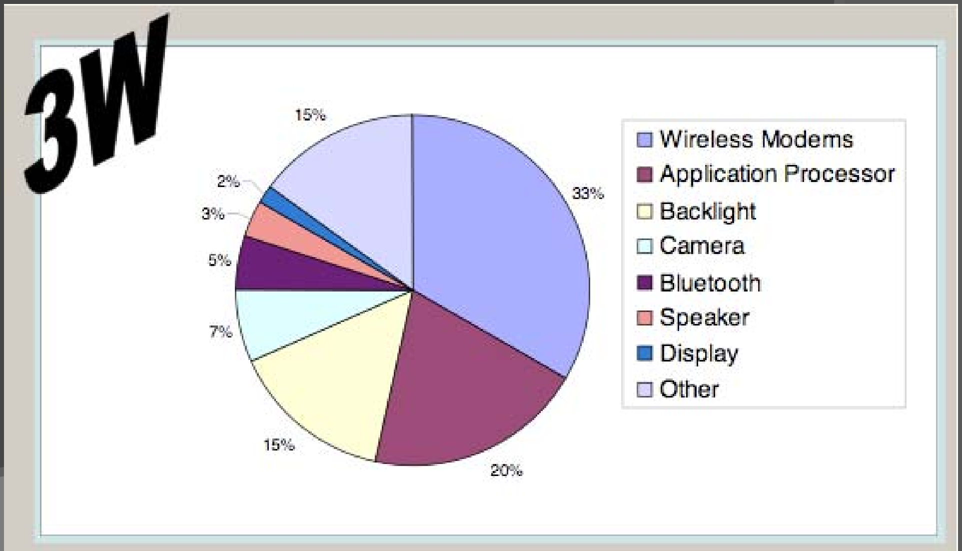
EE Times, Jan. 14 2008

The Mobile Access Device

- The cell phone and its descendants as the "personal" communication and computation device of choice
- Bringing together many different functionalities
- Becoming a "base station" in itself



Power dissipation primary concern … but increasing performance essential



Y. Nuevo, ISSCC 04

Mobile Computing Requirements

Now:	Processing capability	Execution memory	Power	Processing elements	Inter-connect	MIPS/ mW
HPC	360 TFLOPS	32 TB	1.5 MW	65,536 nodes x 2 cores	2.1 GB/s per node	0.24
Media uP	256 GFLOPS	1 GB	60 W	9 cores	200 GB/s	4.2
Laptop uP	4 GFLOPS	1 GB	25 W	1 core	4 GB/s	0.16
Mobile uP	512 MIPS	64 MB	0.5 W	1 core	1 GB/s	1.0

2010-2012:						MIPS/ mW
Mobile uP	4000 MIPS	512 MB	1 W	4 cores	8 GB/s	4
Beyond:						
Mobile uP	256 GIPS	1-2 GB	1 W	? cores	100GB/s	256

Needs factors of 50-60 improvement in energy efficiency!

[Ref: K. Kuusilina, Nokia, DATE 2008]

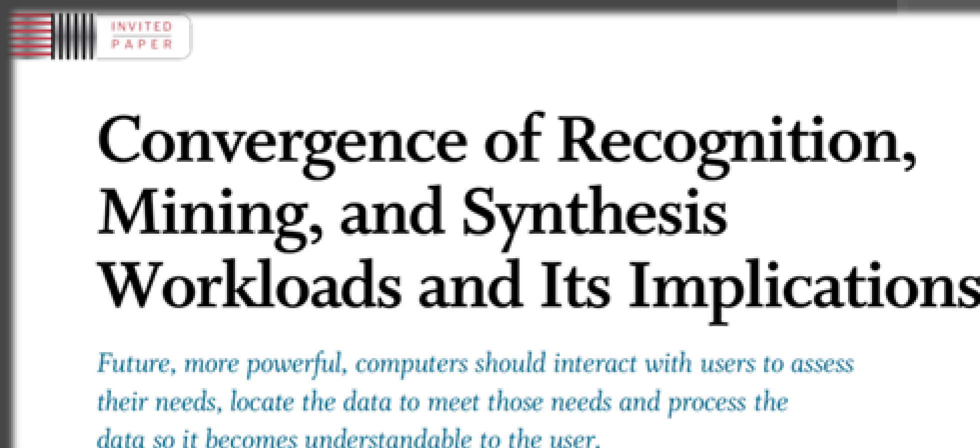
Increasing Mobile Energy Efficiency

- Multi-core platforms only partial answer
  - Energy efficiency quickly saturates
- Some improvement possible with other architectural innovations
  - Heterogeneous architectures with accelerators
  - Better power and activity management
- More efficient communication links enable off-loading of functionality
  - Increasing role of “More than Moore”
  - Needs a system-level perspective!

Increasing Role of User Interfaces

- Innovative interaction paradigms between user and machine becoming increasingly important
  - Recognition, Mining, Synthesis (RMS)
  - A dominant factor of computational requirements in future mobiles
- Opens the door for innovative energy-efficient algorithms and architectures
  - Allowing for error-tolerance
  - “swarm” adds more complex “senses”

YK Chen et al (Intel), IEEE Proc., May 08

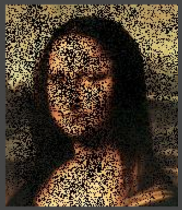


INVITED PAPER

Convergence of Recognition, Mining, and Synthesis Workloads and Its Implications

Future, more powerful, computers should interact with users to assess their needs, locate the data to meet those needs and process the data so it becomes understandable to the user.

Often, It is Ok to Make Errors



Recognition, Mining, Synthesis (RMS)

- Low-order bit data errors OK – known for decades
  - Google MapReduce, Bayesian nets, etc.

☹ RMS on unreliable hardware DOESN'T WORK

☹ Frequent crashes, high-order bit errors

Key observations:

Compute intensive threads → Errors OK

Main (control) thread → No errors allowed

Error Resilient System Architecture (ERSA):

Asymmetric reliability

Hardware prototype:

ERSA → Resilient to > 10^{16} FITs

Linear speedup

ERSA Architecture

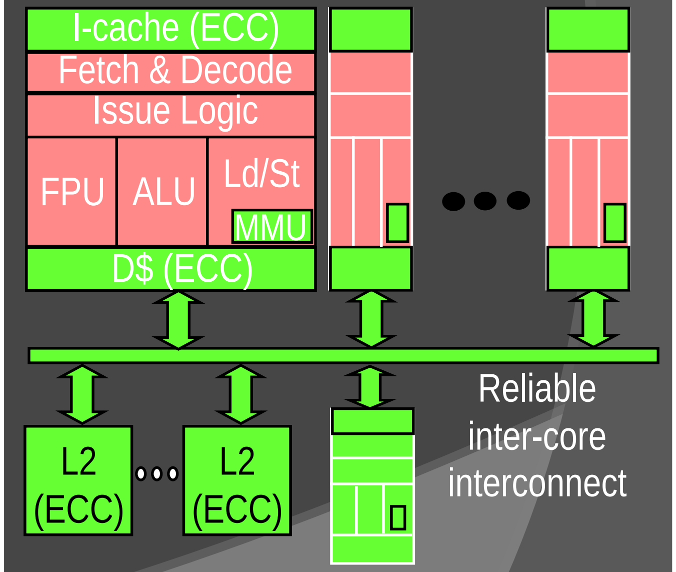
Relaxed Reliability Cores (RRCs)

(> 3,000 errors per sec.)

Sequestered from OS

Compute intensive threads +

"lightweight" memory bounds checks



Strictly Reliable Core (SRC): OS visible

Main (control) thread + timeout checks

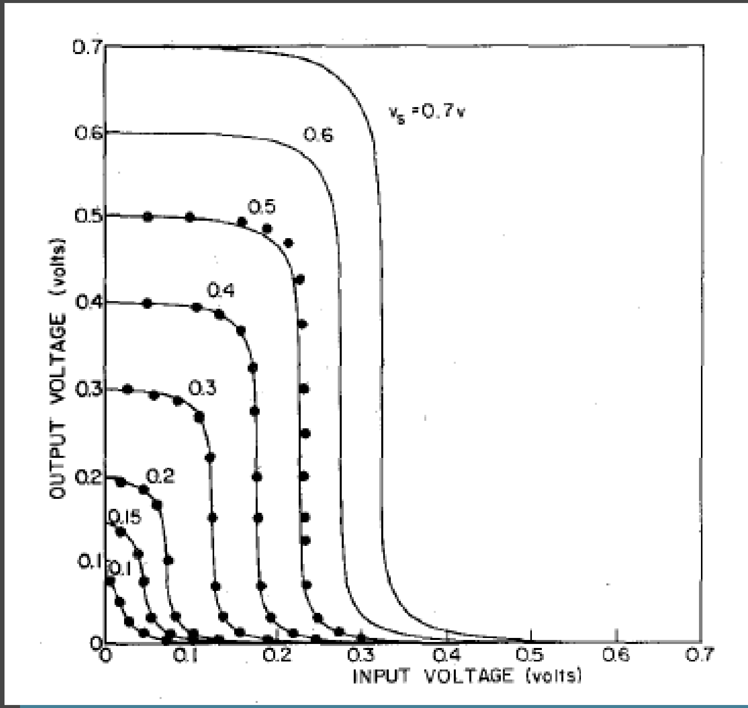
[S. MITRA, Stanford]

Trading Computation for Communication

- In "always-connected" world, energy-intensive tasks can be performed in "power-rich" backbone
  - Use energy when and where available

- This raises the immediate question: where to draw the line?
  - Cost of communication versus computation
  - Impact on performance and energy

Minimum Energy per Operation



[Swanson, Meindl (1972, 2000)]

Minimum operational voltage (ideal MOSFET):

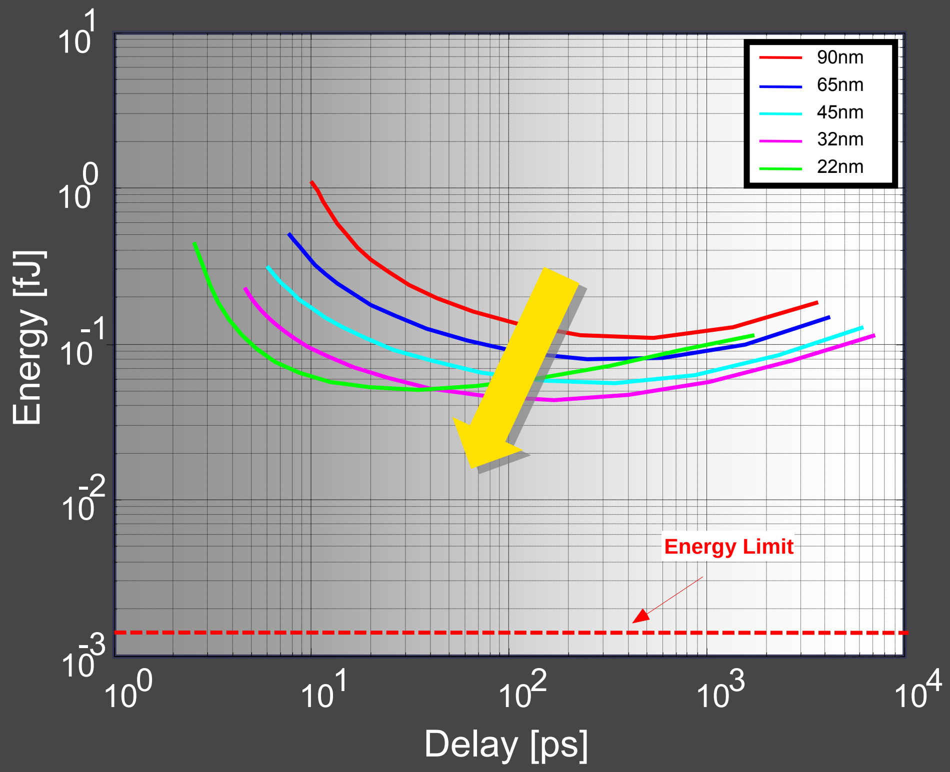
$$V_{dd}(\min) \cong 2\,(\ln 2)kT/q = 1.38kT/q = 0.036 \text{ V at } 300^\circ\text{K}.$$

Minimum Energy/Operation = kTln(2)

5 orders of magnitude below current practice (90 nm at 1V)

[Von Neumann (1966)]

The Limited Return of Further CMOS Scaling



Energy-delay curves for inverter

- 423 stage ring oscillator
- Using predictive models
- Thresholds set to nominal levels

Minimum energy/inversion scales with factor 3 (down to 40 aJ/operation) (Less than linear)

Still factor 40 above energy limit (set at 500 kTln(2))

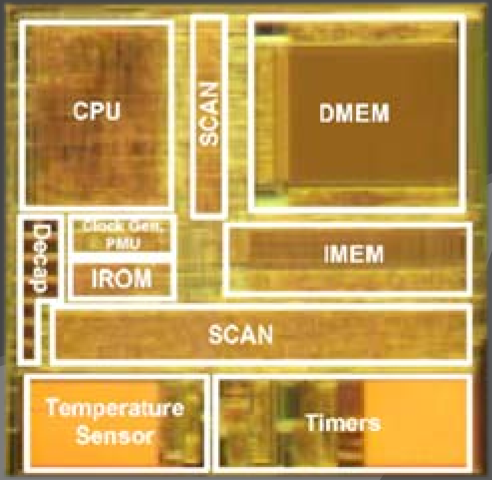
Delay scales with factor 2 (excluding 22 nm)

State of the Art in Computation

TI MSP 430 (90 nm)
3.6 mW @ 12 MHz → 300 pJ /cycle – 0.3 μW deep sleep

Philips (NXP) CoolFlux DSP (90 nm)
640 μW @ 10 MHz, 8 instr/cycle → 8 pJ/instruction

Michigan Phoenix Processor (180 nm) [VLSI08]
0.3 μW @ 106 KHz, 2.8 pJ/cycle – 29.6 pW sleep



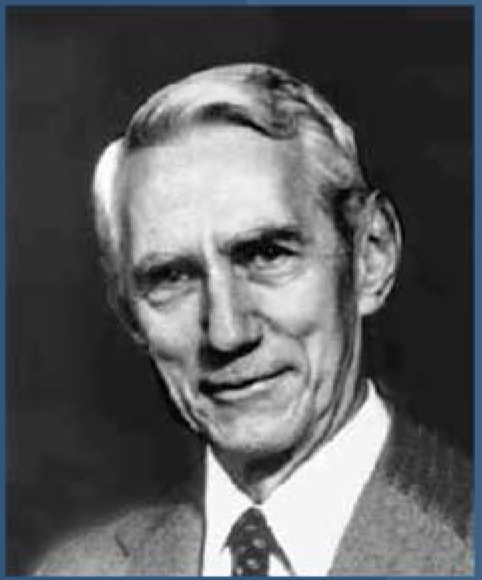
Equivalence between Communication and Computation

Shannon's theorem on maximum capacity of communication channel

$$C \leq B\log_2\left(1+\frac{P_S}{kTB}\right)$$

$$E_{bit} = P_S / C$$

C: capacity in bits/sec
B: bandwidth
P_{S}: average signal power



Claude Shannon

$$E_{bit}(\min) = E_{bit}(C/B \to 0) = kT\ln(2)$$

Valid for an "infinitely long" bit transition (C/B→0)
Equals **4.10^{-21}J/bit** at room temperature

Note: TX considered dominant from energy perspective

State of the art in communication

Power and Energy/bit metrics

802.11.g RX: 315 mW	→ 6 nJ/bit
Bluetooth RX: 67 mW	→ 166 nJ/bit
802.15.4 RX: 42 mW	→ 168 nJ/bit
Infineon RX: 24 mW	→ 160 nJ/bit

But, need to take operation conditions into account:

- Short or long packets
- Random or periodic traffic in bursts
- Spacing between packets
- Distance between nodes (for transmitter)

Hence: Energy-per-useful-bit metric

ULP Radio Comparison

- Compare 6 radios using EPUB metric
- Use same network scenario:
  - Same channel model, BER, and MAC scheme (ξ)

		PNII [2]	PN3 TR [4]	PN3 SR [6]	Favre SR [7]	Chipcon Zigbee [8]	Molnar [9]
Modulation		**DSSS - DQPSK**	**OOK**	**OOK**	**OOK**	**DSSS-OQPSK**	**FSK**
Carrier Frequency	**(MHz)**	**2400**	**1900**	**1900**	**900**	**2400**	**900**
Sensitivity @ given data rate	**(dBm)**	**-84.2**	**-78**	**-101**	**-95**	**-94**	**-94**
Required Transmitted Power @10 m	**(dBm)**	**-1.2**	**5**	**-19.5**	**-18**	**-11**	**-18**
P_{TX} (at above output power)	**(mW)**	**33**	**9.3**	**0.3**	**0.3**	**20**	**0.5**
P_{RX}	**(mW)**	**70**	**3**	**0.4**	**3.75**	**35.5**	**1.2**
Data Rate	**(kbps)**	**1600**	**50**	**20**	**100**	**250**	**20**
Header Length	**(symbols)**	**8**	**18**	**28**	**18**	**32**	**8**
E_{TOT}	**(pJ)**	**67**	**405**	**42**	**47**	**221**	**94**

E_{TOT} = asymptote of EPUB in the case of long packets

Communication/Computation Technology Comparison

- **Computation:**
  - State-of-the-art: 1 nJ/op (TI MSP 430) – 1.2 mW in operation
  - Most aggressive designs: 10 pJ/op (Coolflux) – 0.6 mW in operation
- **Communication:**
  - State-of-the-art: 300 nJ/useful bit (CC2420) – 40 mW when active
  - Most aggressive designs: 40 nJ/bit - 400 μW when active

- Energy ratio of "sending one bit" vs. "computing one instruction": Anything between 300 and 4000
- To communicate (send & receive) one kilobyte == computing at least three million instructions!

But …Are we asking the right questions?

Analysis so far takes a "component" perspective

- Wireless Communication:
  - Considered only point-to-point link
  - Assumed that TX power dominates (true for connections of 100 m and more …)
- Computation:
  - Emerged from "centralized" or "isolated" worldview
  - Assumed that the "cost of energy" is always equal, independent of where it is consumed

Let us now take a systems perspective

Assume now that we embrace the "Immersive IT" Platform vision, do the questions and answers change?

While resource constraints (that is, energy and bandwidth) remain identical

Imagine a Different World

POINT OF VIEW

Digital Wireless—*Regulation and Control Free?*

BY PETER COCHRANE
Cochrane Associates, UK

IEEE Proceedings, July 2008

Imagine for a moment that we had arrived at our present state of technological prowess without the discovery and implementation of wireless systems. Improbable and impossible, I know, but bear with me and also imagine that we had simultaneously missed out on the feast of the analog and copper era and had jumped straight to optical line systems entirely operating in digital mode. What a vastly different world it would be with near infinite bandwidth connecting every fixed node in our networks. A world where people never asked the question; why do people want bandwidth and what will they do with it? But also, a world without any form of mobility.

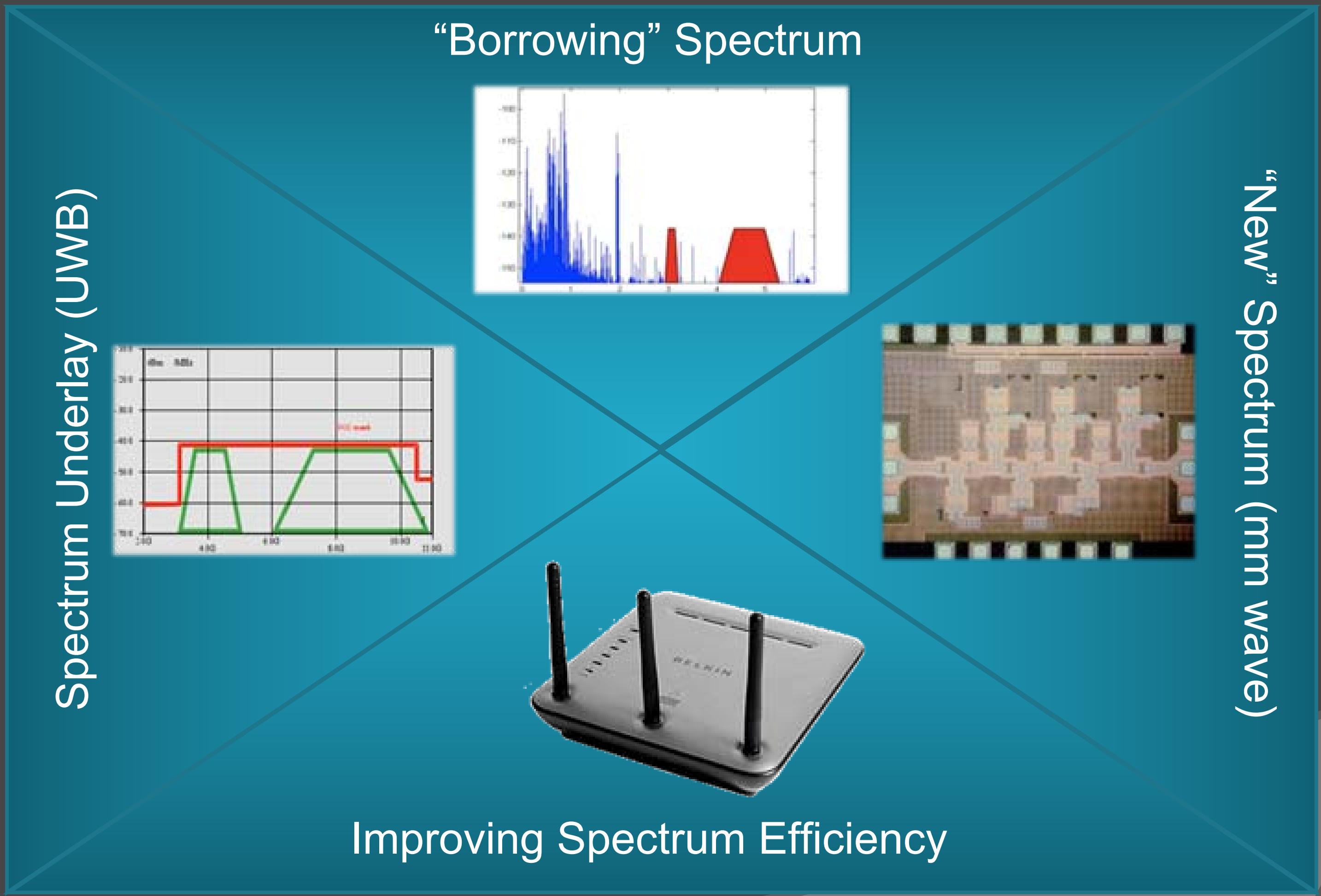
How would you build your wireless network?

A World with Unlimited Wireless Bandwidth and Always-On Coverage?

- Cognitive capabilities of terminals offer prospect of dramatic increase in attainable wireless data-rates
- Collaboration among terminals and infrastructure essential to accomplish cognitive promises, while providing reliability
  - Increases efficiency
  - Provides reliability
  - Opens door for collaboration between heterogeneous services or standards
- Connectivity Brokerage as the new operational (as well as business) paradigm

A Fundamentally Disruptive Technology

A World with Unlimited Wireless Bandwidth and Always-On Coverage?

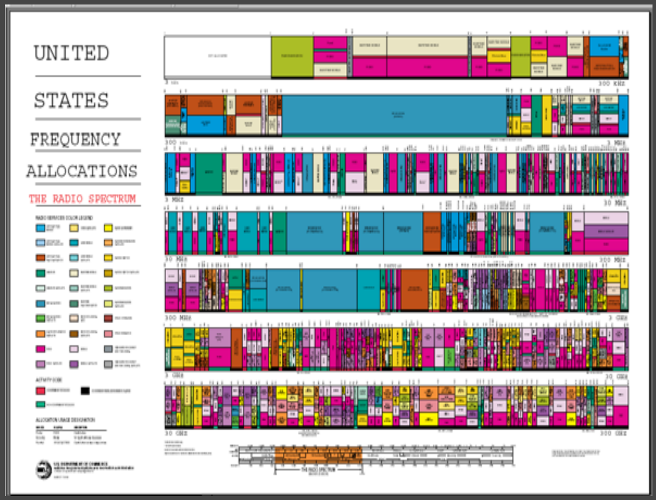


Some exciting technology developments

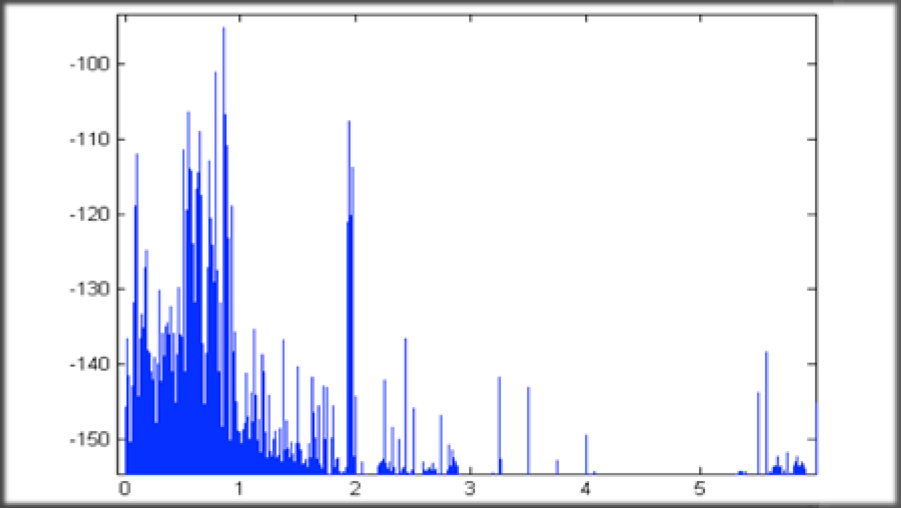
Spectrum Shortage?

- Existing spectrum policy has full allocation but poor utilization

Allocation

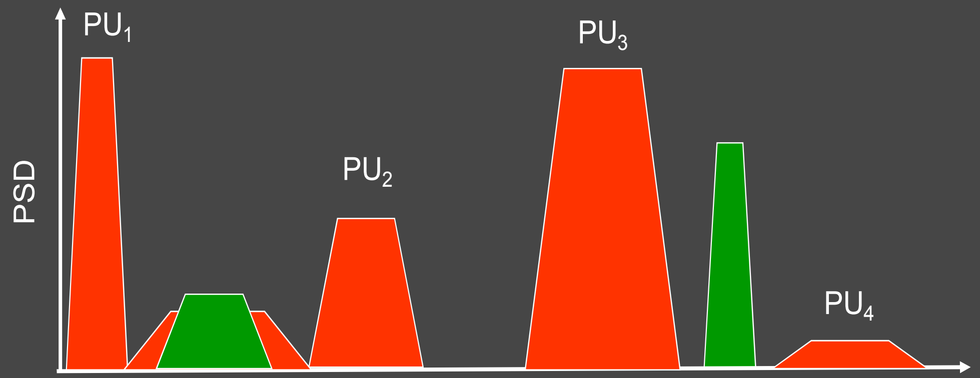


Utilization



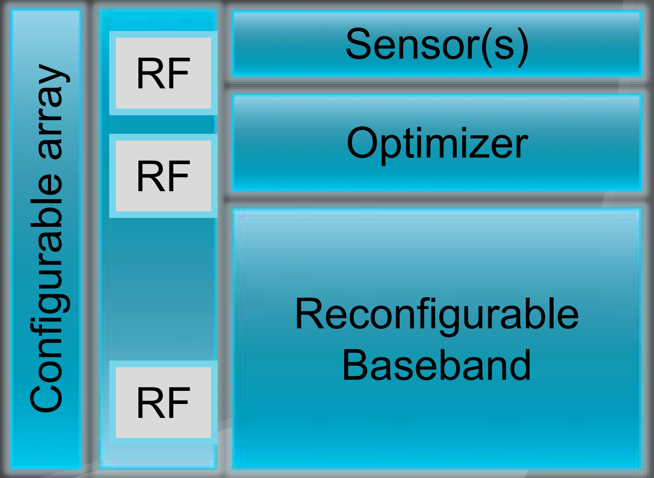
The cognitive radio strategy is to sense the spectrum and to only transmit if there will be no interference

Cognitive Radio to Enable Dynamic Spectrum Allocation



First Experiment in Cognitive:
TV Bands @ 700 MHz
(IEEE 802.22)

- **Sense** the spectral environment over a wide bandwidth
- Reliably **detect** presence/absence of primary users and/or interferers
- **Rules** of sharing the available resources (time, frequency, space)
- **Flexibility** to adjust to changing circumstances (power, freq. band)



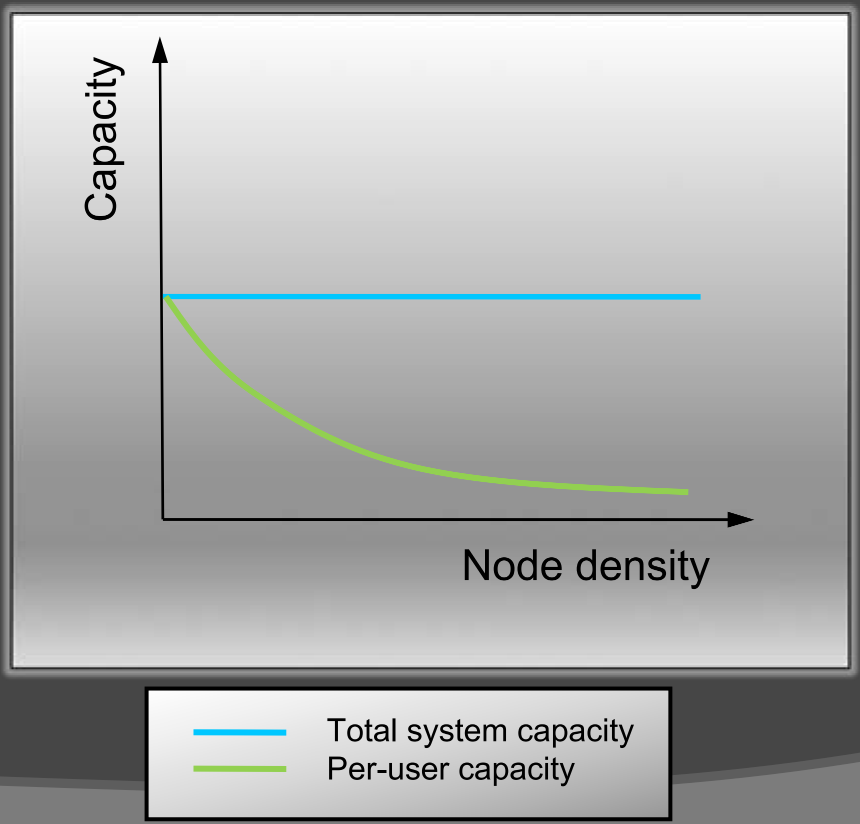
Cognitive terminal

Increased bandwidth availability reduces TX/RX energy cost

The Power of Collaboration

Conventional wireless mindset:

- Services compete!
  - Example: Bluetooth, WIFI and Zigbee
- Adding terminals degrades user capacity



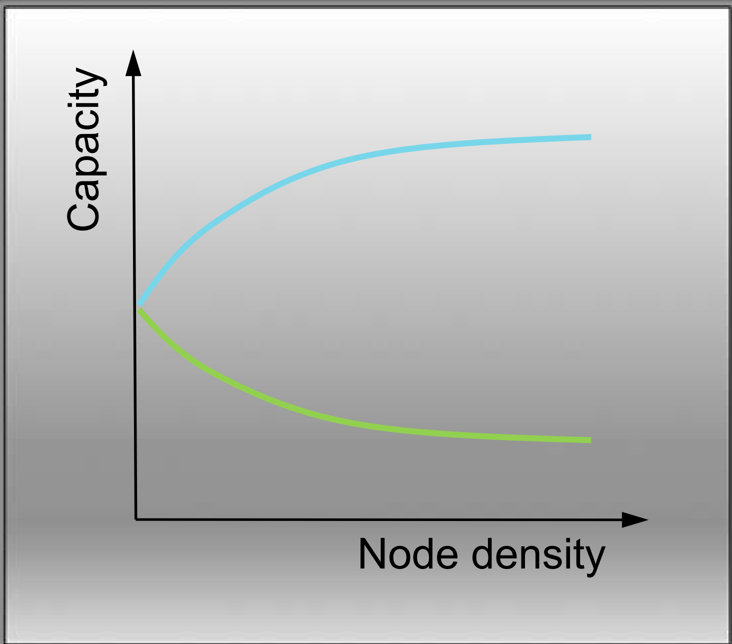
Collaboration is essential for better spectrum utilization

- A single terminal or base-station has only limited perspective
- Working together leads to better capacity, efficiency, coverage and/or reliability

Need to look beyond the single link!

The Power of Collaboration

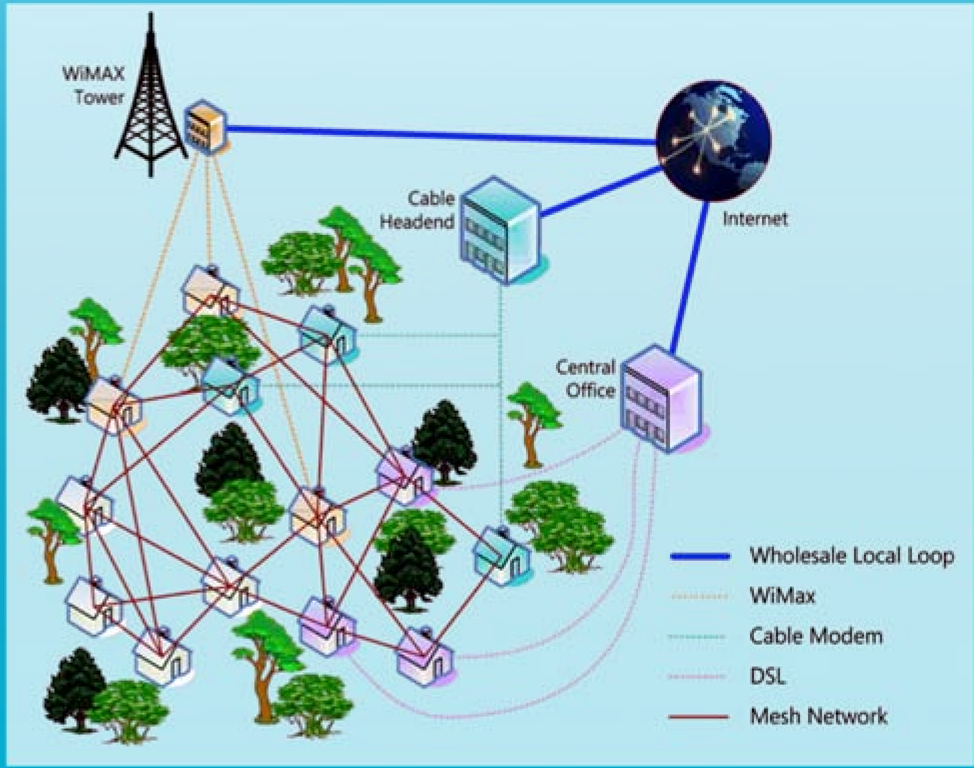
Packet Multi-hop



[Ref: Gupta/Kumar'00]

Reduce the effective link distance

Wireless Meshes

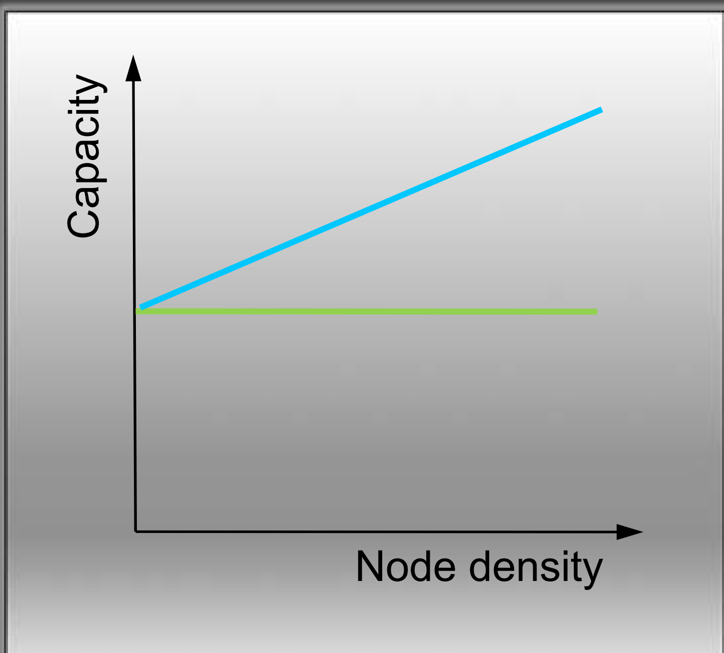


- Connect the unconnected
- Increase "perceived user value"
- Provide reliability in case of failure

[Courtesy: R. Chandra, Microsoft Research]

The Power of Collaboration

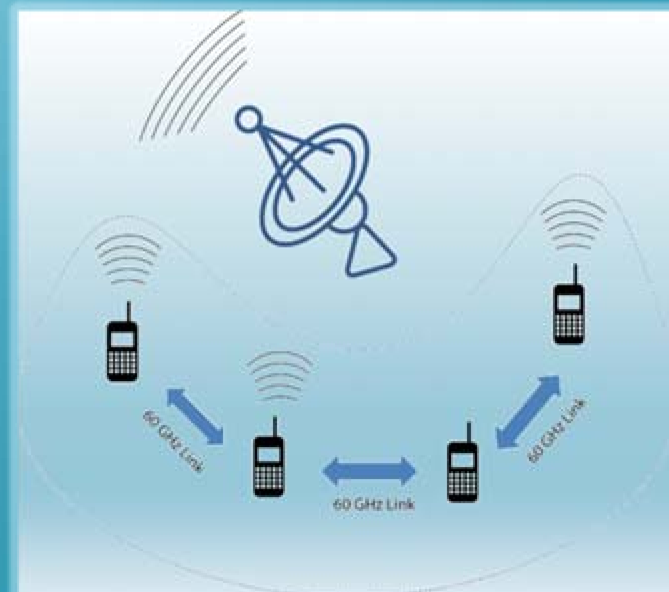
Collaborative Diversity



[Ref: Ozgur/Leveque/Tse'07]

Reduce the effective link distance

Collaborative MIMO



- Construct large effective-aperture antenna array by combining many terminals, increasing throughput or coverage
- Local ad-hoc network between terminals

Cognitive-Collaborative Networks: The Challenges

- How to manage degrees of freedom?
  - Frequency/spatial utilization, collaboration, topology
- So that some global and user goals are met
  - Cost, User experience, Life time
- While …
  - Providing absolute reliability
  - Hiding complexity
  - Providing security and access control
  - Dealing with legacy systems

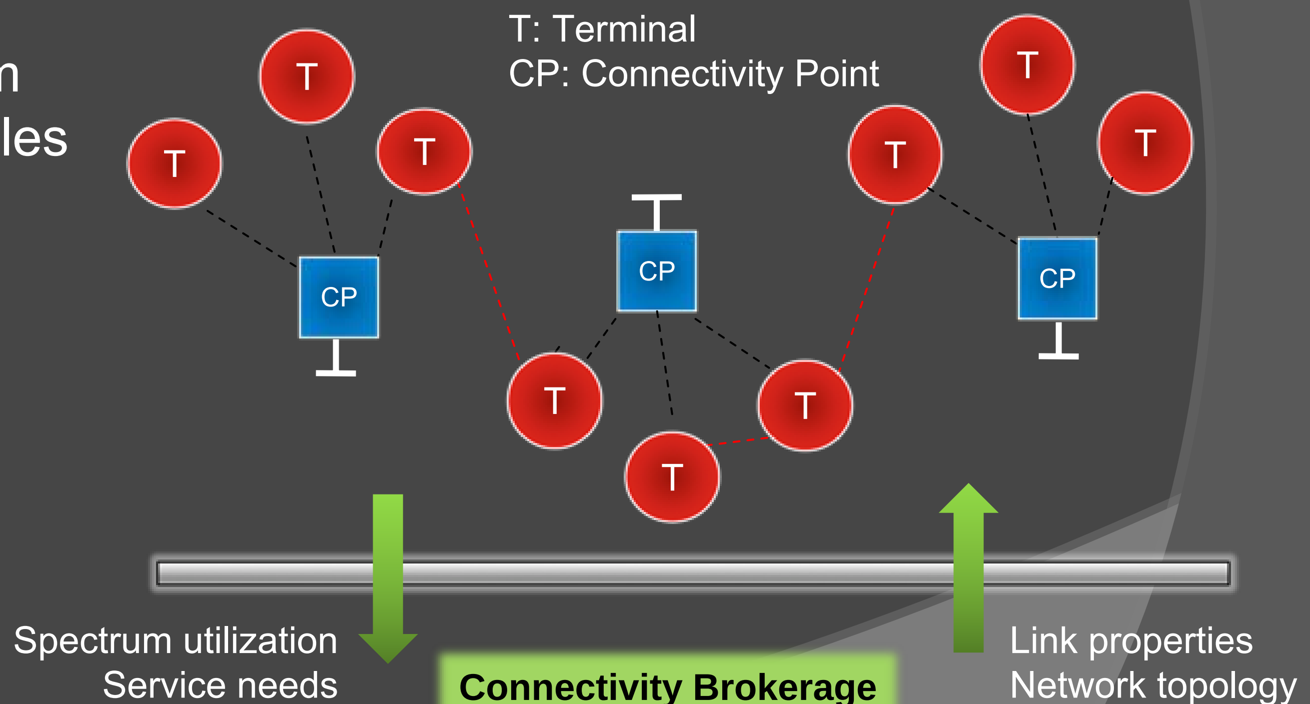
Making Cognitive/Collaborative Work

Connectivity Brokerage (*) as a Distributed OS

Functional entity that enables collection of terminals to transparently connect to backbone network or each other to perform set of services

While optimizing utilization of spectrum under policy rules, rules of engagement and security constraints.

(*) Term first coined by Adam Wolisz (TU Berlin)



A Technical as well as Economic Proposition

A Renewed Look at Ubiquitous Wireless

- A system-level perspective reformulate the computation/communication trade-off questions
  - Network density makes communication cheaper at no cost in capacity
  - Increased bandwidth availability simplifies receiver and transmitter energy cost
  - Cost of computation should be measured (weigthted) by where it is used
- Cost metrics should be redefined from a system perspective

A Call to Arms:
Benchmarks and Metrics of Old Won’t Do

- Mostly based on “high-performance” or personal computer – style applications
  - SpecMarks, EEMBC, SSP
- Or link-level communications
- Traditional quality metrics for design …
  - Performance (e.g. MIPS)
  - Energy efficiency (MIPS/W)

© D. Rosandich, cartoonstock.com

- are second-order or irrelevant in Immersive Wireless Networks
  - Societal IT systems are rarely performance-constrained
  - Energy-efficiency is a function of where it is consumed and when

The New Benchmarks and Metrics

- New Benchmarks Libraries Must Extend Beyond the Component
  - “Workloads of the Future”
- Relevant metrics for SiS Systems
  - “User experience per unit energy”
  - System Latency
  - Reliability/Liability
  - Complexity/composability

Needs Joint Effort By Industry and Academia

Concluding Reflections

- Ubiquitous always-connected wireless radically transforming the Information Technology Arena
  - Towards truly Immersive Systems
  - Cognitive Collaborative Wireless a powerful disruptive paradigm
- Complexity, heterogeneity, reliability and power present formidable challenges
- Efficient realization of Immersive Systems Requires a Systems Vision with Communications and Computations Au Par
- Broad collaboration between systems and semiconductor industries, as well as industry and academia needed
  - Need for new benchmark libraries
  - Need theory of system design

A call to action! These are exciting times again …