
VLSI Chips for future HEP system

A. Marchioro / CERN-EP

Topics

- Key issues for future experiments
 - Requirements for Chips in LHC and beyond
 - Cost per channel
 - How to reduce system cost
 - Power
 - Why is it critical
 - How to improve
 - A structured approach for saving power in the design of chips for HEP
 - Power saving techniques in analog and digital circuits
 - Data compression techniques
 - Examples of different approaches for saving power

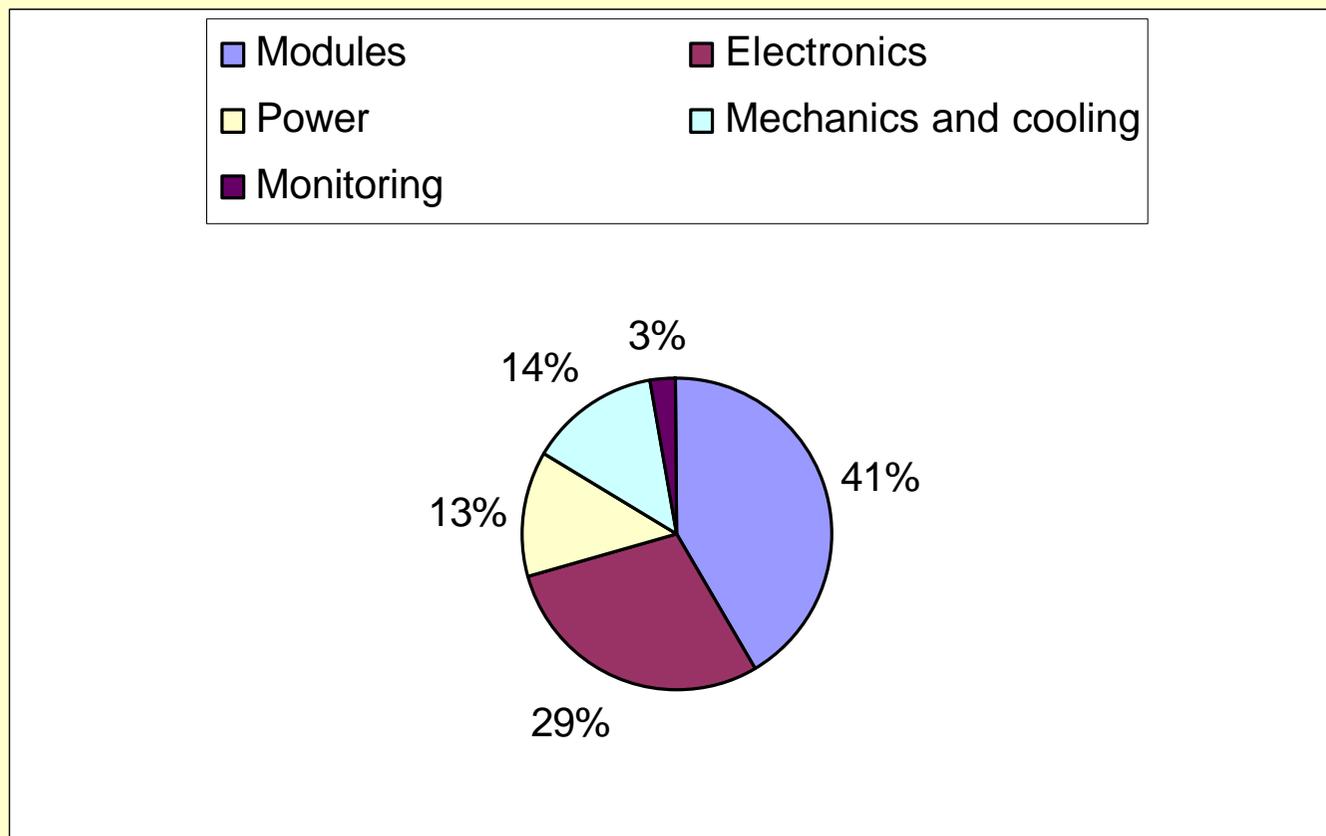
What comes after

- SLHC
 - Luminosity: $\sim 10^{35} \text{ fb}^{-1}$
 - Beam cms energy: \sim same
 - Radiation levels (5 years):
200 Mrad @ 7 cm, 40 Mrad @ 20 cm
 - Compensate for higher intensity through higher segmentation
 - Cost: lower than current !
 - Power/channel must decrease

System Cost 101

- = Cost of FE electronics *(chips + overhead)*
- + Technology Access
- + Cost of Data Links *(type: analog or digital, quality)*
- + Electronics in Counting Room
- + Installation and Operating Costs
- + Infrastructure Overhead *(e.g. PSU, cooling, cabling)*
- + R&D Time
- + Design Time *(complexity, bugs)*
- + Testing & Packaging
- + Maintenance Cost

System cost: the CMS tracker



Module = Si sensor + mechanical mounting + pitch adapter

More than 40% of the cost is in the modules.

About 75% of the cost of the modules is the cost of the sensors.

How to reduce system cost

- Cost reduction can be achieved through:
 - Simply wait for today's "fancy" technology to become tomorrow's "commodity"
 - Sooner or later 0.1 μ m BiCMOS with 10 metal layers will become available through MOSIS ...
 - Use more efficiently a given technology
 - Multiplex more channels into a given link
 - Use microelectronics to lower cost of some other part, invest heavily on VLSI
 - Have a look at some commodity electronics you can purchase today and "think" ...
 - For example: CD players, ADSL modems, Gigabit Ethernet

Power

Why is it important

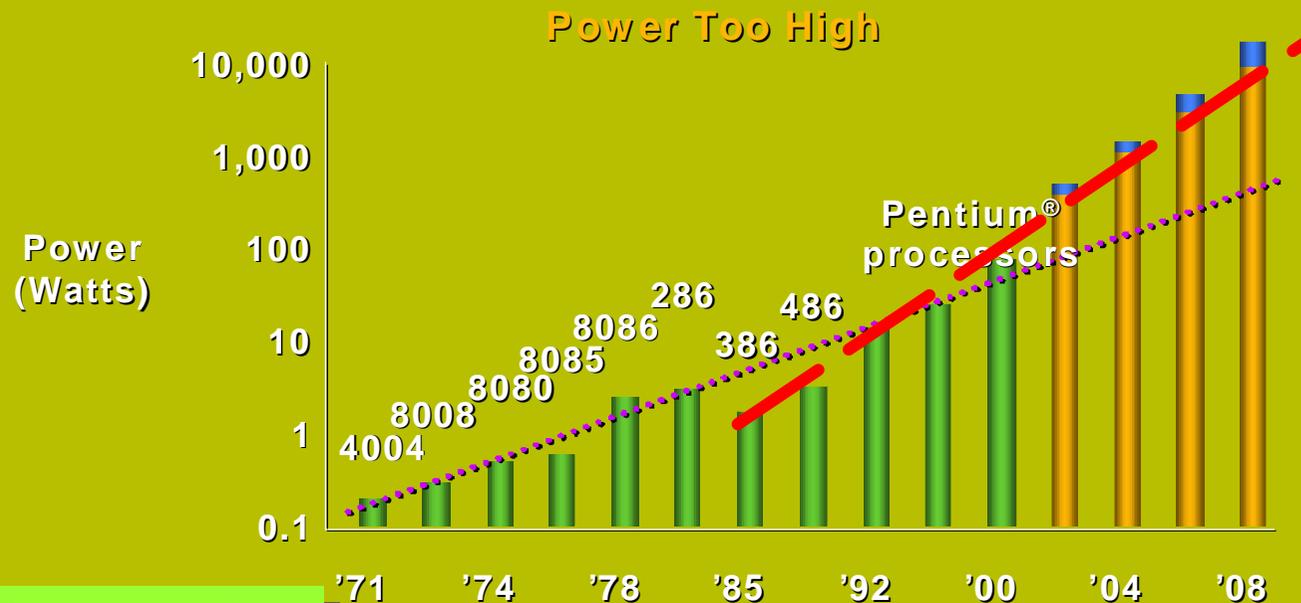
How does it relate with other system aspects

How to reduce it

A four levels approach to combat the problem

Power: Not only our problem...

by End of the Decade



Source: P. Gelsinger, Intel Corp.
Presentation at the ISSCC 2001

But this is not the worst of it...

© 2001 IEEE International Solid-State Circuits Conference

© 2001 IEEE

Issues: what matters

	<i>Power</i>	<i>Dyn-Range</i>	<i>Electronic noise (S/N)</i>	<i>Linearity</i>	<i>Timing Resol.</i>
<i>Pixel</i>	High (< 0.1 mW)	Low (few bits)	High ~200/~5000 e ⁻	Low	Medium (BX ident)
<i>Tracker</i> (<i>Si strips</i>)	High (~ 1 mW)	Medium (6-8 bit)	Medium ~1000/~20000 e ⁻	Medium	Medium (BX ident)
<i>Calorimeter</i>	Medium (100 mW)	High (10-15 bit)	High ~ 500 e ⁻	High	Medium ()
<i>Muon</i>	Low (>10 mW)	Low (1 bit)	Medium	Low	High (drift time)

What to attack

	<i>Technology</i>	<i>Circuit</i>	<i>Architecture</i>	<i>Algorithm</i>
<i>Power</i>				
<i>Dyn-Range</i>				
<i>Noise</i>				
<i>Timing</i>				

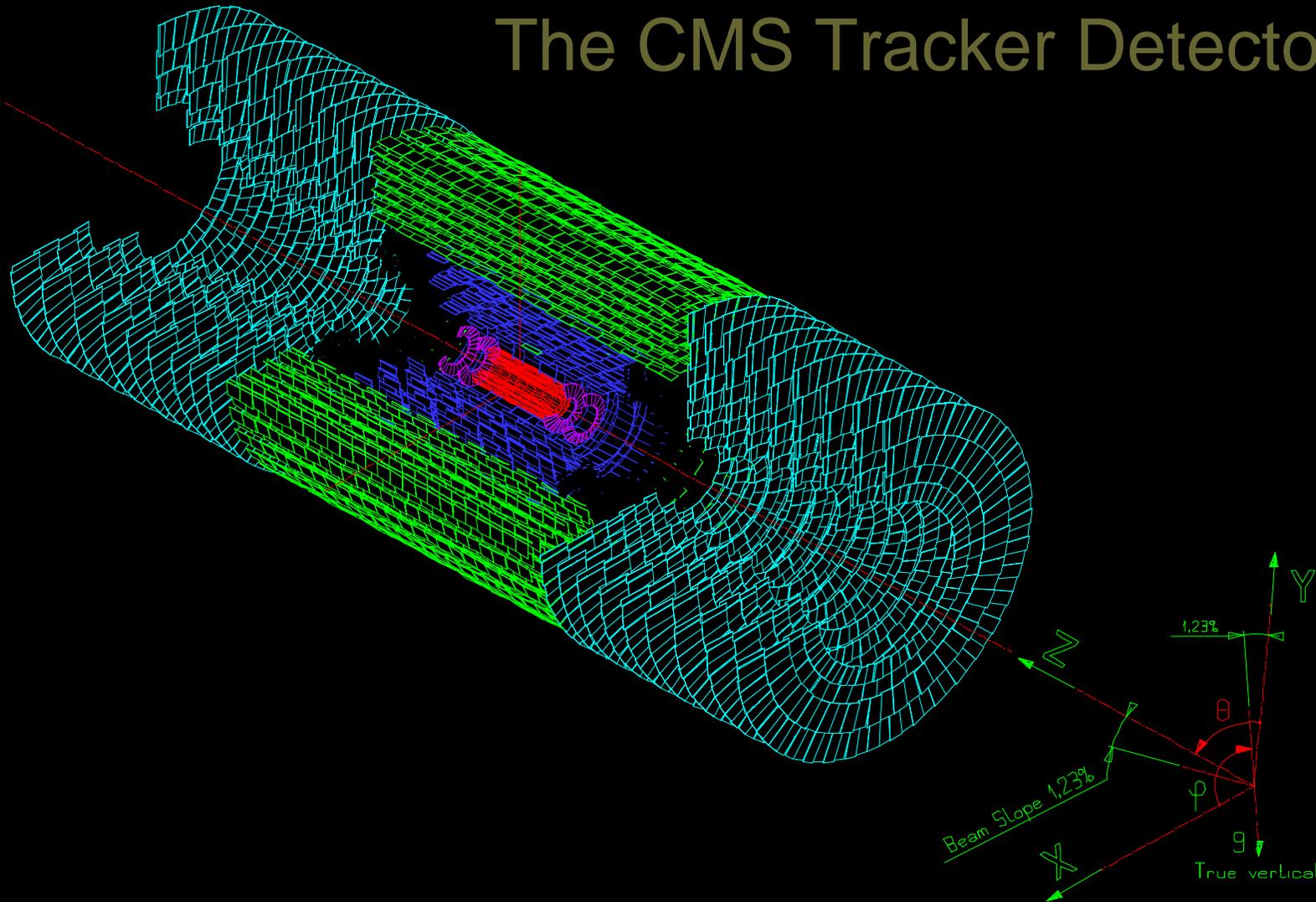


Work necessary



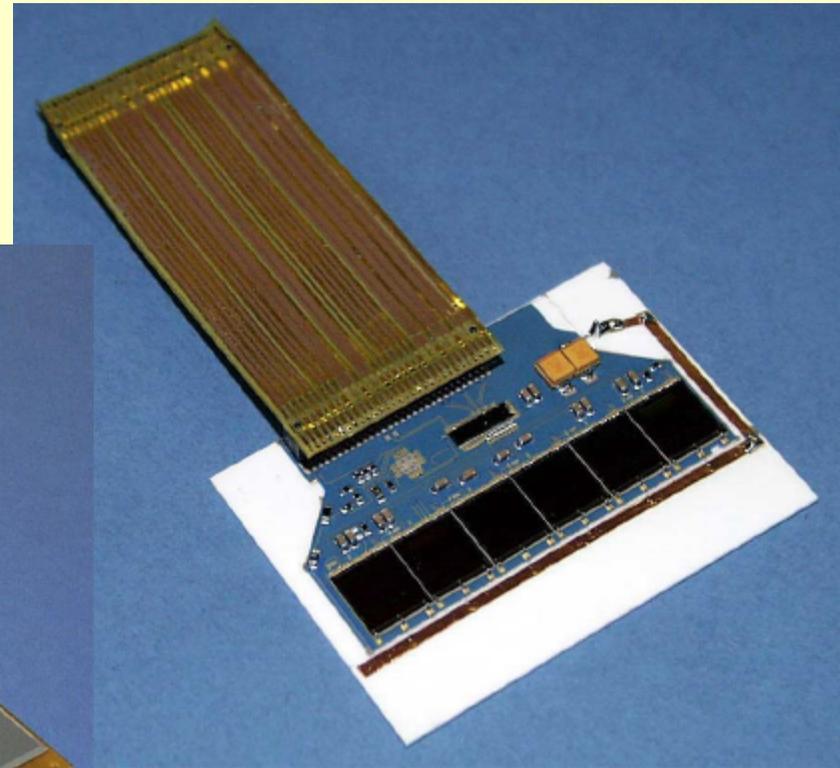
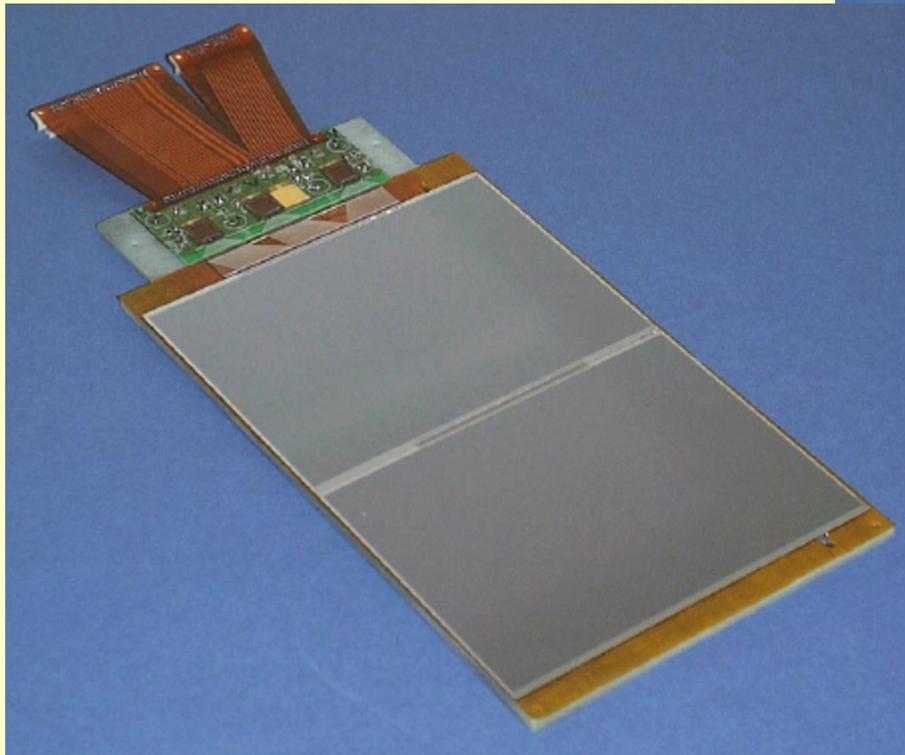
Clever idea necessary...

The CMS Tracker Detector



Fichier<TL012562.pl>

Si Tracker Detector module



Power in CMS Tracker: worst case ¹⁾

- Total # channels: 75,500 FE chips x 128 = ~10M
- Power/FE: 2.3 mW/channel
- Pwr/ch data TX: ~0.6 mW/channel
- Supply: 2.5 V and 1.25 V, $P_{\text{tot}} = \sim 30$ kW
- Total FE currents: $I_{\text{DD125}}: \sim 7.5$ kA, $I_{\text{DD250}}: \sim 6.5$ kA
- Remote supplies
 - ▶ # of service cables: 1,800
- Power in the cables: > 75 kW
- Cross section of power cables and cooling pipes directly proportional to power dissipated !

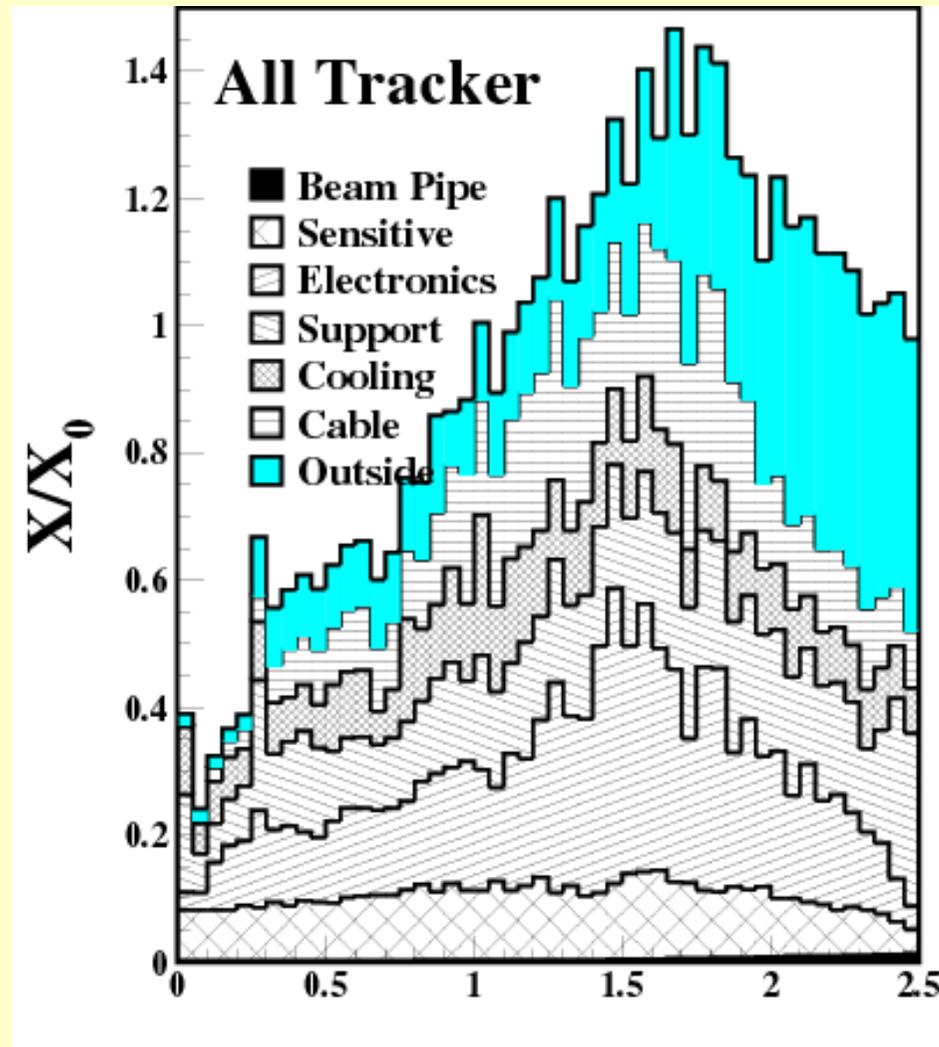
1) Worst case is computed after 10 years of irradiation

What if SLHC ?

- If 5x luminosity ^[1] tracker would require:
 - 2 x speed
 - 2x segmentation ? 20 M channels
 - 25% higher occupancy
- Assuming that (magically) FE power/ch remains the same
 - $P_{\text{tot}} = 60 \text{ kW}$
 - $P_{\text{cables}} = 150 \text{ kW}$
 - Cables_? : **double**, cooling pipes: **double**

[1] This is purely hypothetical, actual numbers may change

Material budget in CMS Tracker



Saving power in VLSI circuits

- Technology scaling
 - Advanced technology, packaging, scaling
- Circuit and logic topologies
 - Device sizing, Logic optimization (digital), Power down (sleep) mode
- Architecture (analog and digital)
 - Signal features (e.g. correlation), Data representation, Concurrency, Partitioning
- Algorithms
 - Regularity, Data Representation, Complexity

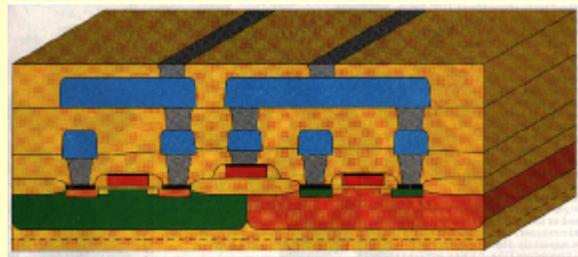
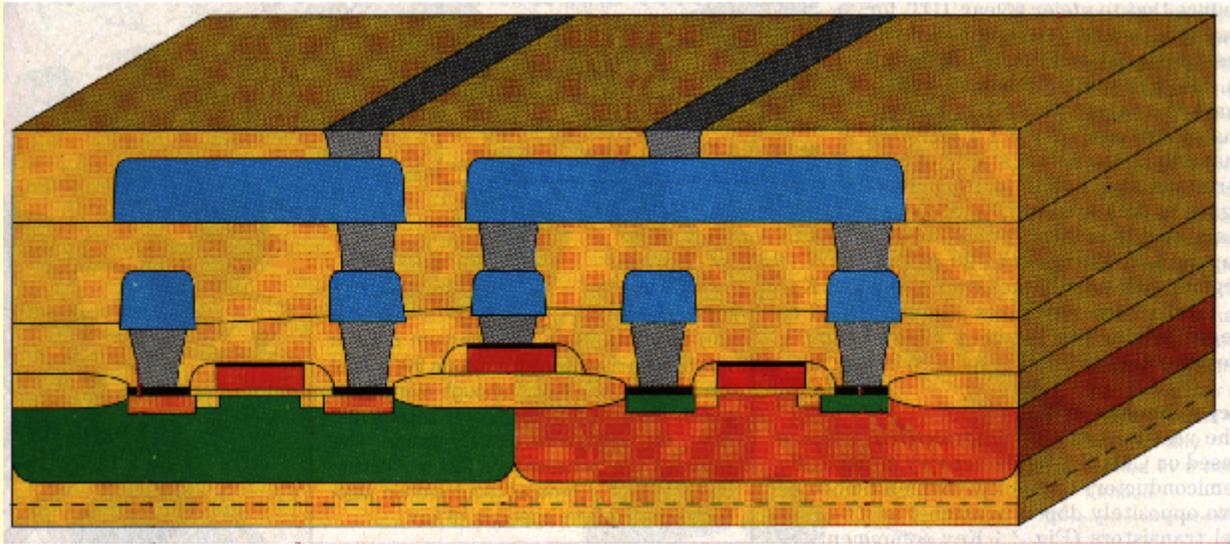
Saving power: Technology

LHC
Start

SLHC
Start

	1997	1999	2001	2003	2006	2009	2012
<u>Overall Characteristics</u>							
Transistor density ⁽²⁾	3.7 M/mm ²	6.2 M/mm ²	10 M/mm ²	18 M/mm ²	39 M/mm ²	84 M/mm ²	180 M/mm ²
Chip size ⁽³⁾	300 mm ²	340 mm ²	385 mm ²	430 mm ²	520 mm ²	620 mm ²	750 mm ²
Local clock frequency ⁽⁴⁾	750 MHz	1.25 GHz	1.5 GHz	2.1 GHz	3.5 GHz	6 GHz	10 GHz
Power supply voltage ⁽⁵⁾	1.8-2.5V	1.5-1.8V	1.2-1.5V	1.2-1.5V	.9-1.2V	.6-.9V	.5-.6V
Maximum power ⁽⁶⁾	70 W	90 W	110 W	130 W	160 W	170 W	175 W
<u>Technology Requirements</u>							
μP channel length ⁽¹⁾	.20 μm	.14 μm	.12 μm	.10 μm	70 nm	50 nm	35 nm
DRAM ½ pitch ⁽¹⁾	.25 μm	.18 μm	.15 μm	.13 μm	.10 μm	70 nm	50 nm
T _{ox} Equivalent ⁽⁷⁾	4-5 nm	3-4 nm	2-3 nm	2-3 nm	1.5-2 nm	<1.5 nm	<1.0 nm
Gate Delay Metric CV/I ⁽⁷⁾	16-17 ps	12-13 ps	10-12 ps	9-10 ps	7 ps	4-5 ps	3-4 ps
Solutions Exist							
		Solutions Being Pursued				No Known Solution	

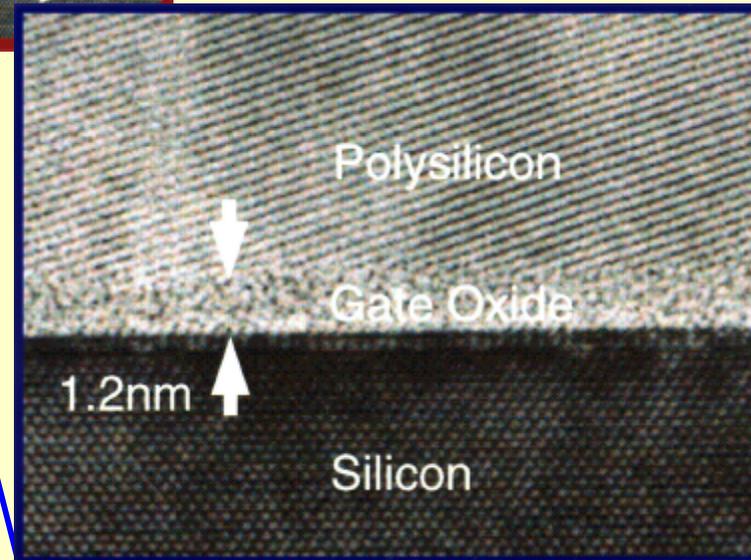
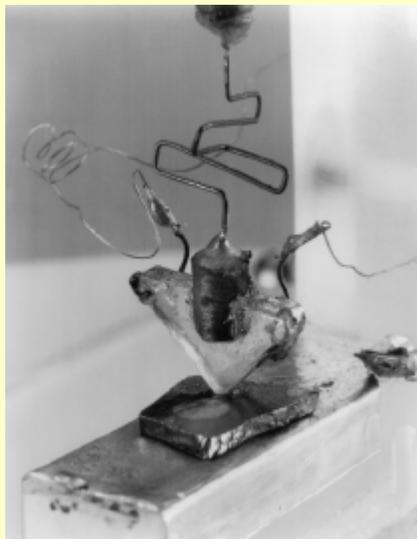
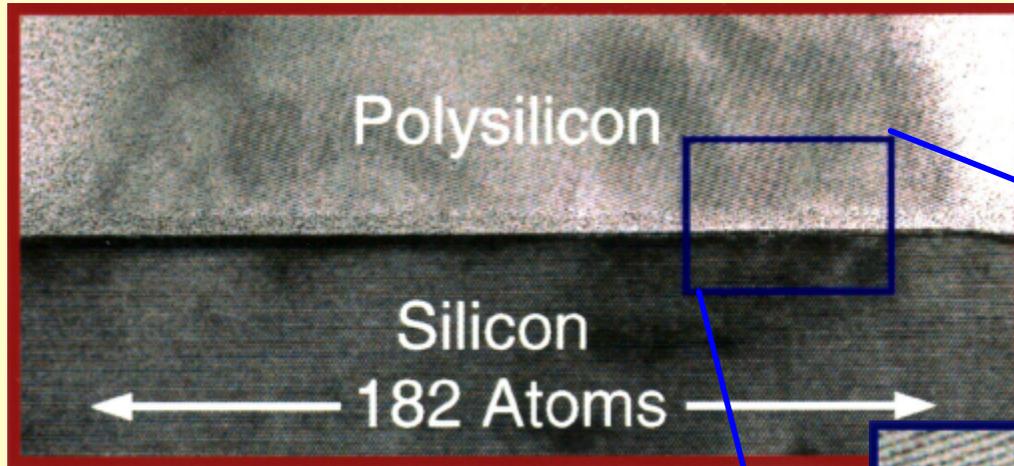
Scaling



Geometrical dimensions scale by a ,
E-field should remain constant,

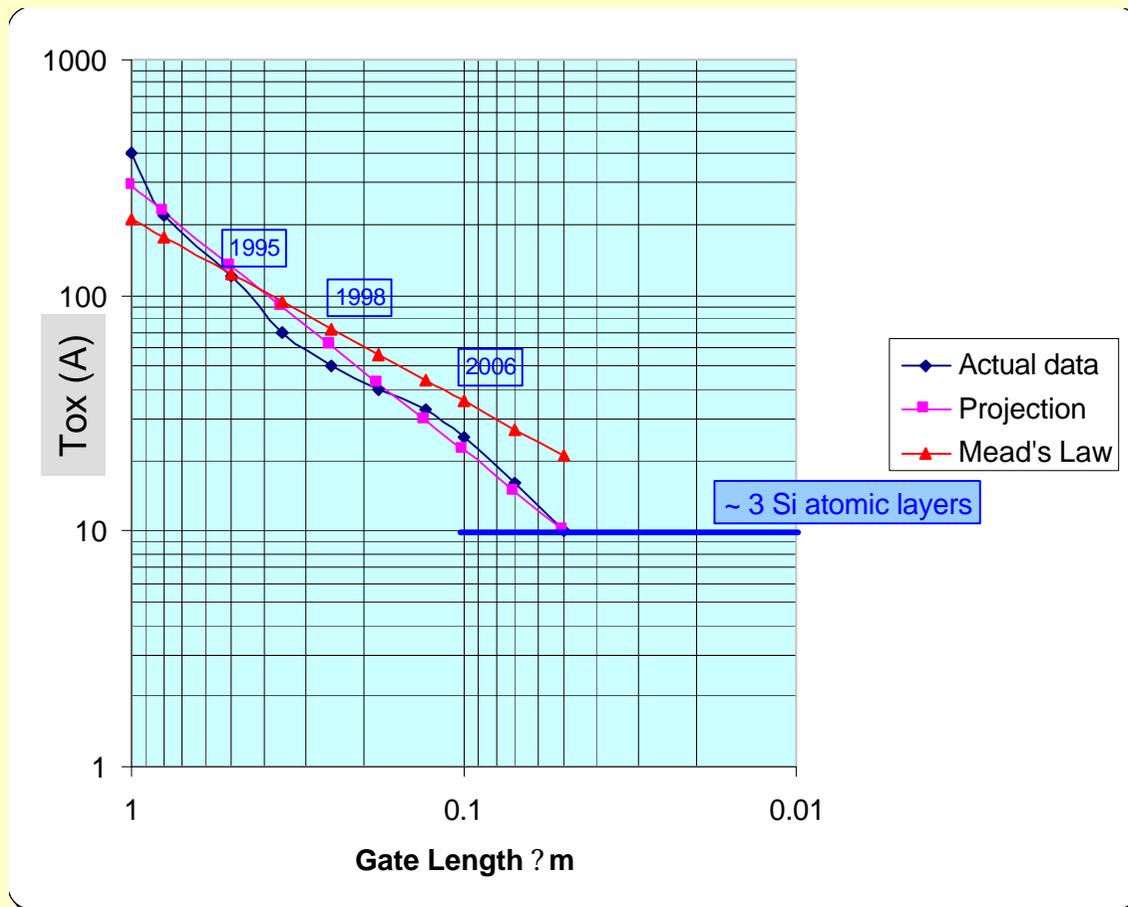
- C increases
- R increases
- Power density increases

Transistors from 1947 to 1997



ATT 0.1 μ m device, 1997

When will it stop ?



Carver Mead's Law

$$t_{ox} = 210 * L^{0.77}$$

from C. Mead, 'Scaling of MOS Technology to Submicron Feature Sizes',
Journal of VLSI Signal Processing,
July 1994

Getting there !



PIPELINE

Today's News and Analysis

The 3-Atom-Thick Transistor

Intel's breakthrough could lead to a 20-GHz chip by 2007.

By Cade Metz
June 11, 2001

Moore's Law shows no signs of abating. This weekend at the Silicon Nano Electronics Workshop in Kyoto, Japan, Intel researchers said they had produced a silicon transistor that measures only 80 atoms wide and 3 atoms thick and is capable of switching on and off 1.5 trillion times a second. Constructed with equipment used to build today's silicon chips, the tiny electronic switch is evidence that basic computer hardware will continue to improve at the current pace until the end of the decade.

Ideal “Analog Technology”

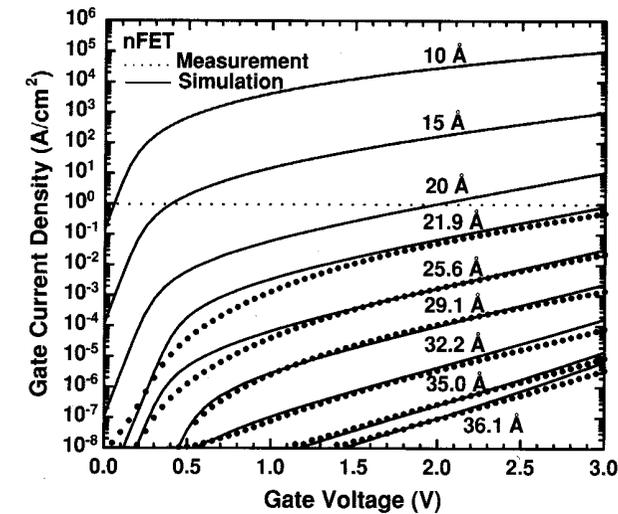
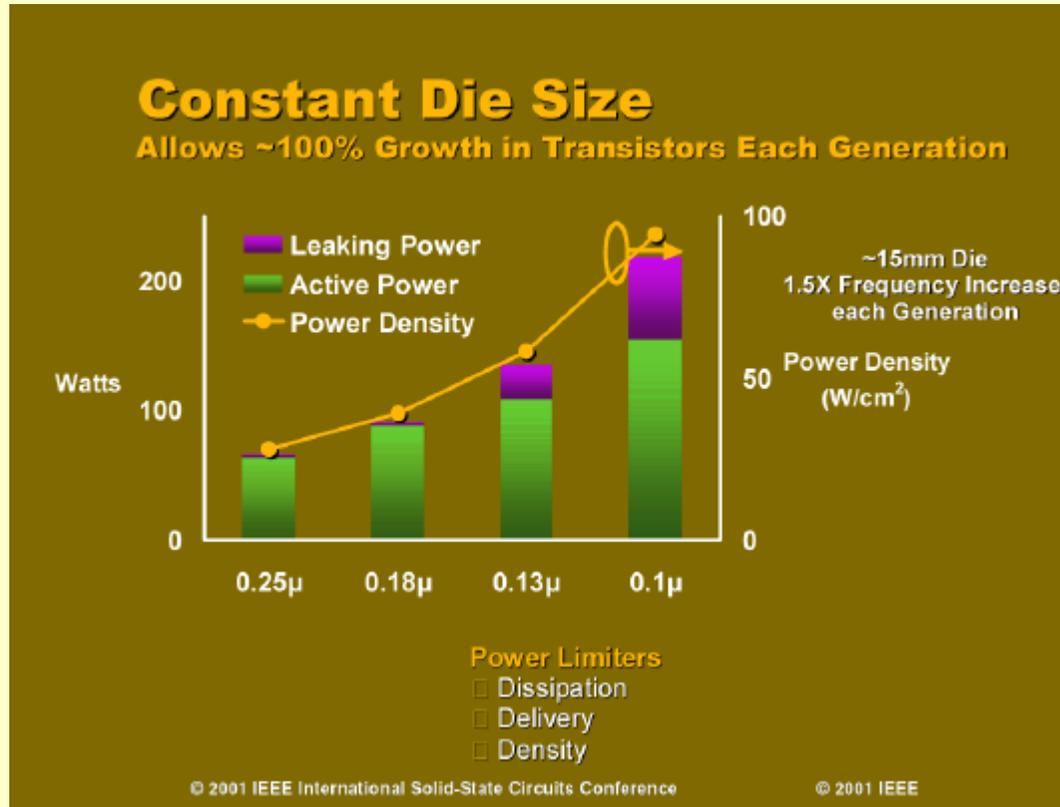
*...Several considerations suggest that the
0.35 μ m or perhaps the 0.25 μ m
[BiCMOS technology] will be adequate...*

B. Gilbert,
“Analog at Milepost 2000”,
Proc. of the IEEE, 3/2001

Reasons:

1. Cost of high performance technologies
2. No need for extreme scaling in analog
3. Limited supply voltage
 - Limited topologies
 - Limited signal swing (dyn-range)

Problem: Device leakage



Source: D. Frank et al.,
Proceedings of the IEEE, 3/2001

Source: P. Gelsinger, Intel Corp.
Presentation at the ISSCC 2001

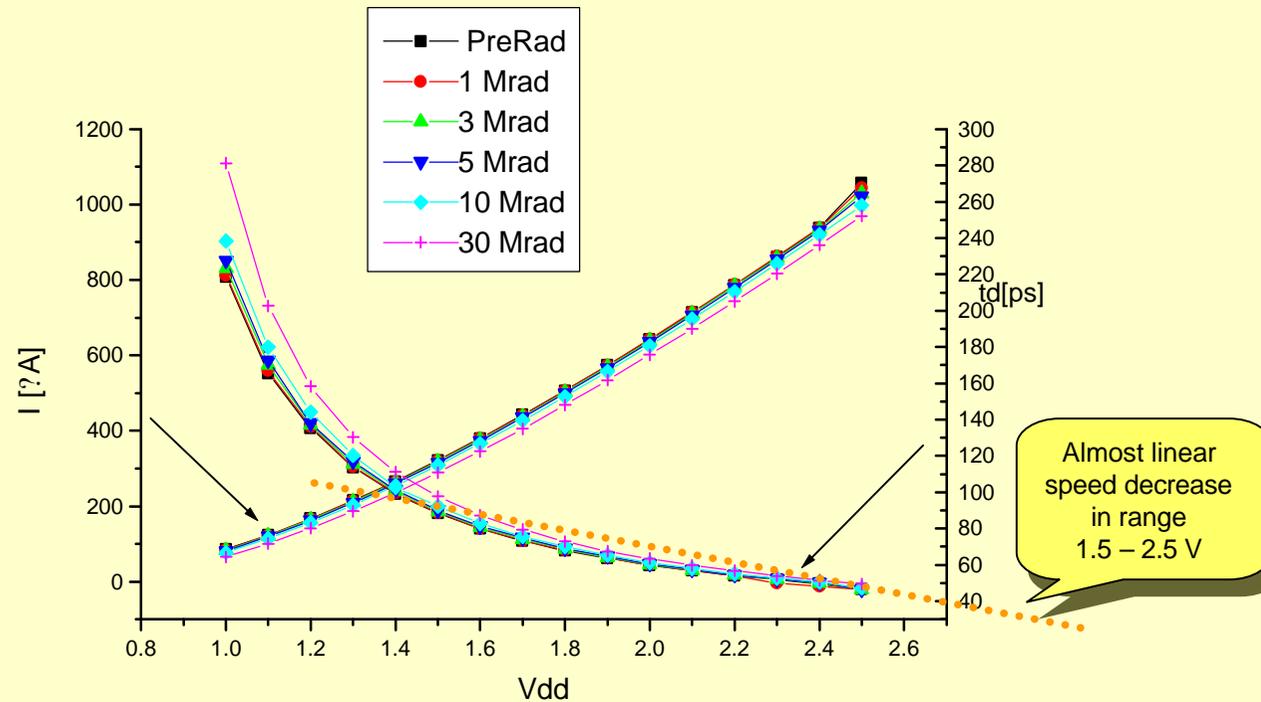
Saving power in digital circuits (1)

$$P = P_{\text{switching}} + P_{\text{short-circuit}} + P_{\text{leakage}}$$

$$P = \alpha_{0>1} * C_L * V_{DD}^2 * f_{\text{clk}} + I_{\text{sc}} * V_{DD} + I_{\text{leak}} * V_{DD}$$

- How to optimize:
 - Reducing V_{DD} helps everywhere !
 - 1st term:
 - ? ? : logic design
 - C_L : technology, circuit design, architecture
 - 2nd term: technology, circuit design
 - 3rd term: technology

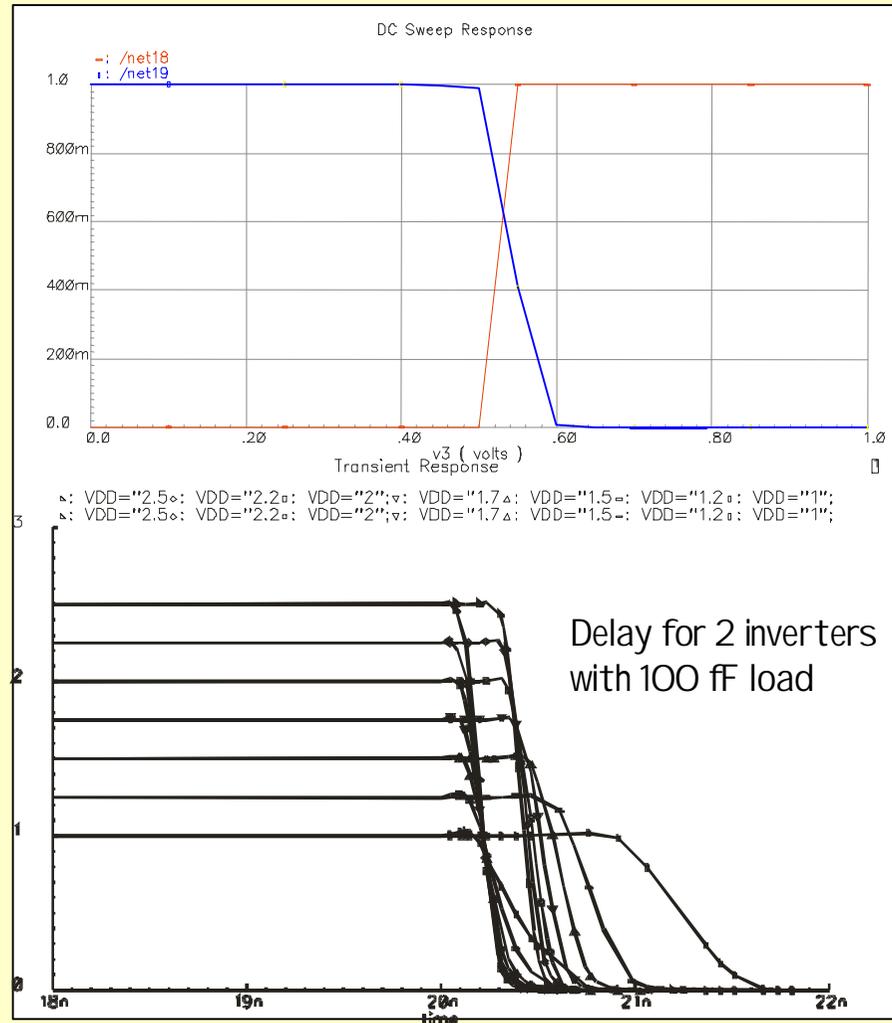
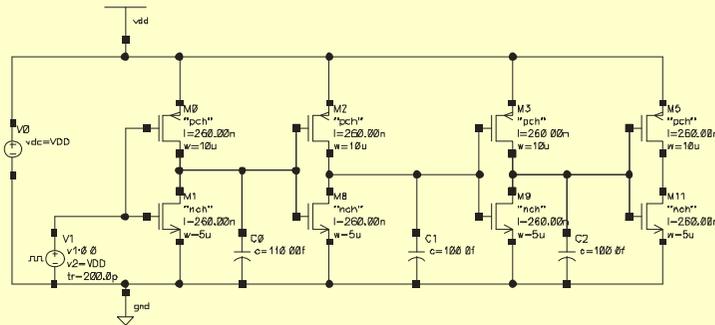
Deep Sub Technology helps !



1001 Elements Ring Oscillator in 0.25 μ m

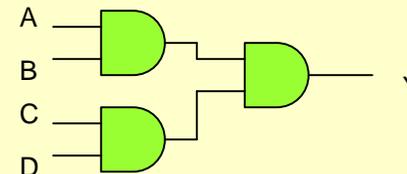
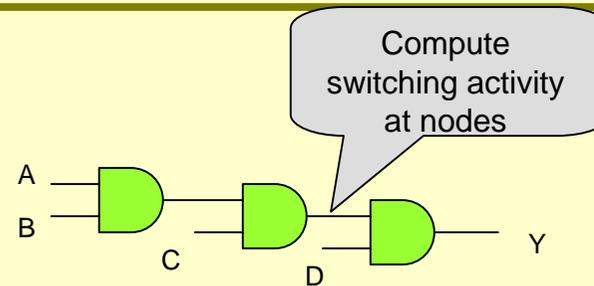
Power: 0.05 μ W/gate*MHz @ 1 V
0.25 μ W/gate*MHz @ 2.5 V

40 MHz operation @ 1V supply



Saving power in digital circuits (2)

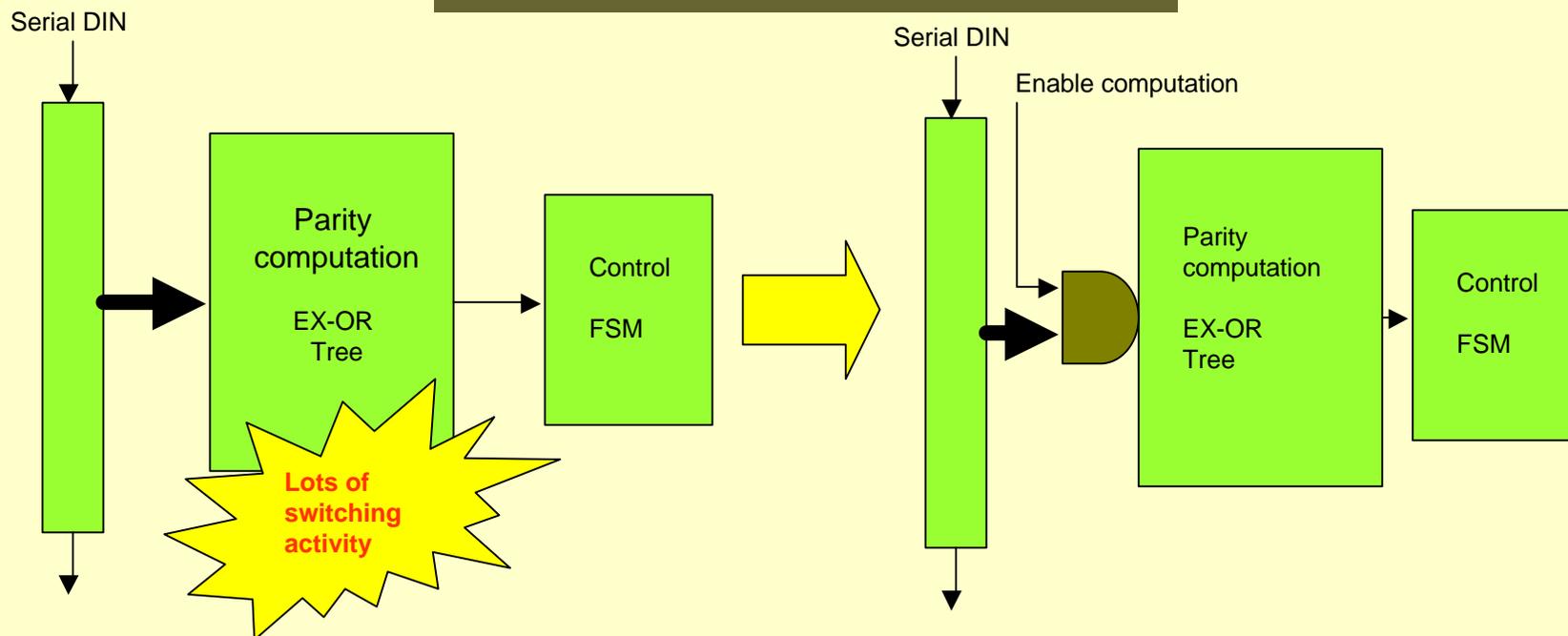
- Optimize logic for power
 - Build equivalent function
 - Compute switching activity at nodes
 - Select configuration with minimum activity
- Seems easy, but:
 - Tools don't work this way (but changing)
 - Timing can be tricky
 - Signals statistics may fool you
 - Don't forget that $C_{\text{wire}} \gg C_{\text{gate}}$



Saving power in digital circuits (3)

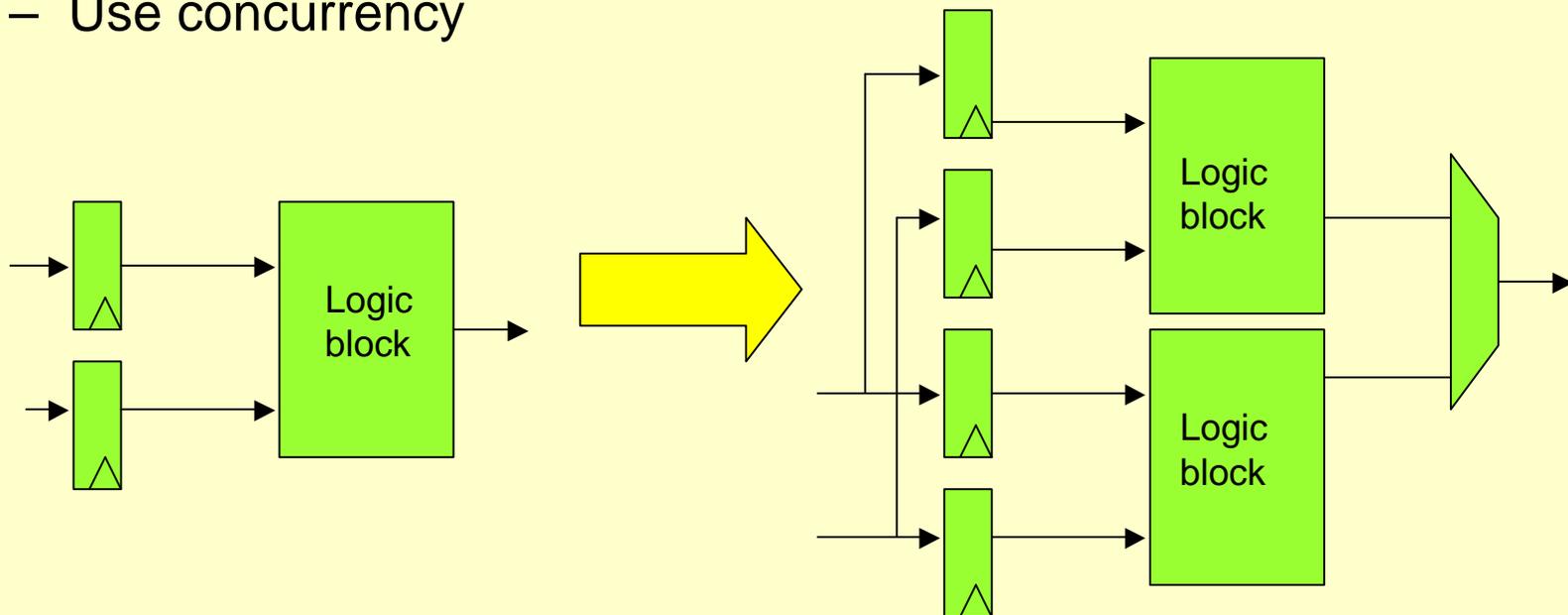
- Compute when needed: results of logic blocks are often not needed
 - Gate input or clocks to blocks

Example: parity checking in serial stream



Saving power in digital circuits (4)

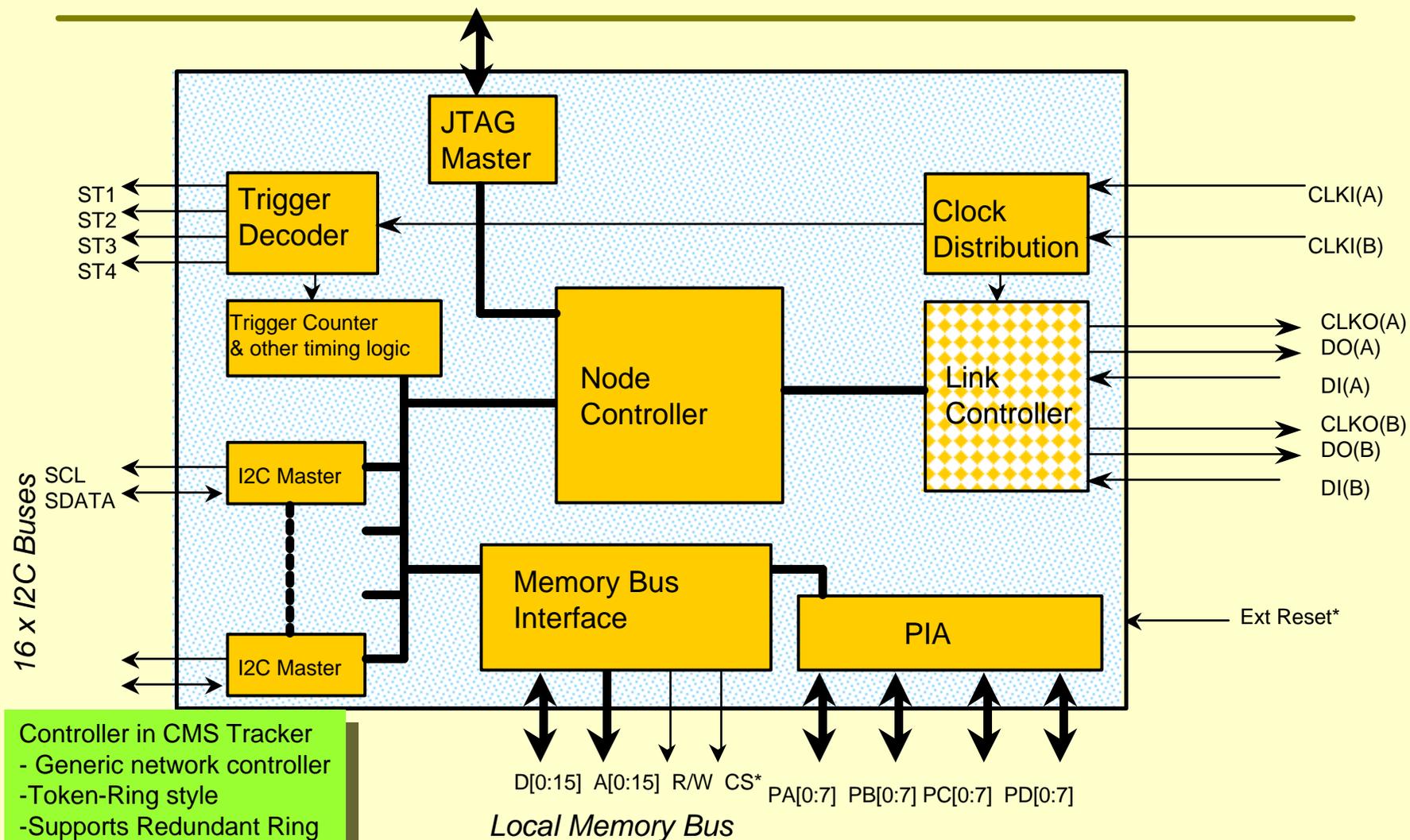
- If your circuit is not time critical ^[1]
 - Trade V_{DD} for speed
 - Use concurrency



**Lower speed,
but much lower V_{DD}^2**

[1] Almost always the case in HEP applications with modern DS technologies

Saving power in digital IC: example (1)



Controller in CMS Tracker
 - Generic network controller
 -Token-Ring style
 -Supports Redundant Ring
 ~ 60,000 gates

SEU in digital circuits requires redundancy

F. Faccio et al.
LEB98 conference
For 0.25 μm tech

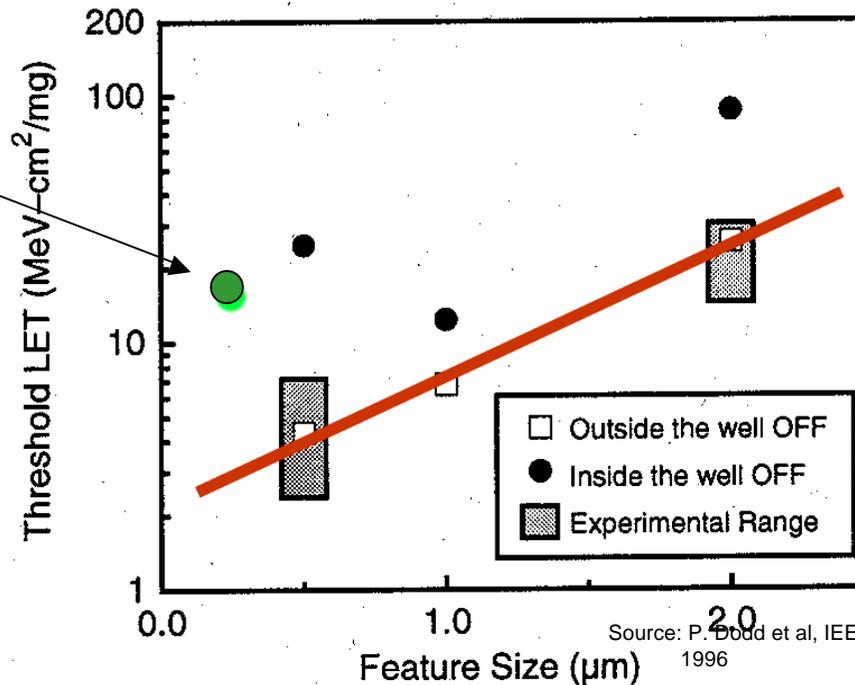


Figure 5. Experimentally-measured and simulated upset thresholds in three Sandia SRAM technology generations (no feedback resistors). The 1- and 2- μm technologies are n-substrate, 5 V; the 0.5- μm technology is p-substrate, 3.3 V.

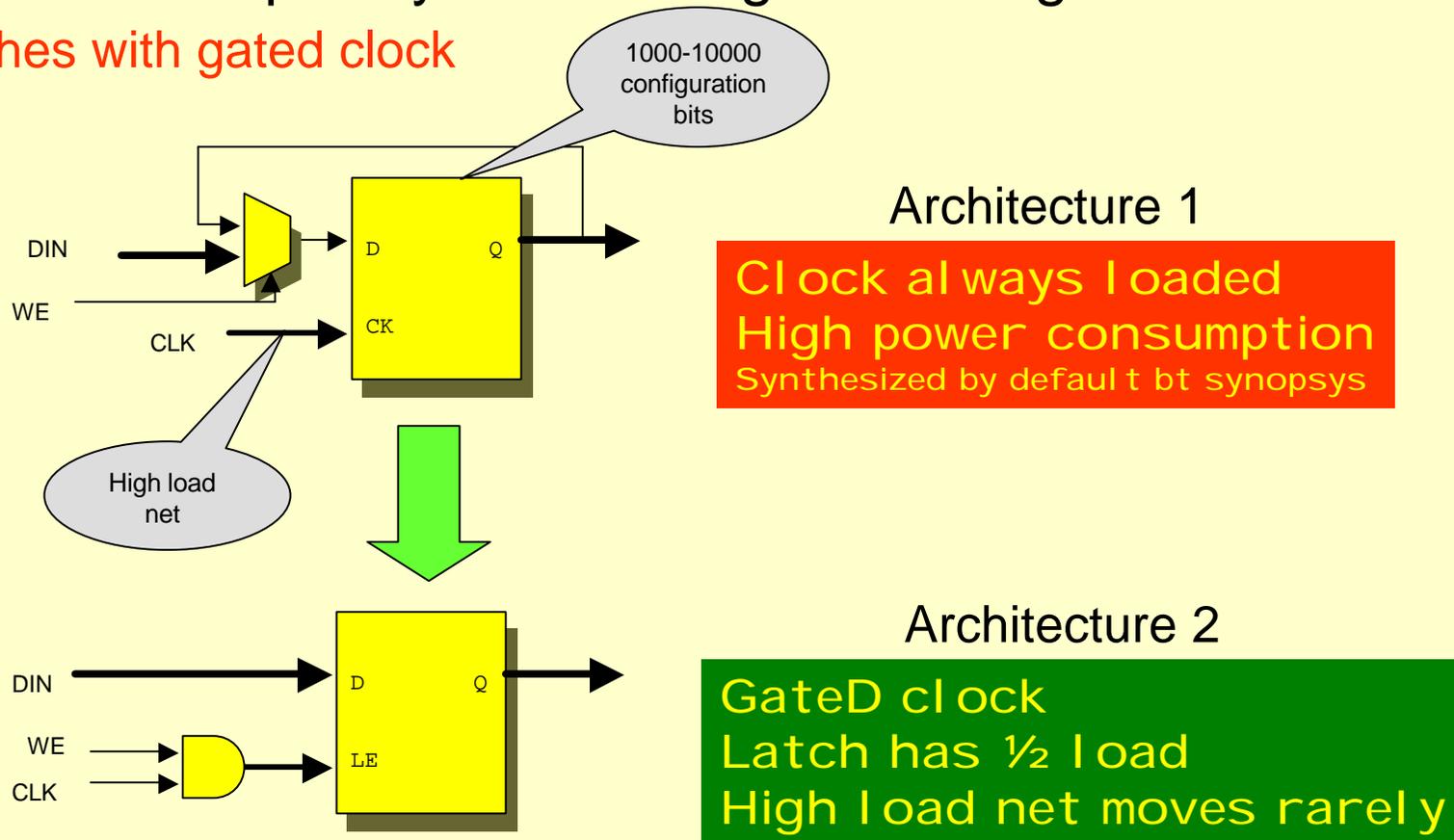
Saving power in digital IC: example(2)

- **CCU** (Controller for Timing and Slow Control Network in CMS Tracker)
 - More than 60K gates
 - Environment: CMS Tracker, 10 Mrad + 10^{14} neutrons
 - **Problem: SEU demands redundant logic**
 - Possible solution: triple everything and vote !
 - Consequence: Higher load in clock tree ? 3 x higher power
- **Better Solution:**
 - **Triple redundant logic only in critical network controller (< 3000 gates)**
 - **Parity check on data path**
 - **3 x One-Hot Finite State machines with state restore**
 - **Replace FF with latches using single phase clock**

Saving power in digital IC: example(3)

Example: un-frequently used configuration registers

Use latches with gated clock



Saving power in digital IC: example(4)

... net power saving

$$P \propto C V^2 f n M$$

C: input capacitance per inverter (20×10^{-15} F) [1]
V: supply voltage (2.5 V)
f: operating frequency (40 MHz)
n: number of Inverters in FF (6)
M: number of FF in system (5000)

Architecture 1: 150 mW
Architecture 2: < 1 mW

[1] Need of enclosed devices prevents use of minimum size transistors

Low power techniques for DSP

A 660- μ W 50-Mops 1-V DSP for a Hearing Aid Chip Set

Philippe Mosch, Gerard van Oerle, Stefan Menzl, Nicolas Rougnon-Glasson, Koen Van Nieuwenhove, and Mark Wezelenburg, *Member, IEEE*

Abstract—This article presents the flow and techniques used to design a low-power digital signal processor chip used in a hearing aid system implementing multiband compression in 20 bands, pattern recognition, adaptive filtering, and finescale noise cancellation. The pad limited 20 mm² chip contains 1.3 M transistors and operates at 2.5 MHz under 1.05-V supply voltage. Under these conditions, the DSP consumes 660 μ W and performs 50 million 22-bit operations per second, therefore achieving 0.013 mW/Mops (milliwatts per million operations), which is a factor of seven better than prior results achieved in this field. The chip has been manufactured using a 0.25- μ m 5-metal 1-poly process with normal threshold voltages. This low-power application-specific integrated circuit (ASIC) relies on an automated algorithm to silicon flow, low-voltage operation, massive clock gating, LP/LV libraries, and low-power-oriented architectural choices.

Index Terms—Algorithm optimization, digital signal processing, gated clock techniques, low power design.

I. INTRODUCTION

THE HEARING AID of which this DSP chip (Fig. 1) is a part does not replace an existing analog counterpart. It is designed from the beginning to be a high-end digital processing system with features which are only feasible with this specific technology. Examples of these features are multiband compression, pattern recognition, adaptive filtering, and noise cancellation.

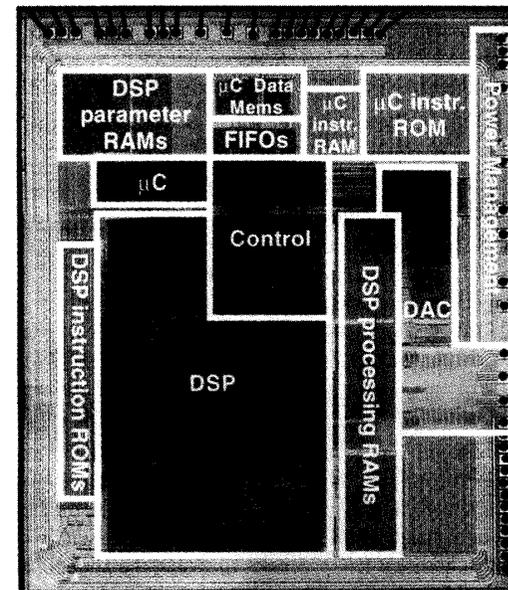


Fig. 1. Chip micrograph (parts of the design are hidden by metal filling structures).

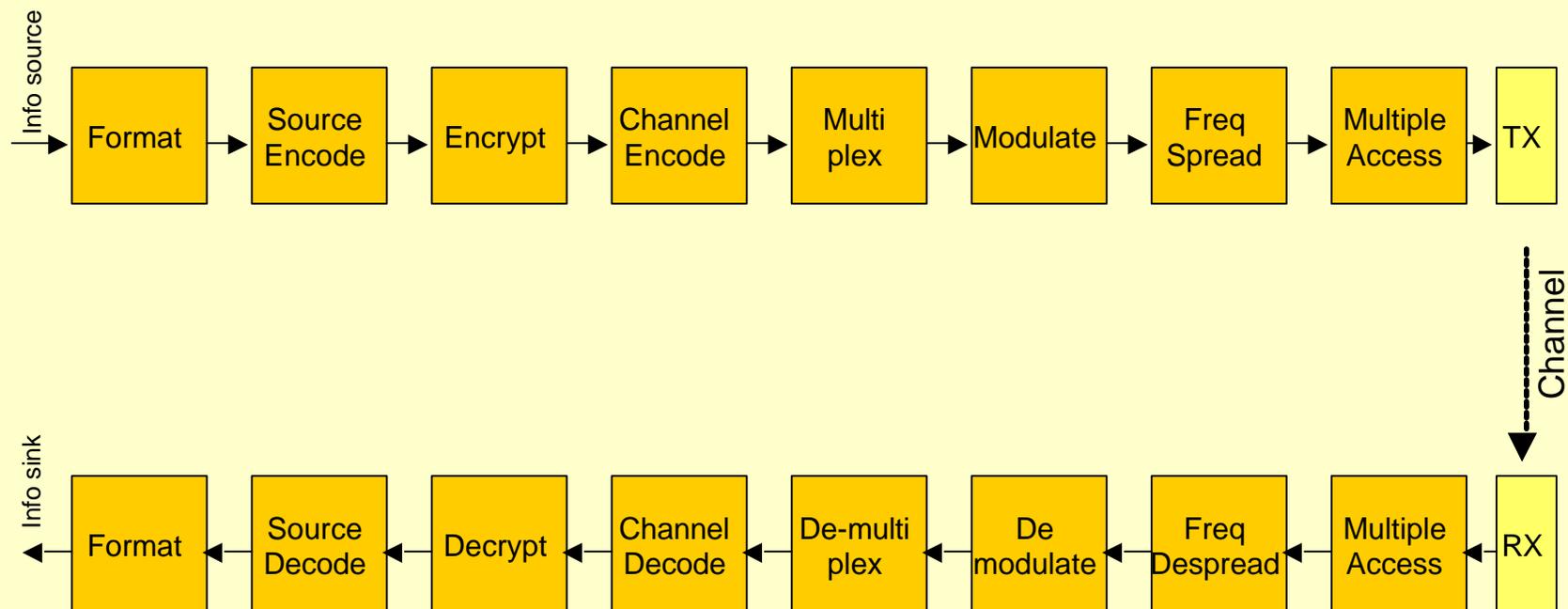
The products based on this chipset must fit in the ear. Volume

A structured approach to
low power design
for HEP applications

... As with most media from which things are built, whether the thing is a bacterium, a sonnet, a fugue or a word processor, architecture dominates material ...

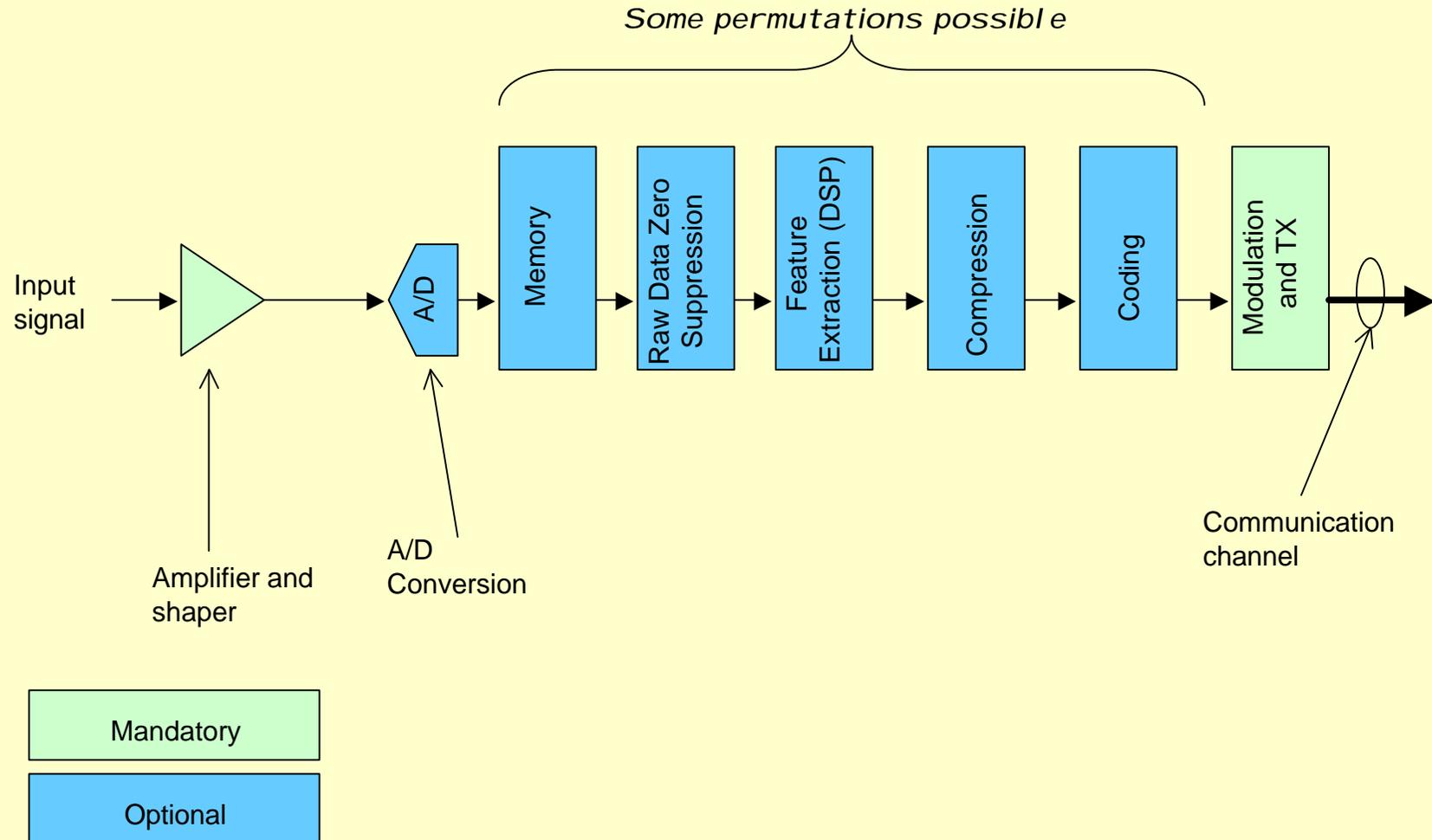
Alan Kay

Generic digital communication system

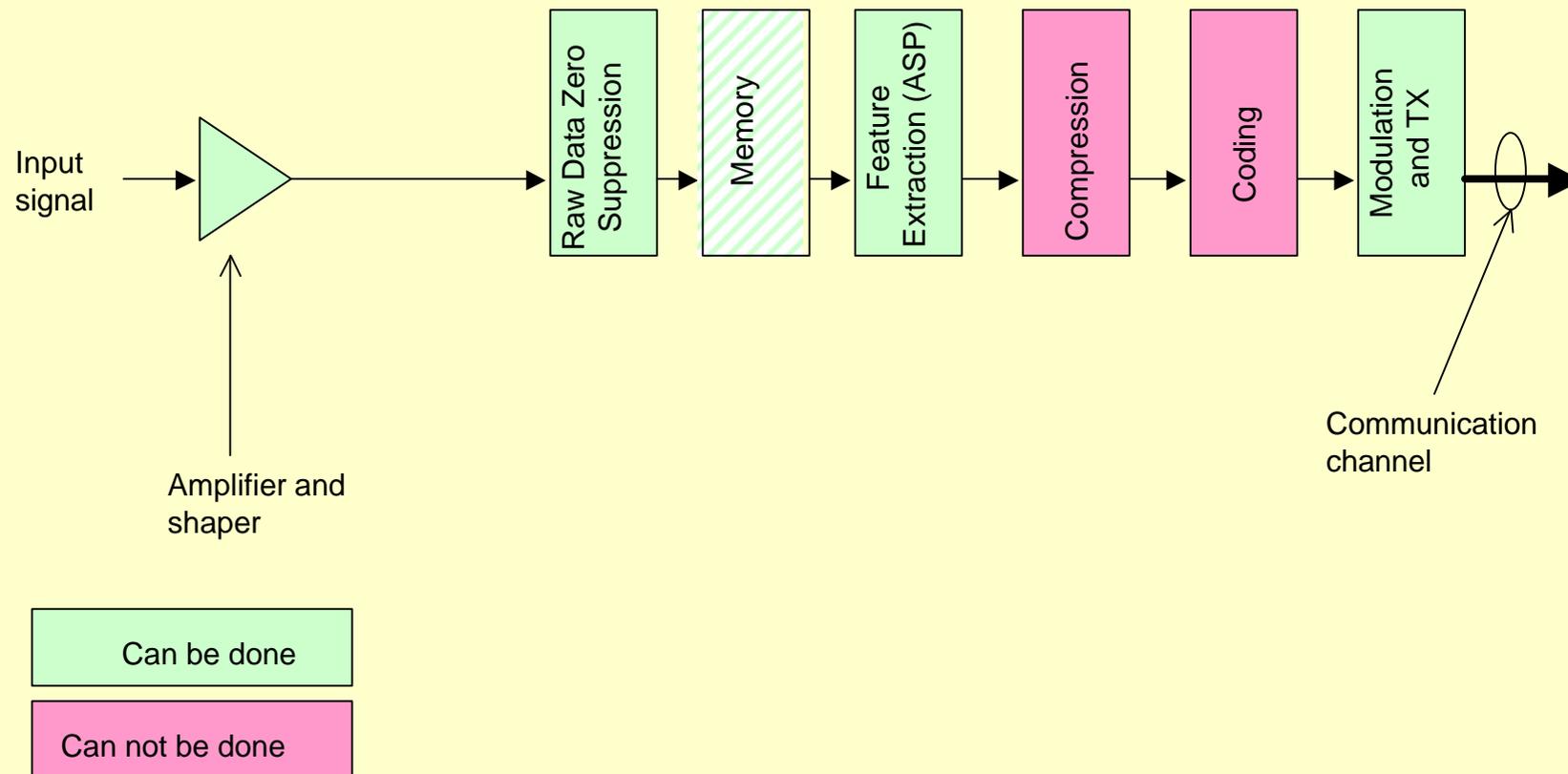


Picture from B. Sklar: "A Structured Overview of Digital Communications", IEEE Communication Mag., August 1983

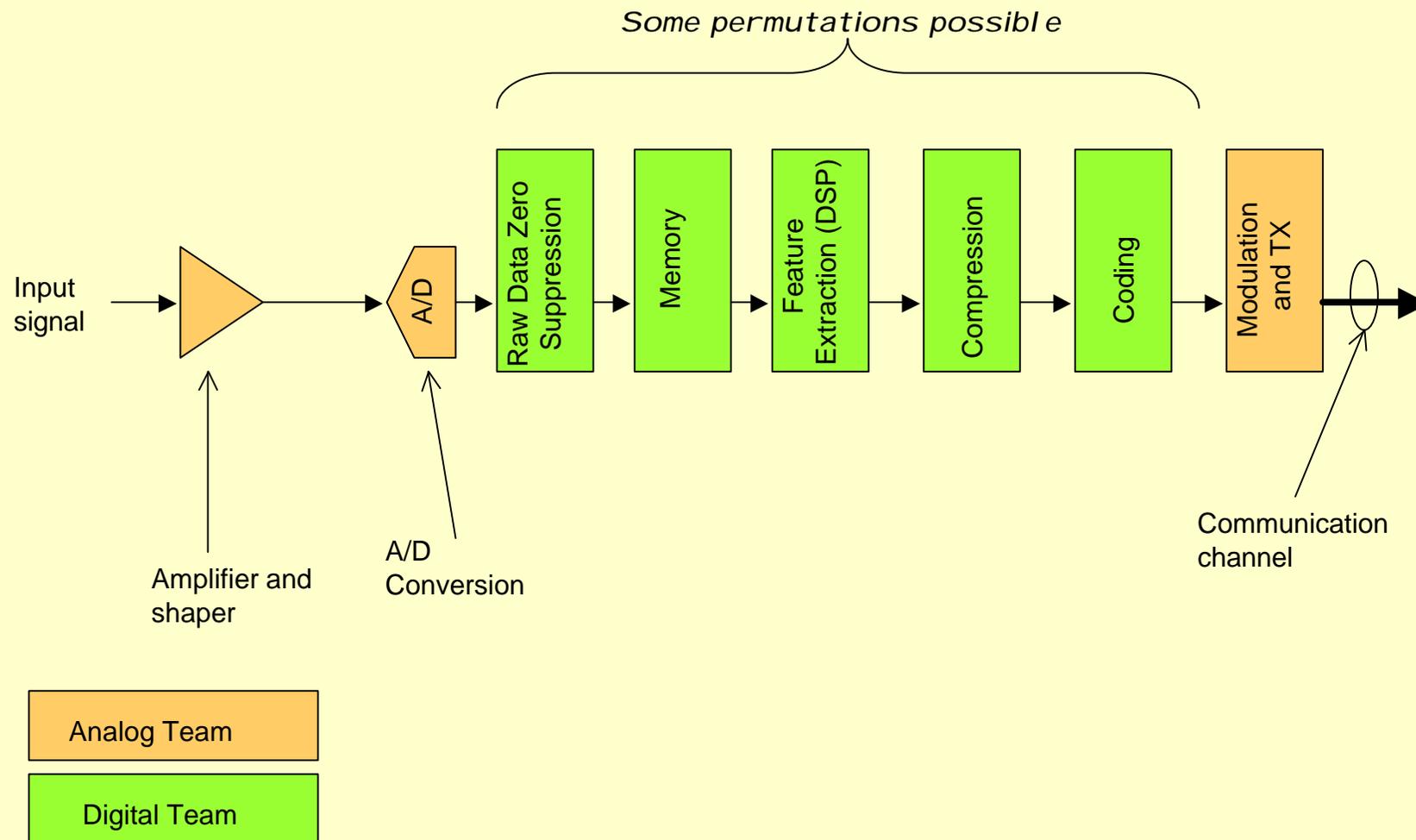
The “text-book” HEP channel architecture



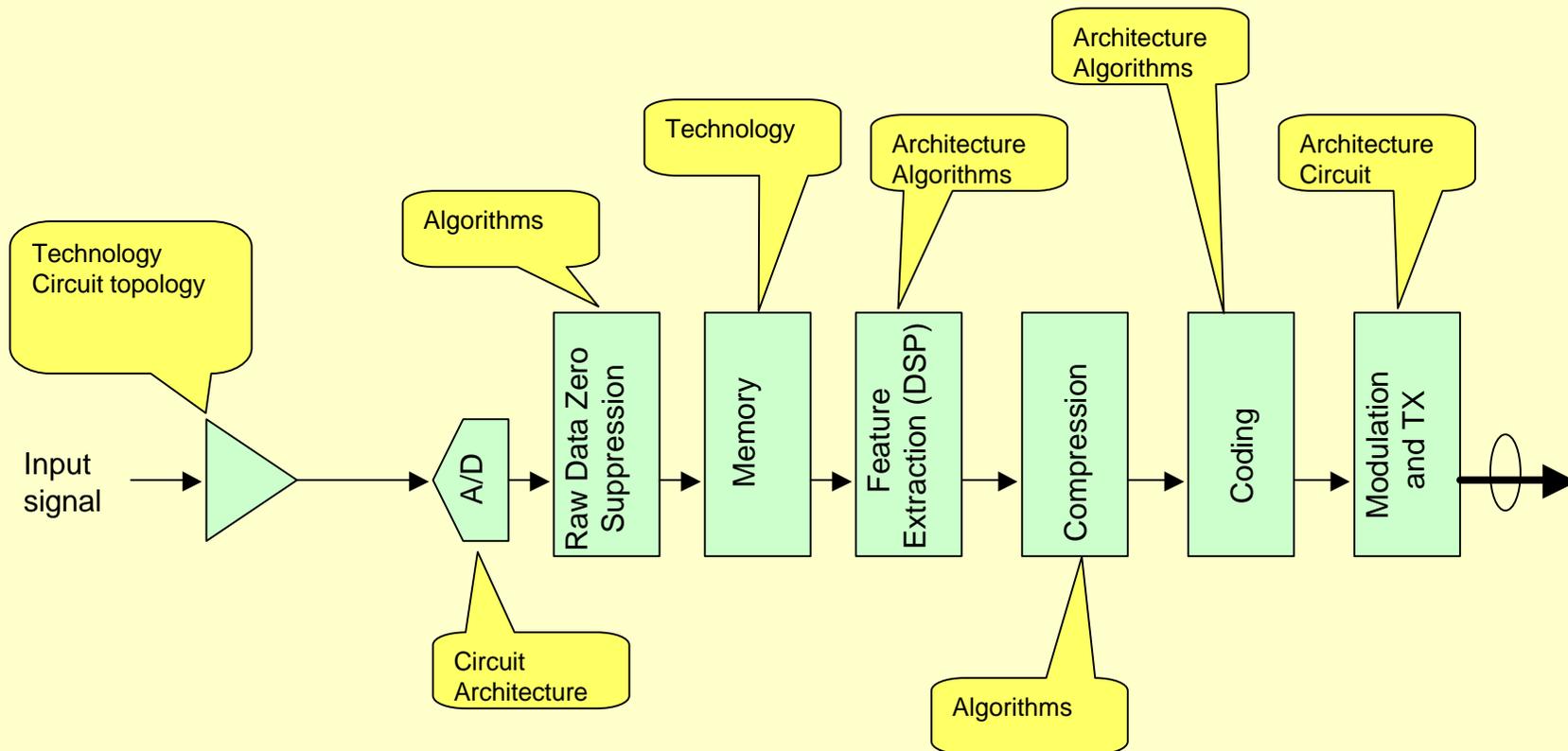
The “all-analog” option



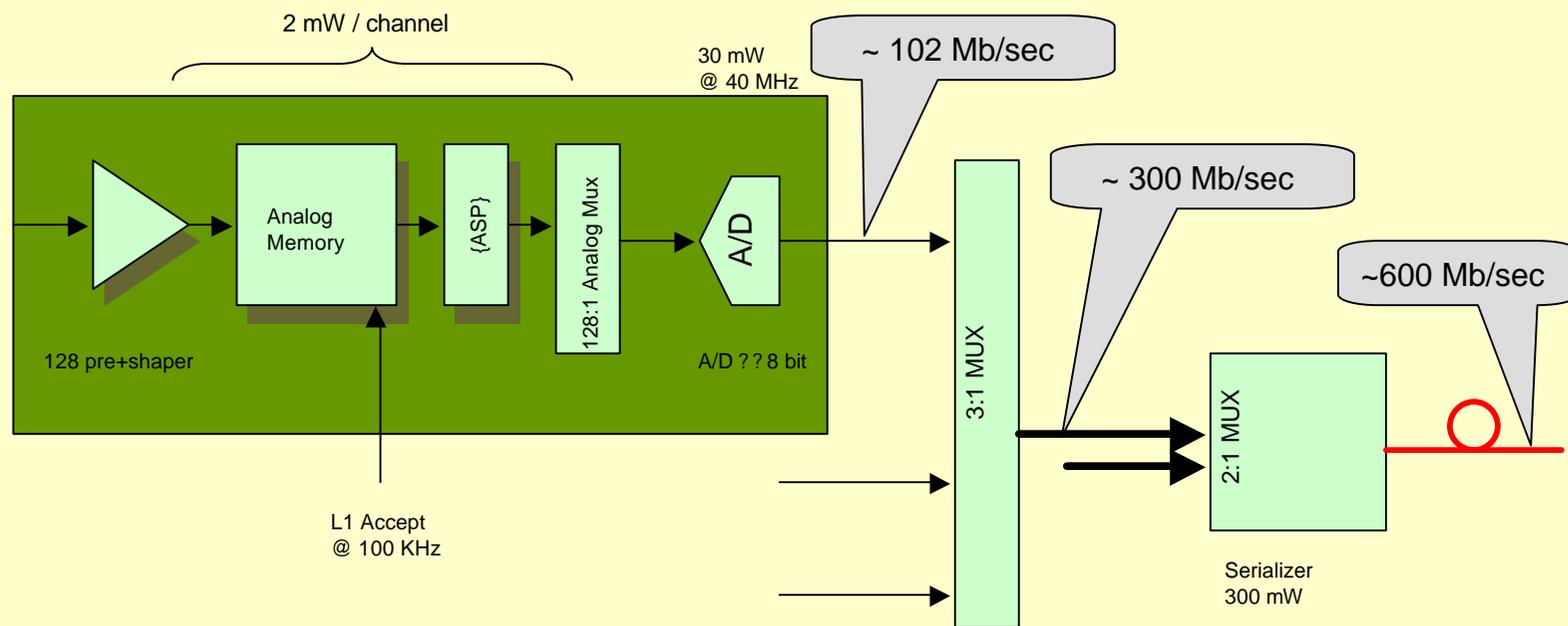
The “all-digital” option



Save power using...



A (not so far into the) future FE chip for tracker

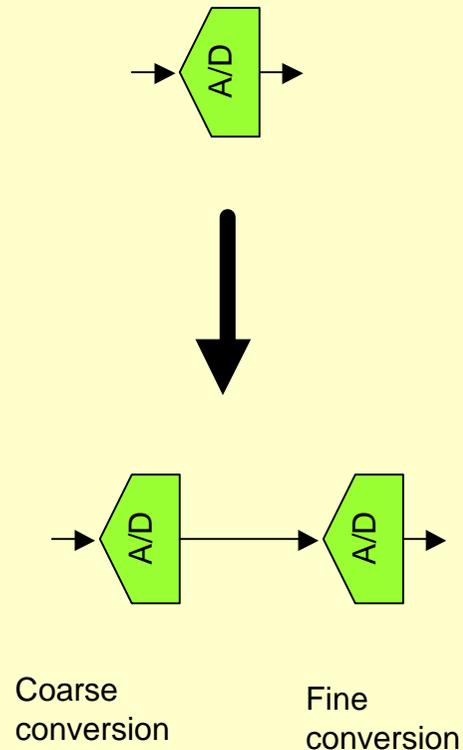


Design parameters:
 100 KHz trigger rate
 occupancy 2%
 avg. # of strips hit: 3

Total power:
 $128 * 6 * 2.0 \text{ mW} = 1536 \text{ mW}$
 $6 * 30 \text{ mW} = 180 \text{ mW}$
 $1 * 300 \text{ mW} = 300 \text{ mW}$
 Total = ~ 2.6 mW /ch

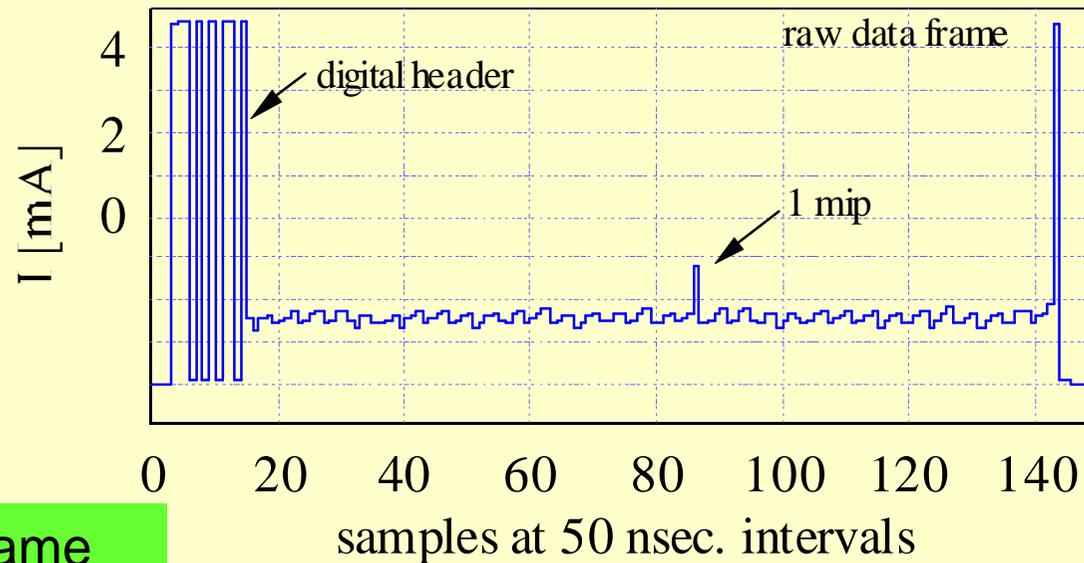
Specific A/D development

- Requirements in HEP are no different from those in industry, but:
 - *Dynamic range is large but signals change little most of the time (actually precision requirement does not change, only dynamic range does)*



Fundamental observation

Detector signals change little
most of the times



Data frame
from Tracker
APV read-out

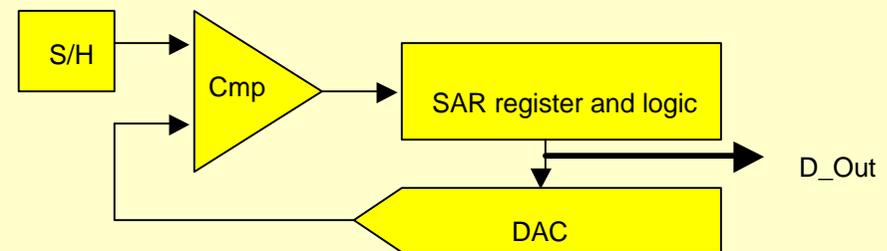
Successive Approximation Converter

Standard Algorithm (full range, 10 bit)

$$V_{DA} = V_{ref}/2$$

$$M = 10$$

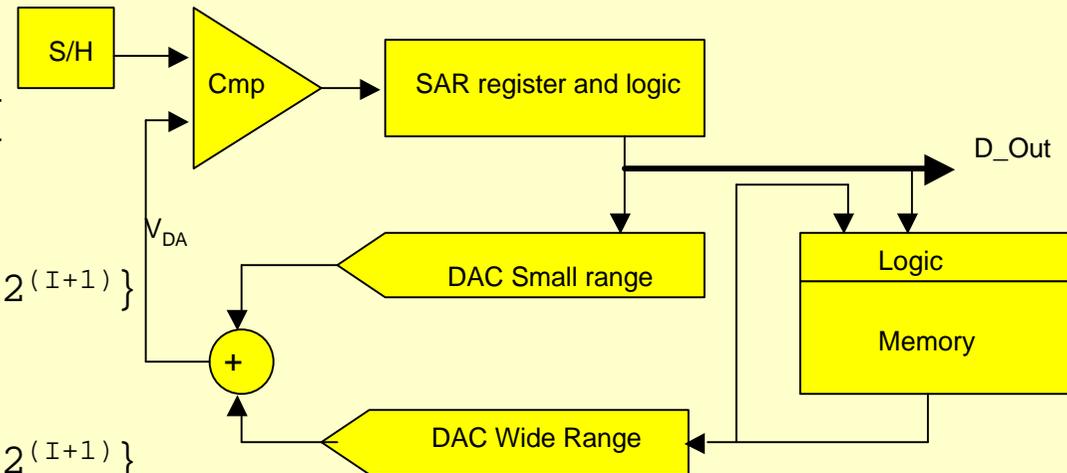
```
for (I = 0; I < M; I++){  
  if( $V_{in} > V_{DA}$ ) then{  
     $b_i = 1$   
     $V_{DA} = V_{DA} + V_{ref}/2^{(I+1)}$   
  }  
  Else{  
     $b_i = 0$   
     $V_{DA} = V_{DA} - V_{ref}/2^{(I+1)}$   
  }  
  result = result << 1 +  $b_i$   
}
```



Modified SA Converter

Low Power Algorithm (3 bit search)

```
VDAw = VPrevious  
VDAs = Vref / 2  
M = 10; N = 3;  
for (I = 0; I < N; I++){  
  if(Vin > VDA) then{  
    bi = 1  
    VDAs = VDAs + Vref/2(I+1)}  
  Else{  
    bi = 0  
    VDAs = VDAs - Vref/2(I+1)}  
  result = result << 1 + bi  
}  
if (... check out of range ...)  
result = resultprevious + tresult
```



Crazy ?



US PATENT & TRADEMARK OFFICE PATENT FULL TEXT AND IMAGE DATABASE

[Home](#) [Boolean](#) [Manual](#) [Number](#) [Help](#)
[Hit List](#) [Prev](#) [Next](#) [Bottom](#)
[View Shopping Cart](#) [Add to Shopping Cart](#)
[Images](#)

(9 of 10)

United States Patent
Fernald

5,543,795
August 6, 1996

Hybrid analog-to-digital convertor for low power applications, such as use in an implantable medical device

Abstract

A method and apparatus is disclosed for an analog-to-digital converter (ADC) to minimize power consumption. The ADC of the present invention minimizes the number of clock cycles required to determine the correct digital code for a particular sample point on an electrogram signal, thus making it possible to turn off some or all of the ADC logic during idle periods. The ADC includes prediction logic that provides a starting point for subsequent digital code representations of the electrogram signal. The prediction logic receives recent code conversions values to predict a current digital code value. This predicted digital code is converted to an analog value and compared with the actual electrogram signal. Next, the ADC adds (or subtracts) a constant value (C) to (or from) the predicted code and compares the result to the electrogram signal. If the ADC determines that the predicted value is within the constant value (C) of the correct digital code, then the ADC counts in the proper direction until the comparator changes output state. If the ADC determines that the predicted value is not within the constant value (C), then the successive approximation logic is enabled and used to find the correct code.

Inventors: **Fernald; Kenneth W.** (Lake Jackson, TX)

Assignee: **Intermedics, Inc.** (Angleton, TX)

Appl. No.: **460139**

Filed: **June 2, 1995**

Current U.S. Class:

341/163; 341/158; 341/162

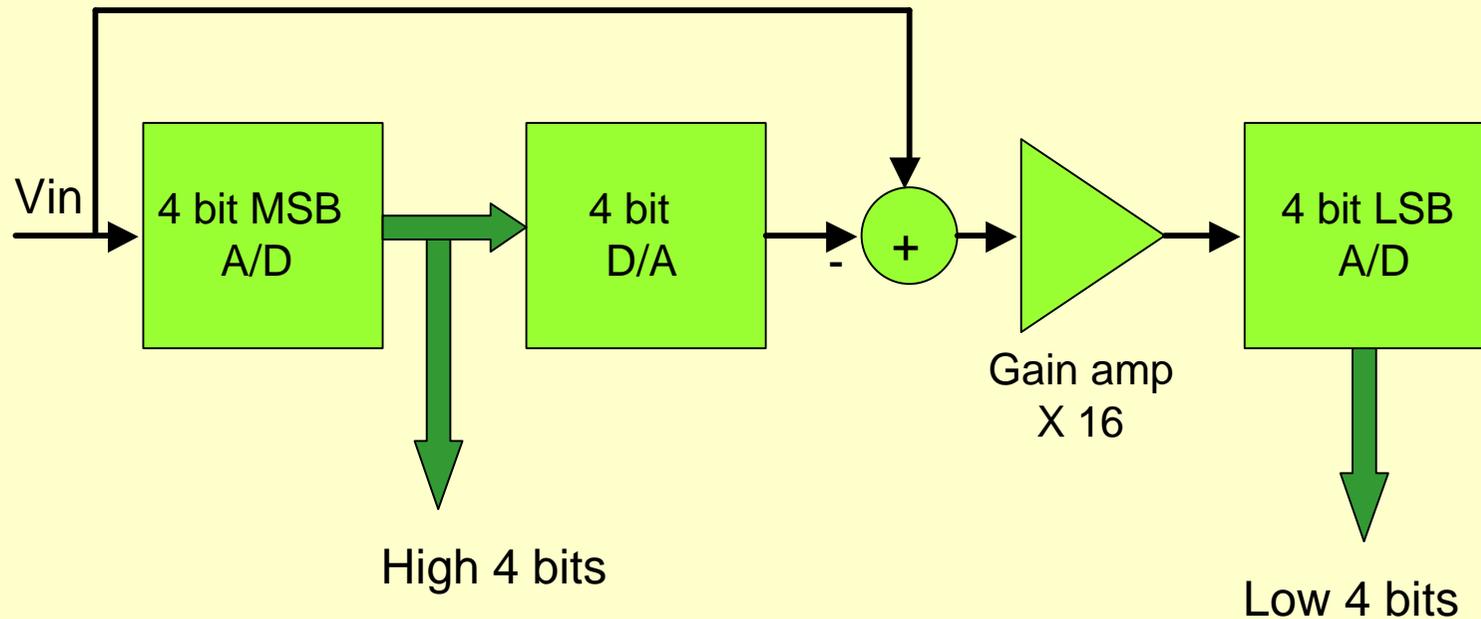
Intern'l Class:

H03M 001/40

Field of Search:

341/164,165,163,141

Power saving sub-ranging A/D



Low Power Algorithm

```
If  $V_{in}$  close_to  $V_{previous}$  then  
  Convert using low A/D  
Else  
  Convert using full A/D
```

Consequences...

- For trackers:
 - *Virtually none, as sometimes there is already an analog memory holding the data*
- For Calorimeters:
 - *Need a more complex Sample/Hold structure (buffer memory) to compensate for the longer conversion time occasionally needed*

Saving power: Data TX Architecture

- Data transmission type:
 - Analog vs. digital
 - Industry is massively digital ⁽¹⁾:
 - datacom, telecom, radio broadcasting, TV, audio (HiFi)
- Coding scheme currently in use in HEP:
 - Analog (sampled data)
 - Amplitude modulation
 - Digital
 - Straight data with timid attempts to use compression



⁽¹⁾This refers to the encoding scheme, but signals are always “analog”, advantage: regeneration, errors are not propagated

Compression: dream or reality ?

- Fact: works very well in some applications
 - Computer data:
 - Text: loss-less
 - Images: lossless and lossy
 - sound (music and voice): lossy
- Fact: interesting data from detectors are “rare” (i.e. occupancy is often low, few %)
- Compression makes data-flow asynchronous
- Random data can't be compressed well ...
 - Need good knowledge about baseline and noise

Compression: how does it work ?

- Most techniques adaptable to HEP would be variations of the Huffman coding algorithm
- Huffman coding (and other loss-less algorithms) works by ordering input data according to frequency, and then coding most frequent characters into shorter sequences of bits

Huffman code example

Alphabet to be transmitted:

a1, a2, a3, a4, a5
000, 001, 010, 011, 100

3 bits necessary in normal binary coding

Actual frequencies in data block

: f(a1) = 0.2
f(a2) = 0.4
f(a3) = 0.2
f(a4) = 0.1
f(a5) = 0.1

Codeword assigned (sorted by length):

h(a2) = 1
h(a1) = 01
h(a3) = 000
h(a4) = 0010
h(a5) = 0011

Average code-length: $0.4 \times 1 + 0.2 \times 2 + 0.2 \times 3 + 0.1 \times 4 + 0.1 \times 4 = \mathbf{2.2 \text{ bits/symbol}}$

While the **entropy** (i.e. the best theoretical compression) would have been: **2.12 bits/symbol**,

Entropy defined as: $H = - \sum_i P_i \log_2 (1 / P_i)$

Compression

Alice Silicon Drift
(preliminary data)

<i>Type</i>	<i>Coding</i>	<i>Compression factor</i>
<i>Fixed table</i>	Huffmann + RLE	2-4

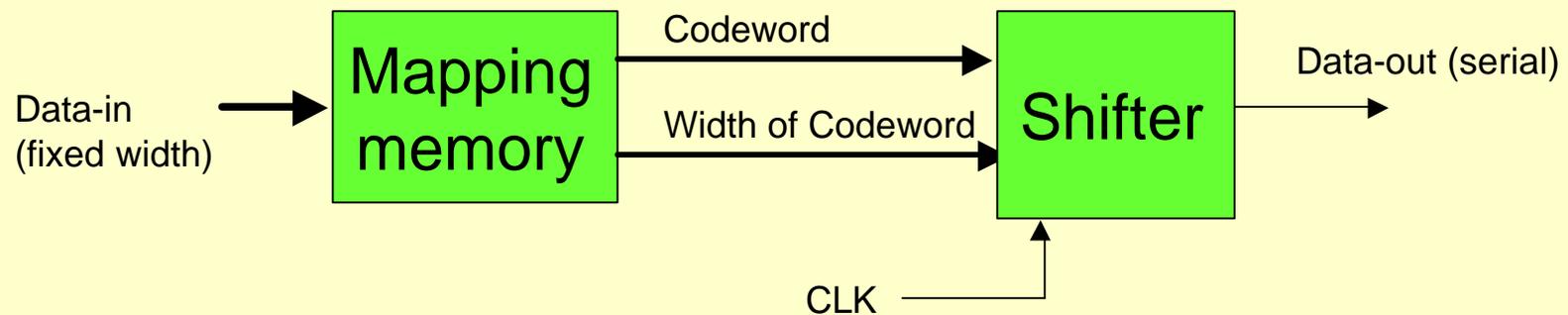
CMS Ecal compression

<i>Type</i>	<i>Coding</i>	<i>Compression factor</i>
<i>Fixed table</i>	Huffman	4.2
<i>Variable table</i>	Huffman	3.1
<i>compress</i>	dictionary	3.6
<i>gzip</i>	dictionary	3.3
<i>dynamic</i>	8/16 bits	2.0
<i>ALDC</i>	dictionary	2.6
<i>DCLZ</i>	dictionary	3.3

Compression factors for CMS ecal calorimeter data.
From J. Badier Et al, CMS Note 1999/068

Huffman coding in hardware

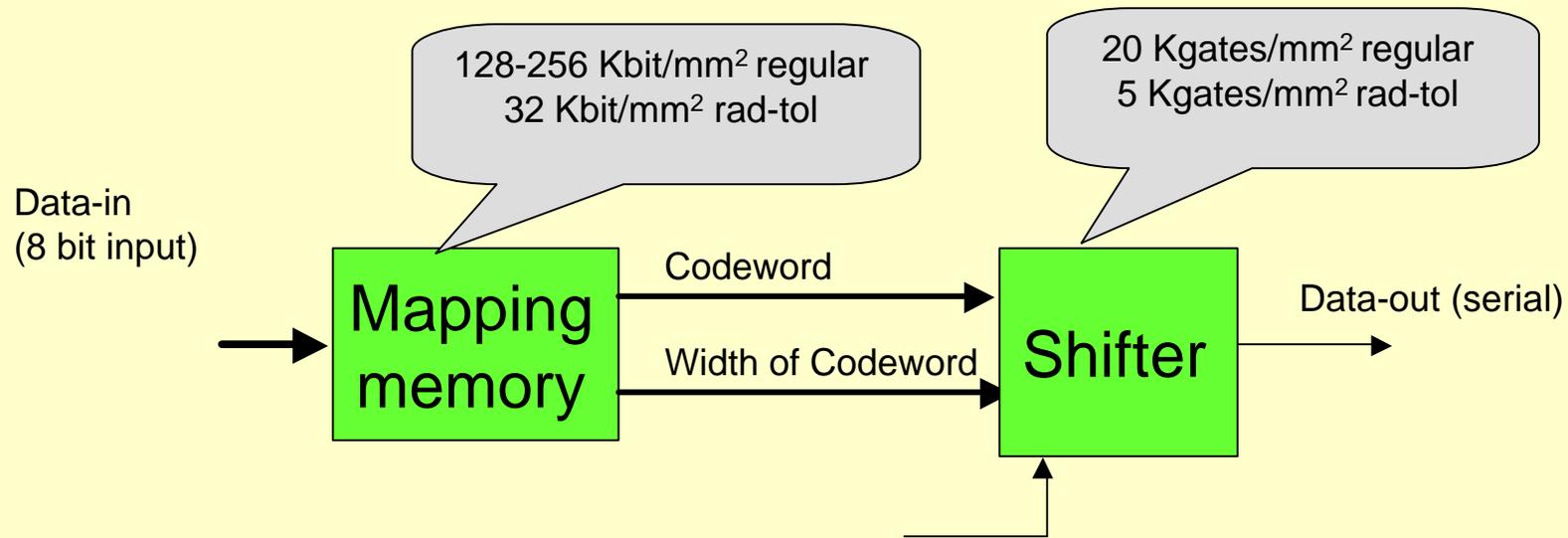
- Need only mapping memory + shifter
- Easy to implement in VLSI
- Can be programmed per channel



Errors on links can be disastrous

Huffman coding in hardware (2)

Example for 0.25 μ m technology



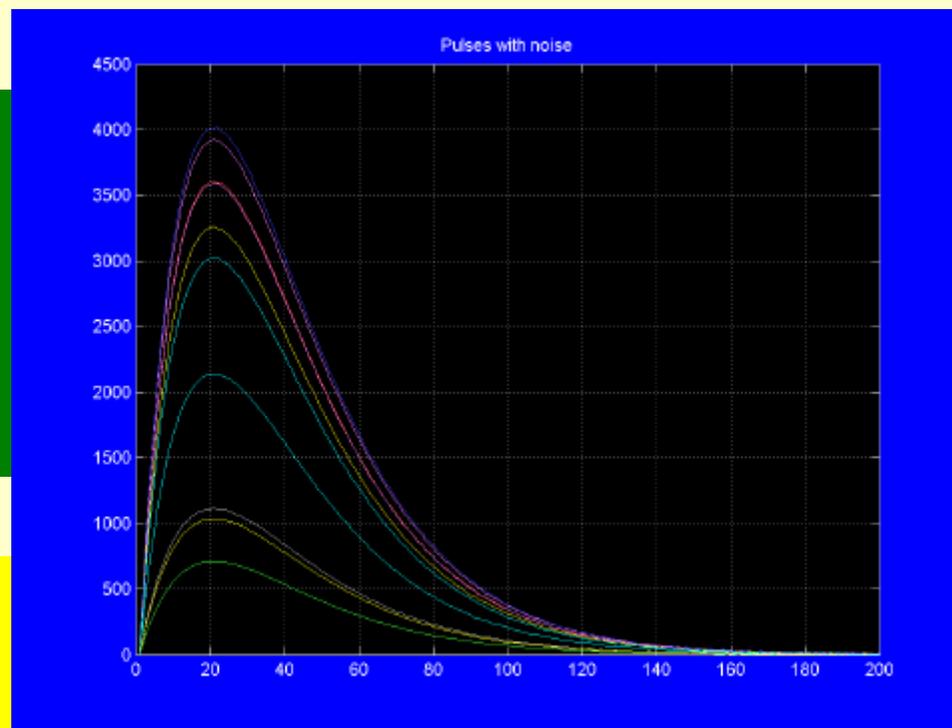
Total area: well below 1 mm²

Compressing calorimeter data

- Observation

Signal Amplitude
Has little
regularity

... but Signal
SHAPE
Does not change
much



Typical pulse shape for CR-RC filter

An efficient compression algorithm for calorimeters

1. Normalize pulse to max reference pulse at peak

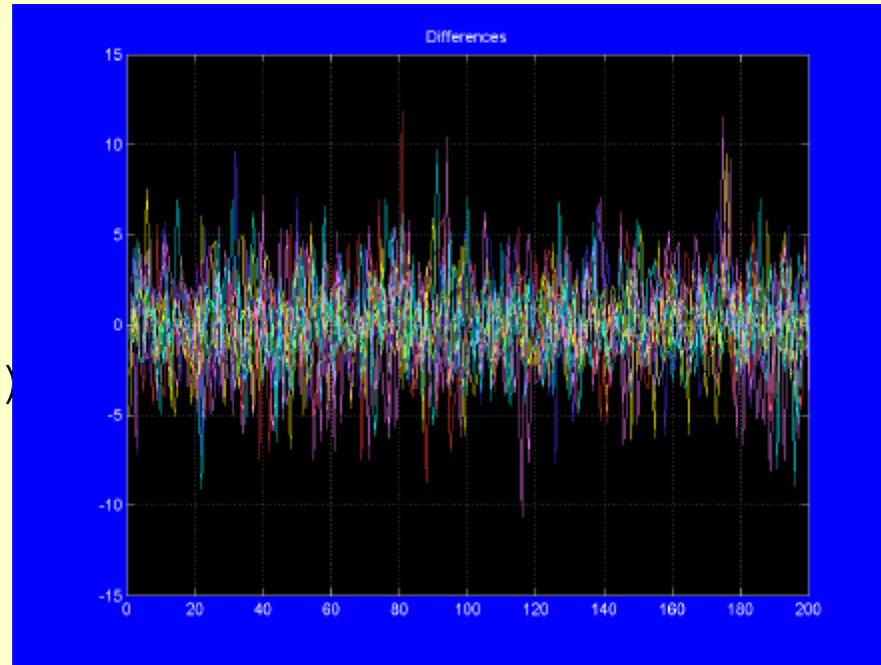
$$p(i_{\text{peak}}) = s * d(i_{\text{peak}})$$
$$i = 1..5$$

2. Compute differences:

$$?p(i) = r(i) - s * d(i)$$

between actual data and reference pulse

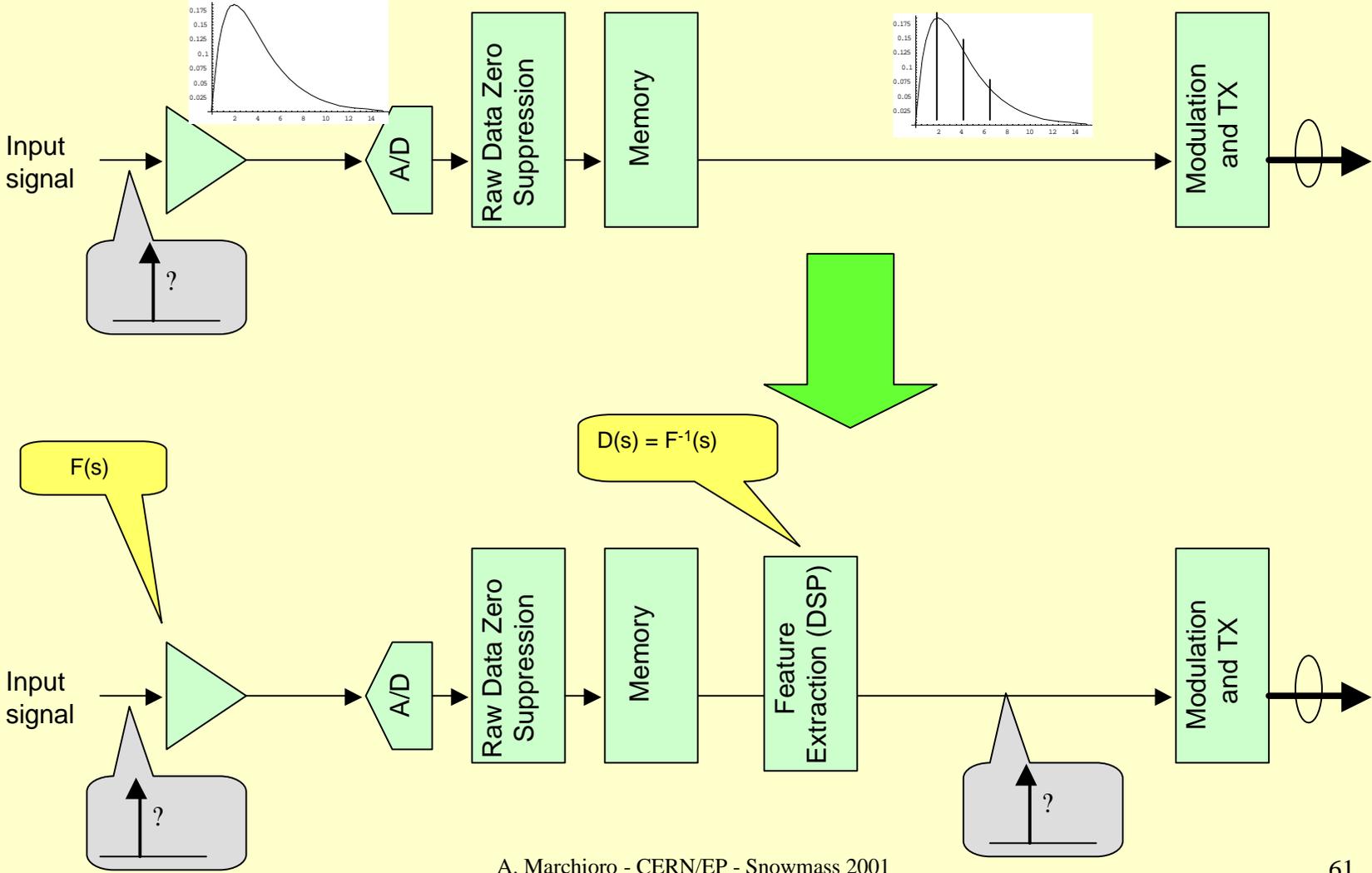
3. Apply Huffman compression to $?p(i)$
4. Transmit only scaling factor s and $?p(i)$



Example:

- ▶ 12 bit A/D
 - ▶ Calorimeter $\sigma = 5\% * E^{1/2}$
 - ▶ 5 samples
- > entropy of differences:
~2.5 bits

The best compression scheme: feature extraction



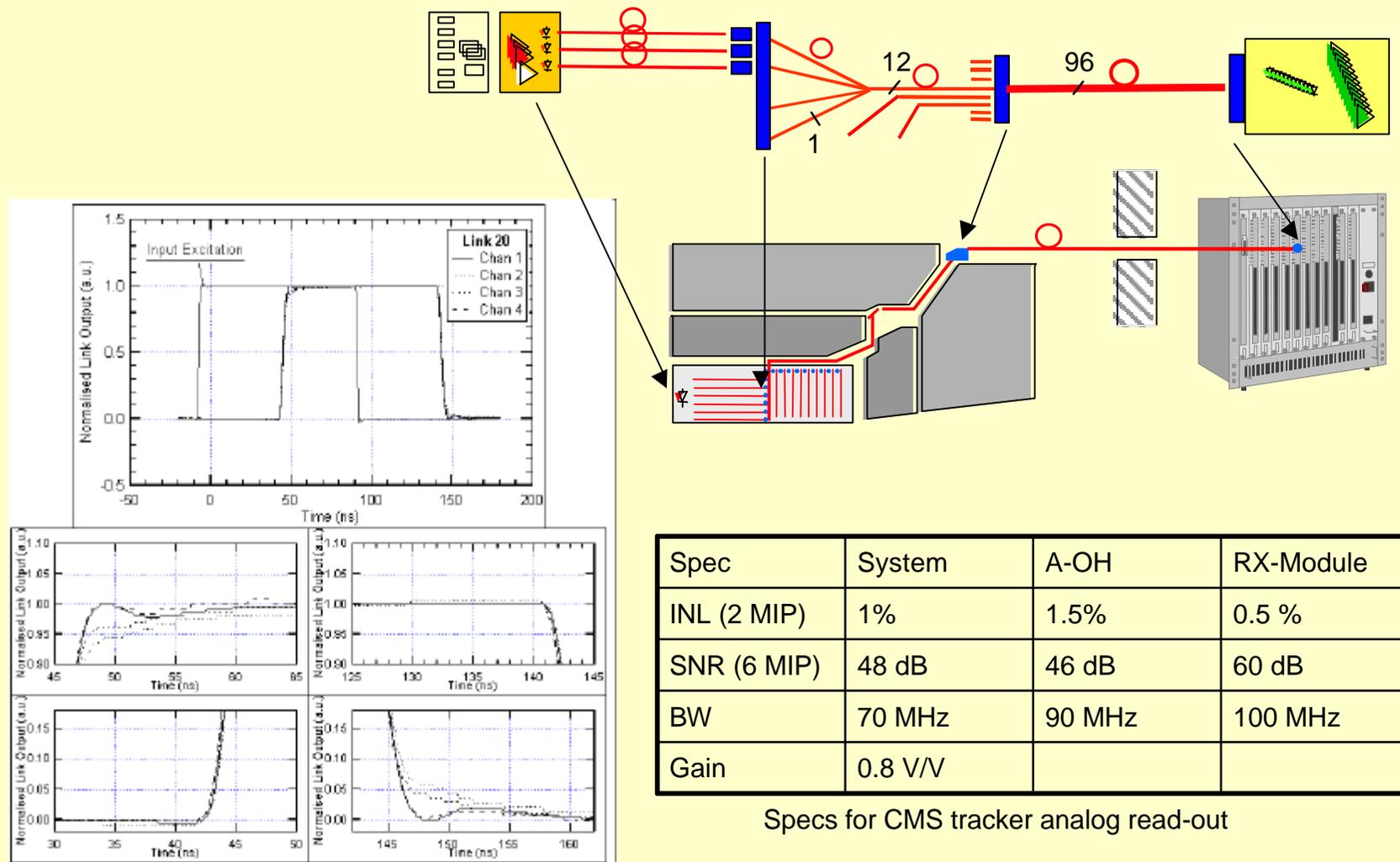
Feature extraction

- Just read the number you really want !
- Why not done more often:
 - Detectors are poorly characterized at start-up
 - Noise, systematic effects etc.
 - Algorithm is almost never fully clear during construction
 - People are “afraid” of losing data
 - Can’t be un-done
- Need high level of programmability
 - Efficient filters are often hardwired
- If money for high speed (quality) links is already spent at the beginning, motivation is low

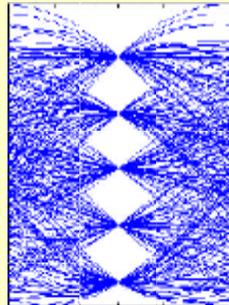
Modulation schemes

- Analog: why are we using only pulse height modulation?
 - Hard to achieve good S/N
 - Look at alternate coding schemes:
 - Frequency or other modulation
 - Very robust to noise and gain variations
 - May not need A/D conversion
 - May be easier to calibrate
- Digital:
 - Easy to multiplex
 - Easy to store/buffer
 - Easy to compress

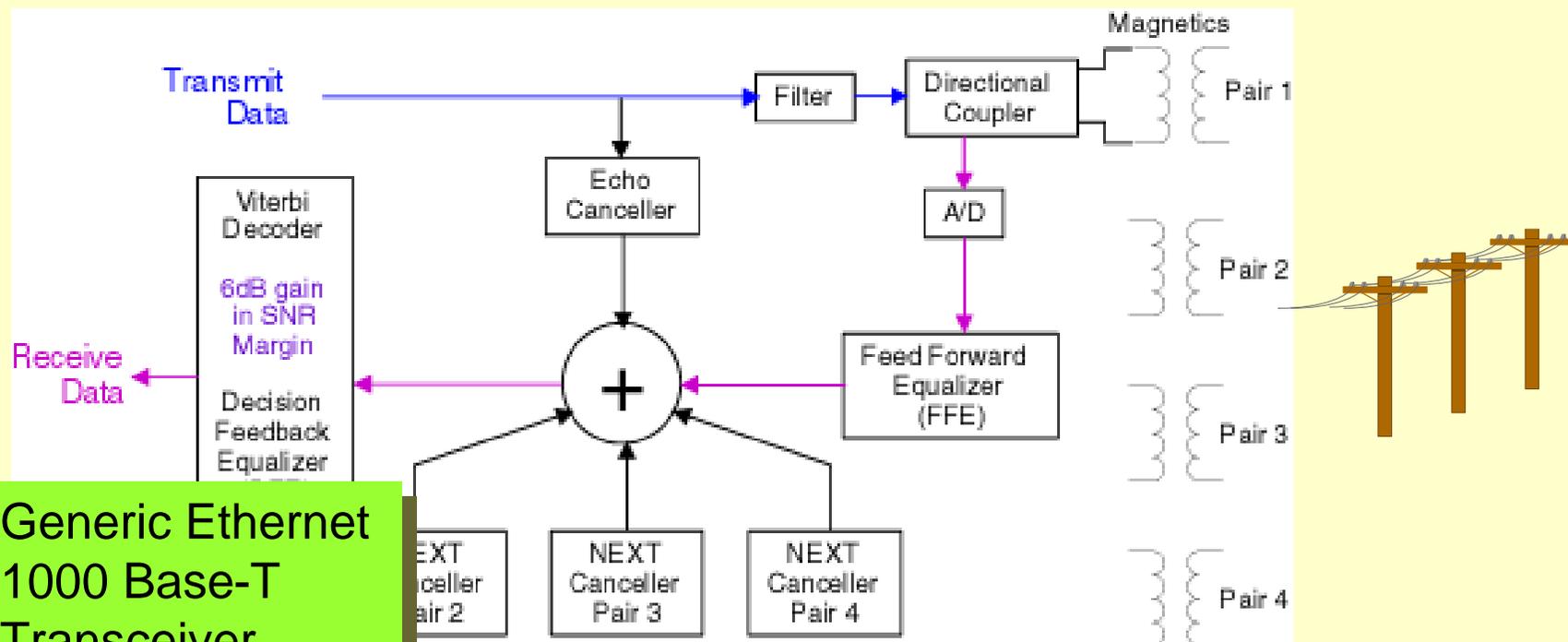
Transmitting data ... our way



Transmitting data ... another way



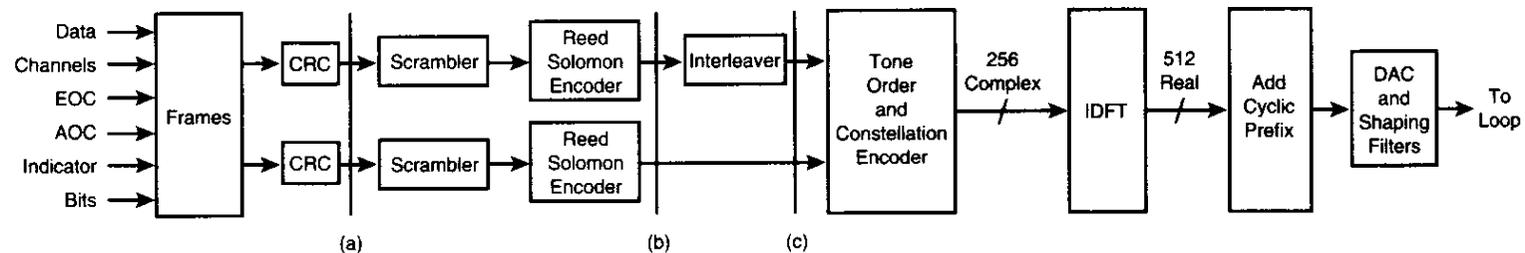
PAM-5 Modulation scheme



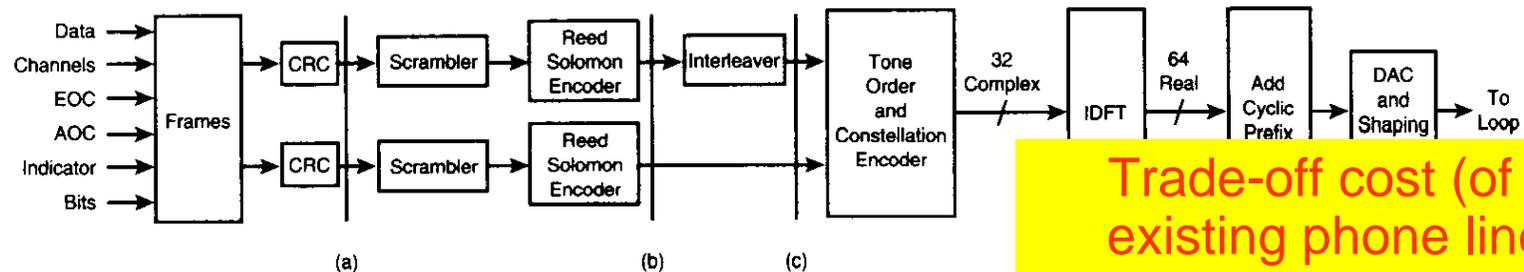
Generic Ethernet
1000 Base-T
Transceiver

ADSL: 1 Mb/sec on AWG24 wires

Block diagram for an upstream ADSL transmitter.

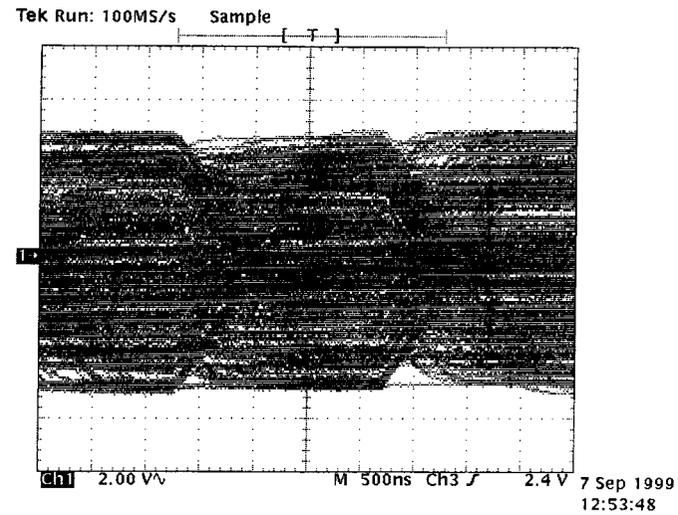
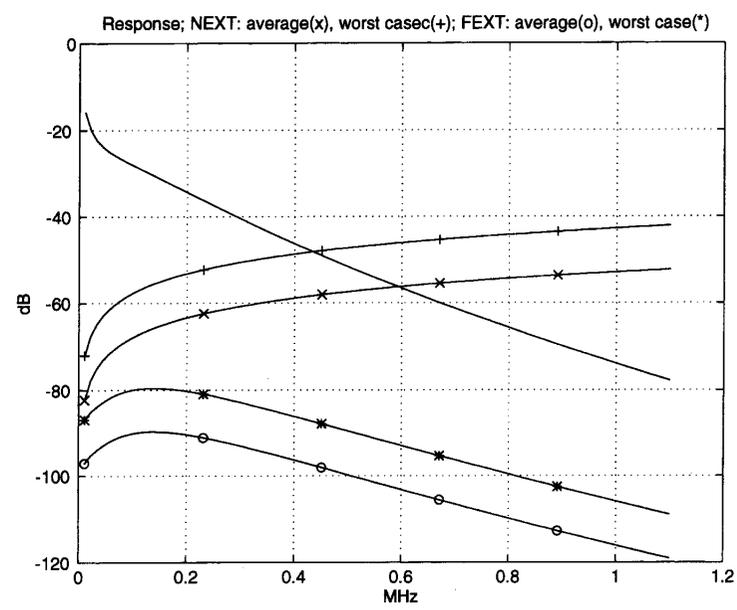


Block diagram for a downstream ADSL transmitter.



Trade-off cost (of existing phone lines) vs. complexity (of new VLSI circuits)

ADSL signals (!)



Future outlook

- The first generation of chips for HEP started with a bottom-up approach
 - Took the requirements from FE detector signals and built a system up
 - Some systems where built around chips
- Second generation will have to exploit global optimization
 - e.g.: demand on power supplies to be delivered to FE has to be factorized in cost of system
 - Design for low-power is a must
 - Chips must be designed on system specs

Conclusions

- VLSI has plenty of potential to improve instrumentation for HEP
- ... and perhaps the “only” key to another generation of experiments, if power reduction can be mastered
- Still, it does not come for free:
 - Complexity of large systems is difficult to manage
 - Projects must start from global perspective, top-down system design, and not from “FE preamp-shaper” specs
 - Our community has to learn (= invest + try) how to exploit techniques developed in the data-communication world to save power and reduce cost of systems