

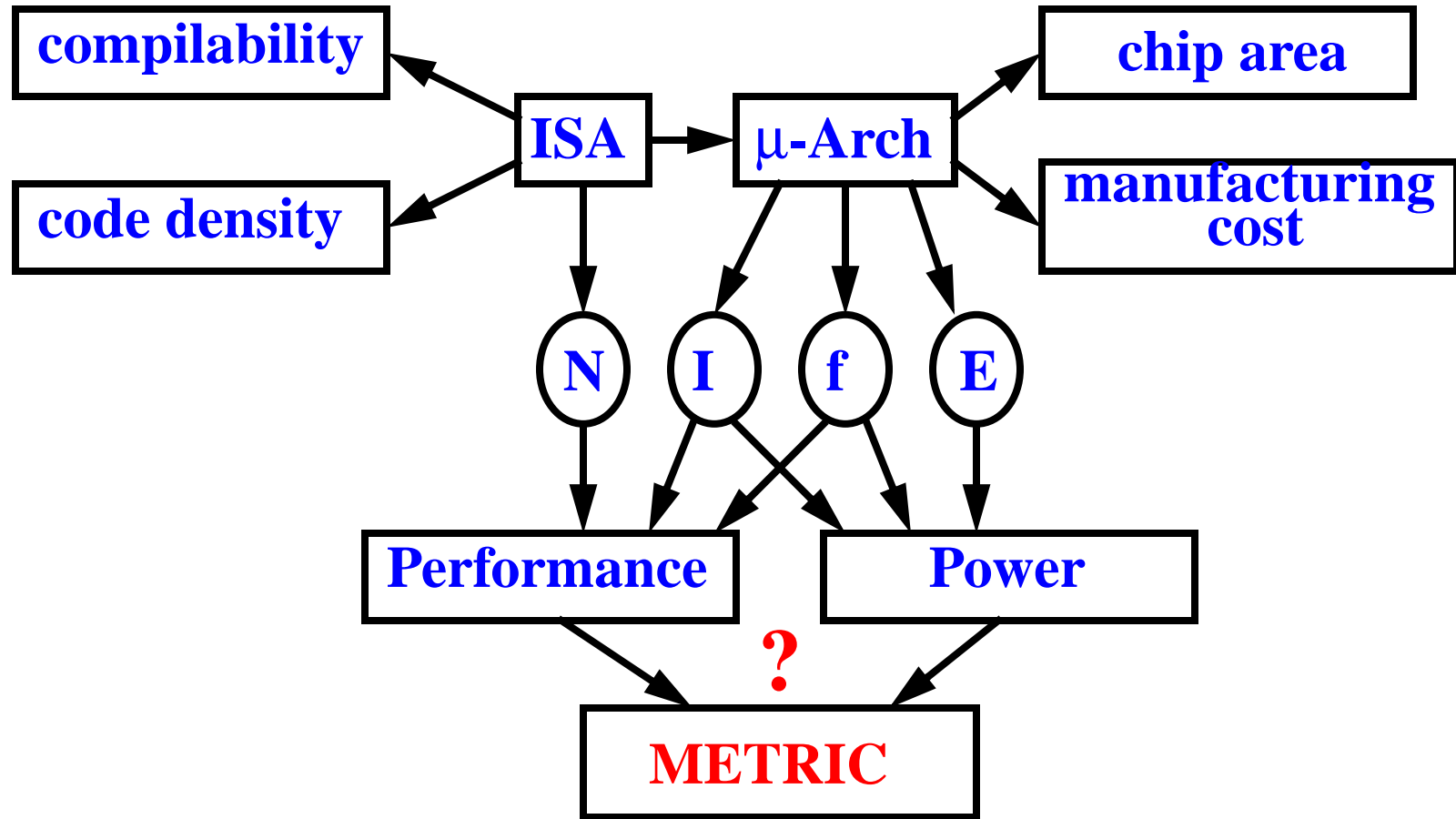
# Unified Architecture Level Energy-Efficiency Metric

**Victor Zyuban**

## Motivation

- Considering power at early stages of microprocessor development presents more opportunities for power-performance tradeoffs
- Architectural changes affect most of processor characteristics
- Architectural energy-efficiency metric is needed to move from qualitative to quantitative analysis.
- Metric must be easy to use, adaptable to project's needs and targeted market. All parameters must have clear physical meaning, and a method for measuring them.

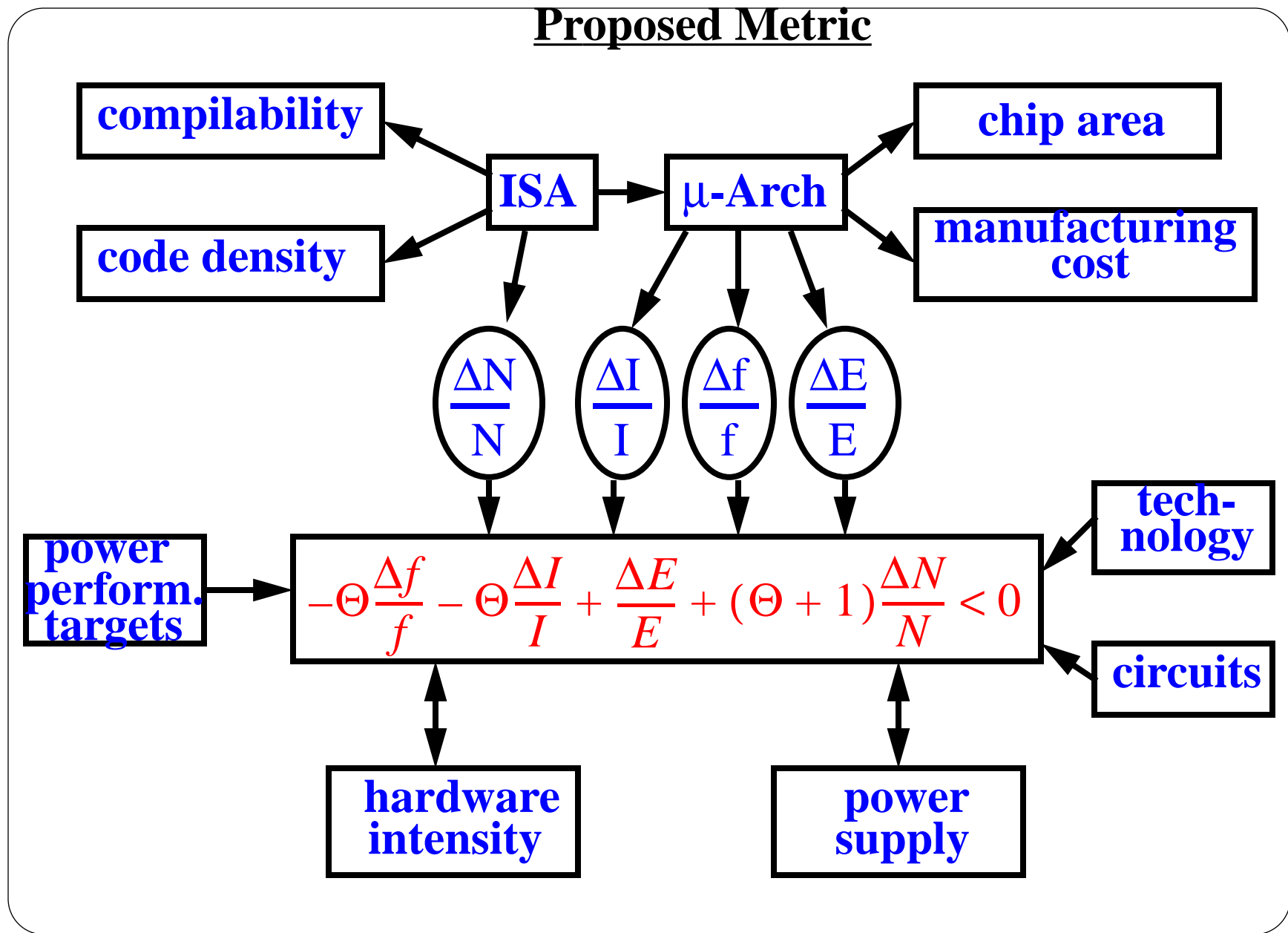
Effects of ISA / Microarchitecture changes



<b>N</b>	<b>dynamic instr. count</b>	<b>f</b>	<b>maximum clocking rate</b>
<b>I</b>	<b>architectural speed, IPC</b>	<b>E</b>	<b>average energy per instr</b>

## State of the Art

- $\frac{\text{MIPS}}{\text{Watt}}$  , reverse of “energy per operation”
  - used for comparing low-end products
  - incorrectly used for “fixed throughput mode” and “power-limited mode”
- $\frac{\text{MIPS}^2}{\text{Watt}}$  , reverse of “energy-delay product”
  - used for comparing mid-range products
- $\frac{\text{MIPS}^\gamma}{\text{Watt}}$  , with “educated” guessing about the value of  $\gamma$ 
  - with  $\gamma = 3$  used for comparing high-end products



## Proposed Metric

- involves only relative changes in processor characteristics
- subsumes all prior art metrics as special cases
- parameters are well defined, with a simple method for measuring them
- parameters take into account
  - technology characteristics
  - circuit style
  - power and performance targets
  - power supply voltage
- derived by solving a multidimensional optimization problem

## Formulation of the Optimization Problem

- independent variables
  - architectural complexity (discrete)  $\xi$
  - power supply voltage  $V$
- functions
  - Performance  $P(\xi, v) = \frac{f(\xi, v)I(\xi)}{N(\xi)}$
  - Power  $W(\xi, v) = f(\xi, v)I(\xi)E(\xi, v)$  or  $W(\xi, v) = f(\xi, v)E(\xi, v)$
- optimization problem
  - A: minimize power  $W$  given a performance requirement  $P = P_o$
  - B: maximize performance  $P$  subject to a power constraint  $W = W_o$

## Alternative Formulation

- evaluate energy efficiency of architectural modification  $\Delta\xi$
- $\Delta\xi \rightarrow \Delta V$  to satisfy the constraint  $P = P_o$  or  $W = W_o$
- $\Delta\xi \rightarrow \Delta N, \Delta f, \Delta I$  and  $\Delta E$
- find relation between relative increments such that
  - A:  $\Delta W < 0$  given a performance requirement  $P = P_o$  ( $\Delta P = 0$ )
  - B:  $\Delta P > 0$  subject to a power constraint  $W = W_o$  ( $\Delta W = 0$ )

**Derivation**

- assume small values of  $\Delta\xi$ ,  $\Delta v$  to neglect second order terms.

- formulation A:

$$\left. \frac{\Delta P}{\Delta \xi} \right|_v \Delta \xi + \frac{\partial P}{\partial v} \Delta v = 0$$

$$\left. \frac{\Delta W}{\Delta \xi} \right|_P = \left. \frac{\Delta W}{\Delta \xi} \right|_v + \frac{\partial W}{\partial v} \frac{\Delta v}{\Delta \xi} \Big|_P < 0$$

- formulation B:

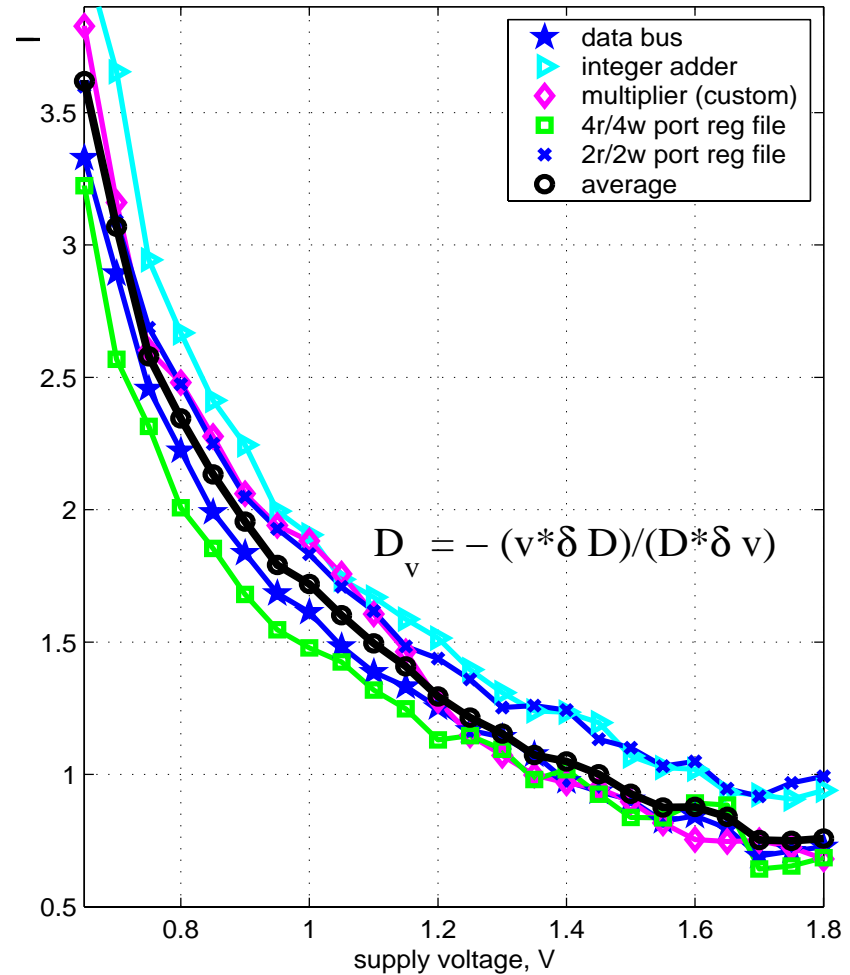
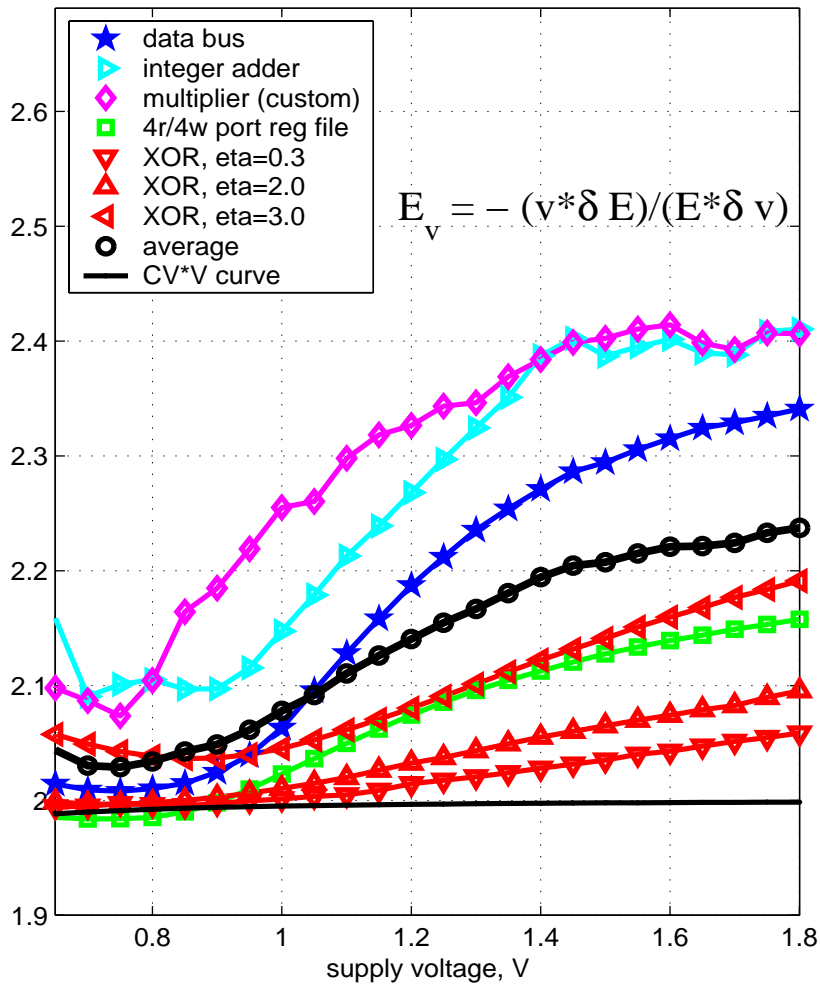
$$\left. \frac{\Delta W}{\Delta \xi} \right|_v \Delta \xi + \frac{\partial W}{\partial v} \Delta v = 0$$

$$\left. \frac{\Delta P}{\Delta \xi} \right|_W = \left. \frac{\Delta P}{\Delta \xi} \right|_v + \frac{\partial P}{\partial v} \frac{\Delta v}{\Delta \xi} \Big|_W > 0$$

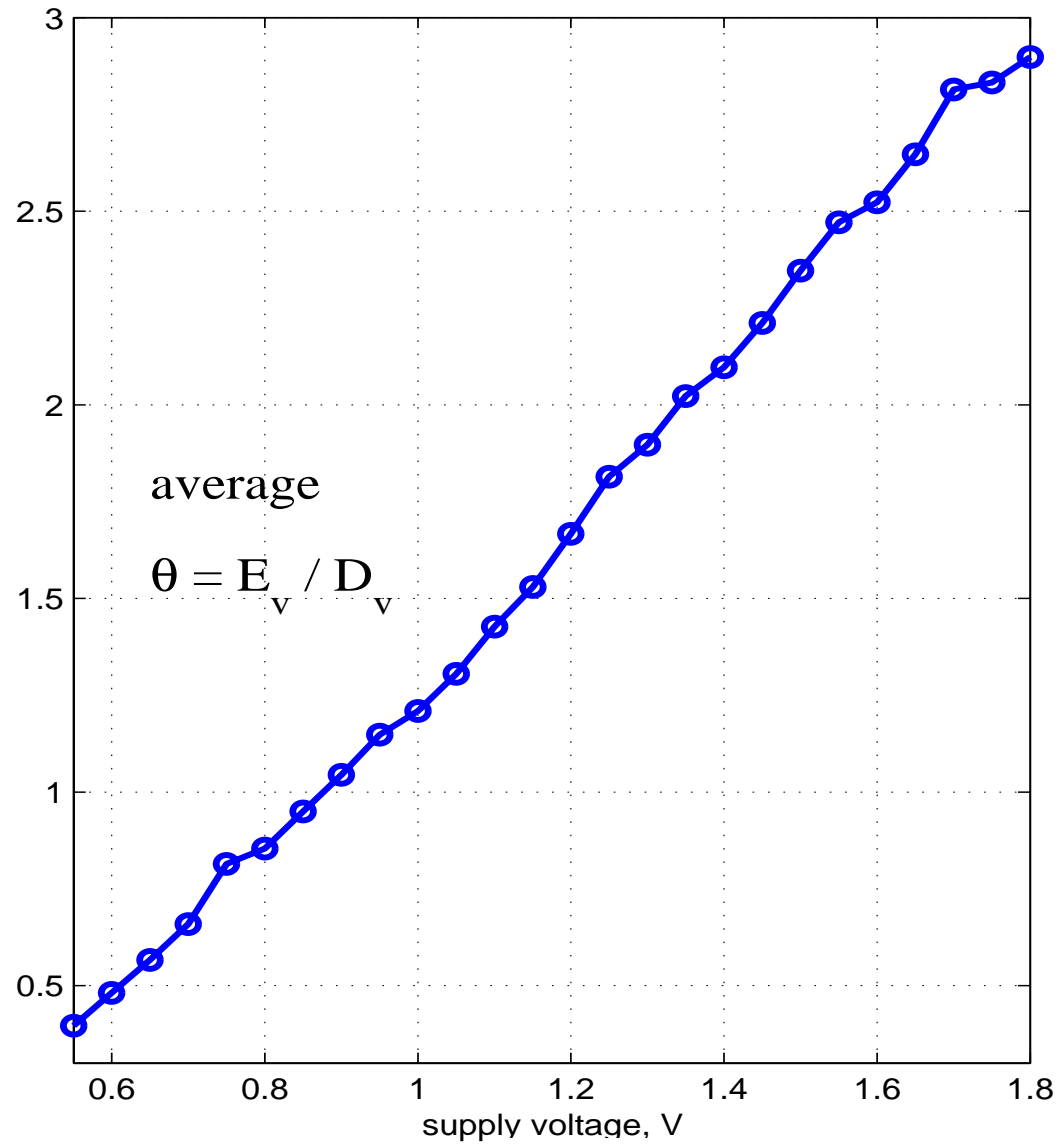
## Derivation

- all terms can be expressed in closed form
- designate  $\theta = \mathbf{E}_v / \mathbf{D}_v$
- both formulations A and B lead to the same expressions
- ideal clock gating 
$$-\Theta \frac{\Delta f}{f} - \Theta \frac{\Delta I}{I} + \frac{\Delta E}{E} + (\Theta + 1) \frac{\Delta N}{N} < 0$$
- no clock gating 
$$-\Theta \frac{\Delta f}{f} - (\Theta + 1) \frac{\Delta I}{I} + \frac{\Delta E}{E} + (\Theta + 1) \frac{\Delta N}{N} < 0$$
- *Vdd* - constrained 
$$-\eta \frac{\Delta f}{f} - \eta \frac{\Delta I}{I} + \frac{\Delta E}{E} + (\eta + 1) \frac{\Delta N}{N} < 0, \eta > \Theta$$

# Measuring Parameters $\theta = E_v / D_v$

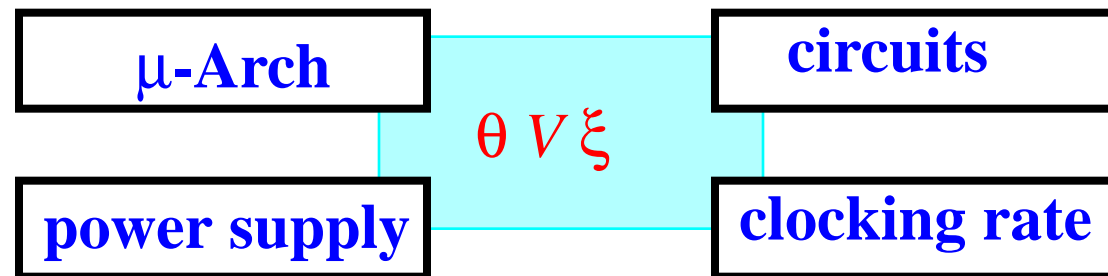


Measuring Parameters  $\theta = E_v / D_{v-}$



## Consequences

- To achieve energy-efficient design architectural choices must be balanced with circuit-level decisions



- Different architectural decisions are optimal for different designs
- Relation to the prior art

$$\frac{\text{MIPS}^\gamma}{\text{Watt}} \quad \gamma = \theta + 1 \quad \text{consistent method for determining } \gamma$$

## Relation to the Prior Art

**Table 1: Special Cases of the Derived Metric**

condition	metric	equivalent prior art
$F_v \gg E_v$ ( $V_{dd} < 0.5V$ ) $\theta \ll 1$ (ultra-low power design)	$\frac{\Delta E}{E} + \frac{\Delta N}{N} < 0$	$\frac{\text{MIPS}}{\text{Watt}}$
$\theta = 1$ ( $V_{dd} = 0.9V$ ) (eLite design point)	$-\frac{\Delta f}{f} - \frac{\Delta I}{I} + \frac{\Delta E}{E} + 2\frac{\Delta N}{N} < 0$	$\frac{\text{MIPS}^2}{\text{Watt}}$
$\theta = 2$ ( $V_{dd} = 1.4V$ ) (nominal CU11 design point)	$-2\frac{\Delta f}{f} - 2\frac{\Delta I}{I} + \frac{\Delta E}{E} + 3\frac{\Delta N}{N} < 0$	$\frac{\text{MIPS}^3}{\text{Watt}}$
$\theta = 3$ ( $V_{dd} > 1.9V$ ) (ultra-high performance design)	$-3\frac{\Delta f}{f} - 3\frac{\Delta I}{I} + \frac{\Delta E}{E} + 4\frac{\Delta N}{N} < 0$	$\frac{\text{MIPS}^4}{\text{Watt}}$
$\theta = 2.7$ $\eta > \theta$ ( $V_{dd} = 1.7V$ ) (power supply constraint mode)	$-\eta\frac{\Delta f}{f} - \eta\frac{\Delta I}{I} + \frac{\Delta E}{E} + (\eta + 1)\frac{\Delta N}{N} < 0$	$\frac{\text{MIPS}^{\eta + 1}}{\text{Watt}}$

## Example 1

- processor A: low power microprocessor with fine-grain clock gating,

$$V_{dd} = 1V, \theta = 1.23$$

Evaluate energy-efficiency of adding the execution bypass.

- Suppose  $\frac{\Delta I}{I} = 0.07$ ,  $\frac{\Delta f}{f} = -0.05$ ,  $\frac{\Delta E_{\text{bypass}}}{E_{\text{total}}} = 0.02$

80% of dynamic instructions read operands from the RF, then  $\frac{\Delta E}{E} = 0.016$

$$-1.23 \left( \frac{\Delta f}{f} + \frac{\Delta I}{I} \right) + \frac{\Delta E}{E} = 1.23(0.05 - 0.07) + 0.016 = -0.023 < 0$$

Execution bypass is energy-efficient.

However, if  $V_{dd} < 0.8V$ , the same feature is not energy-efficient.

## Example 2

- processor B: high-performance dynamic issue microprocessor with no clock gating,  
 $V_{dd} = 1.7V, \theta = 2.7$

Evaluate energy-efficiency of adding an extra read port to a multiported int. RF.

- Suppose  $\frac{\Delta I}{I} = 0.005$ ,  $\frac{\Delta f}{f} = 0$ ,  $\frac{\Delta E_{RF}}{E_{RF}} = 0.1$

int. RF is responsible for 15% of the CPU Power, then  $\frac{\Delta E}{E} = 0.015$

$$-(2.7 + 1)\frac{\Delta I}{I} + \frac{\Delta E}{E} = -3.7 \cdot 0.005 + 0.015 = -0.004 < 0$$

adding an extra port is energy-efficient.

However, if  $V_{dd} < 1.3V$ , the same feature is not energy-efficient.

## Conclusions

- New architectural energy-efficiency metric derived.
- All parameters in the derived metric are well defined, with a clear method for measuring them in every situation.
- Derived metric establishes conditions for a balance between architectural and circuit-level decisions.
- Subsumes prior art as special cases of a more general relation.
- Still has limitations. Some of important factors are not taken into account, such as code size, area, implementation cost