

KOALA

A PLATFORM FOR OS-LEVEL POWER MANAGEMENT

D.C. Snowdon

E. Le Sueur

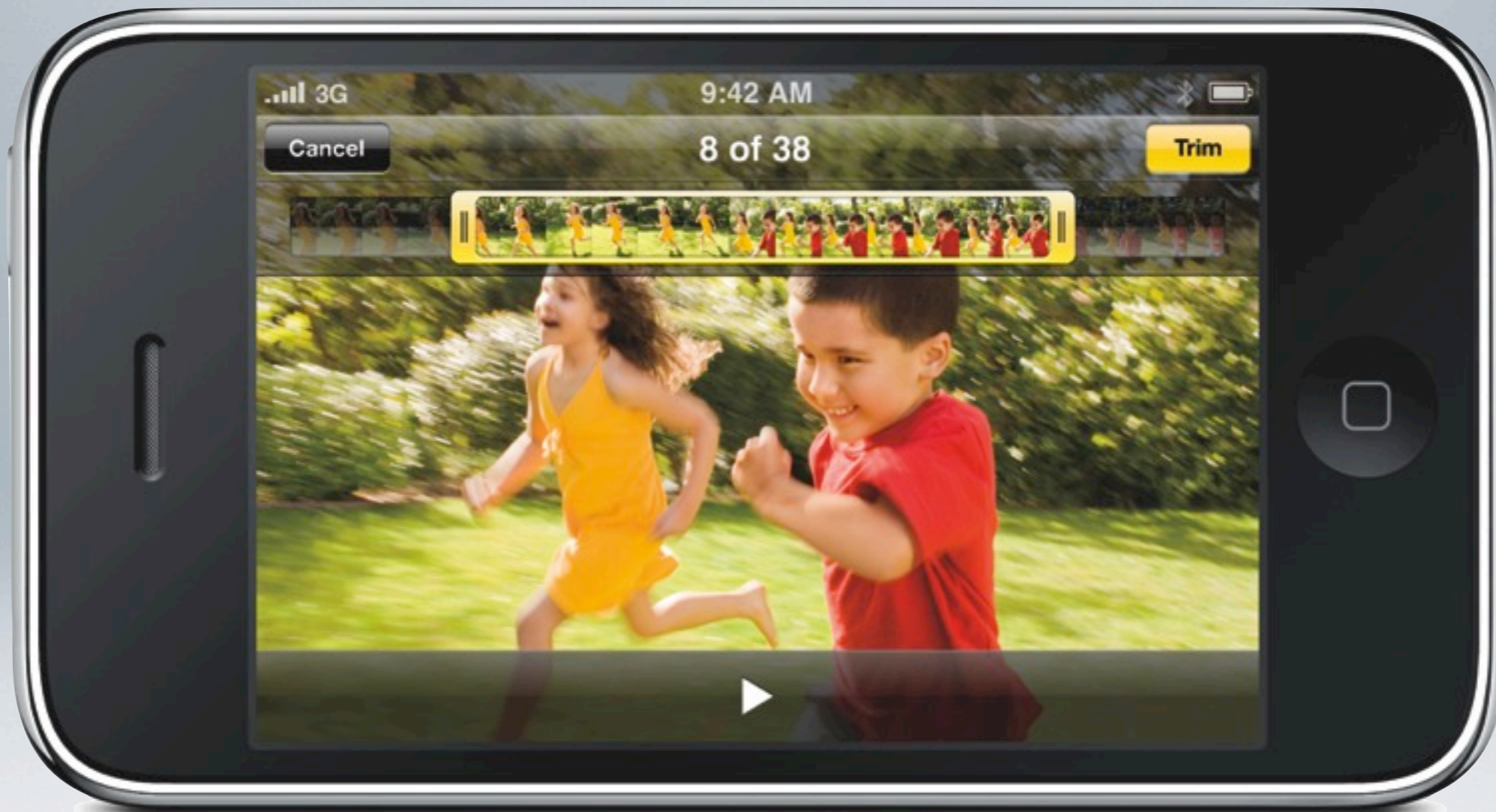
S.M. Petters

G. Heiser





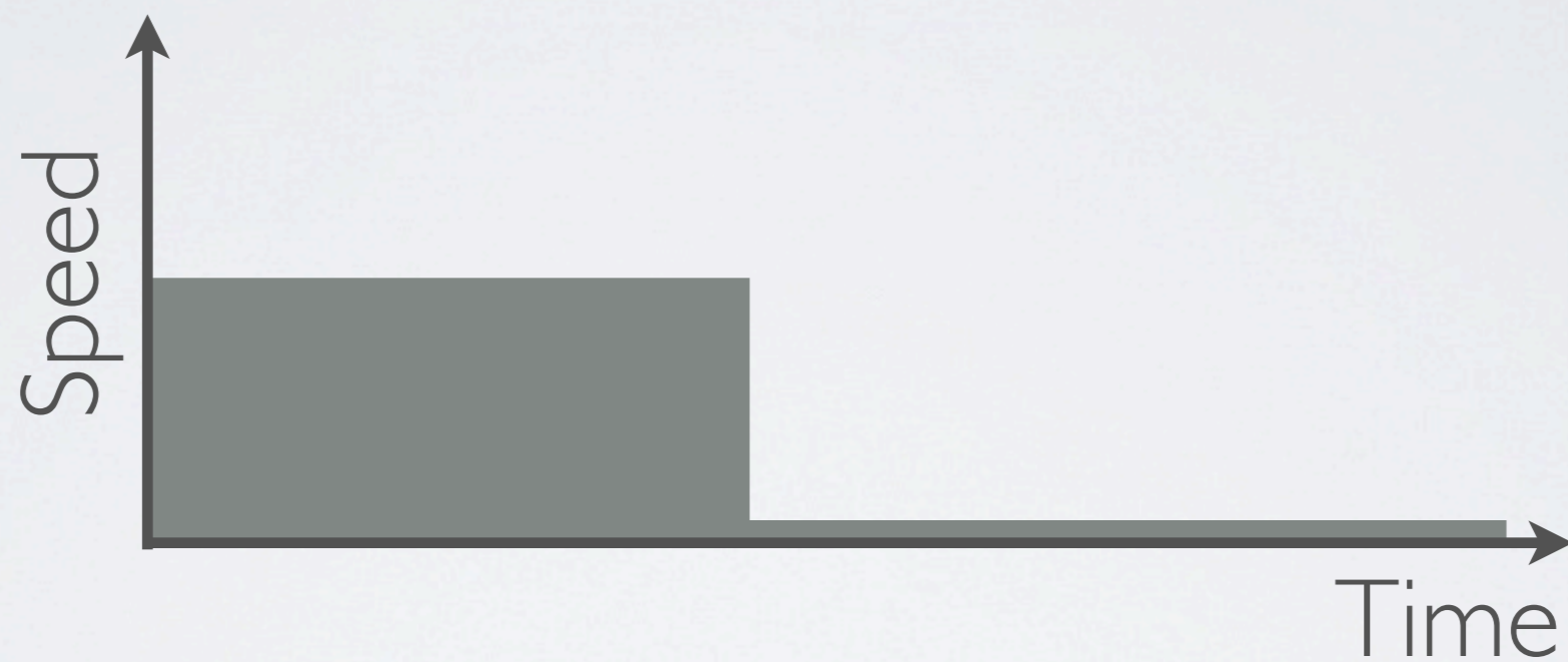
Image by Diliff under CC license



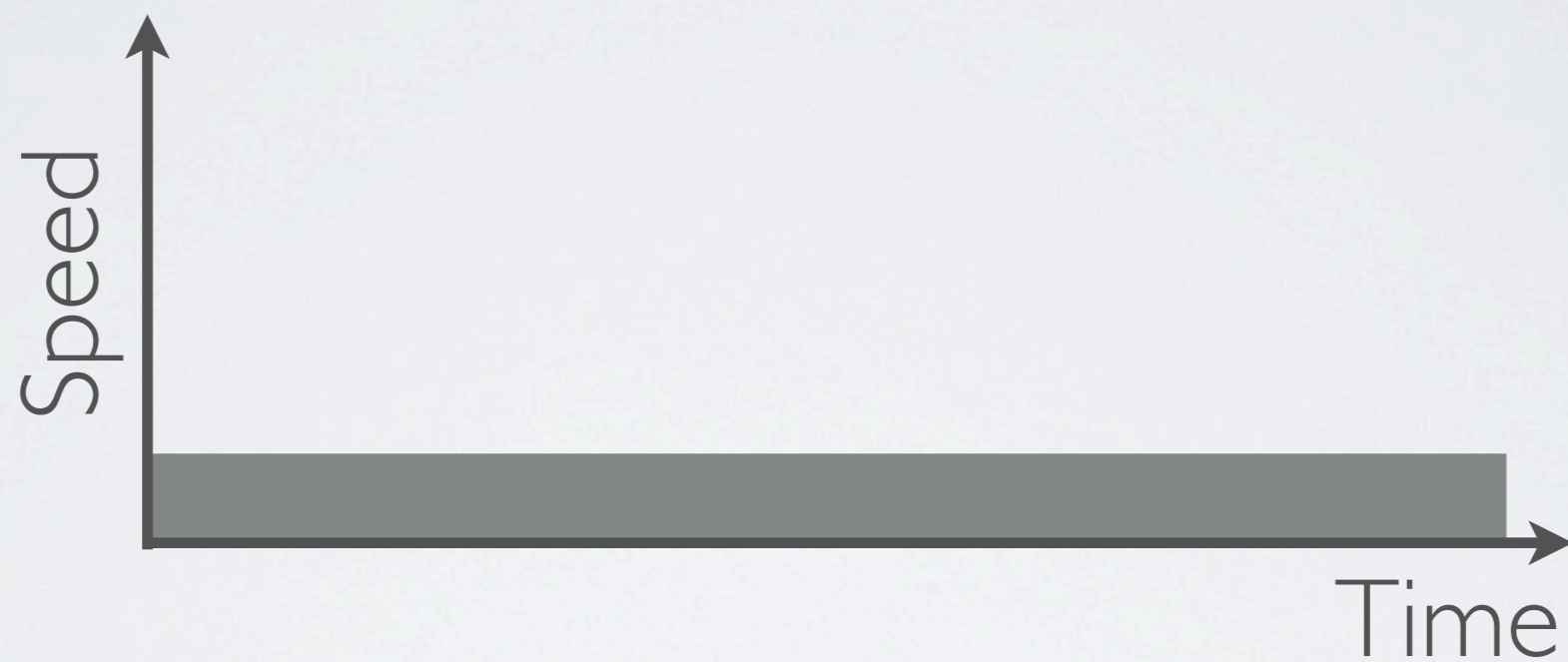
KNOBS

- CPU frequency
- CPU voltage
- CPU sleep states
- memory and bus frequency
- power states of IO devices (not considered here)

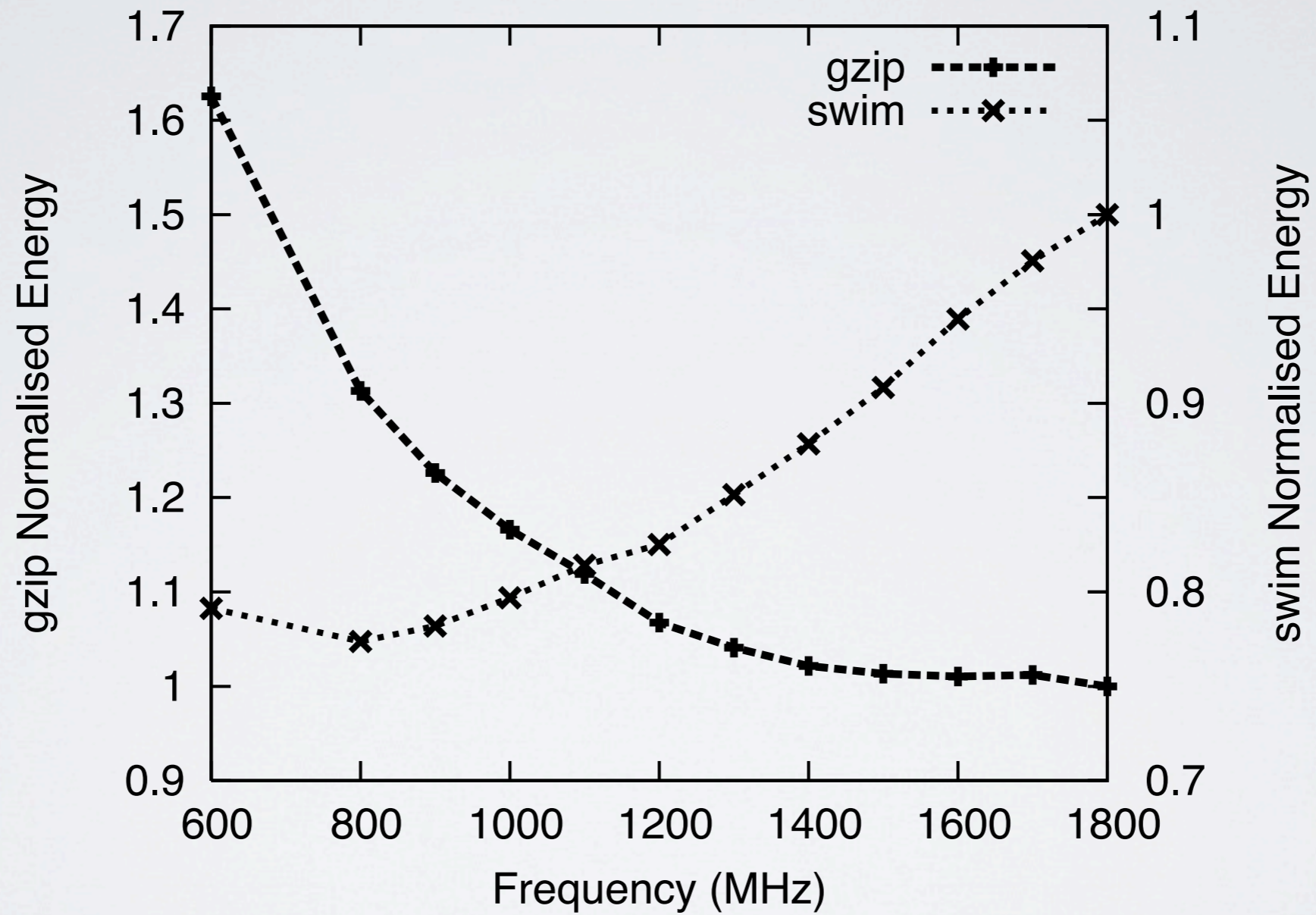
SIMPLISTIC POLICIES



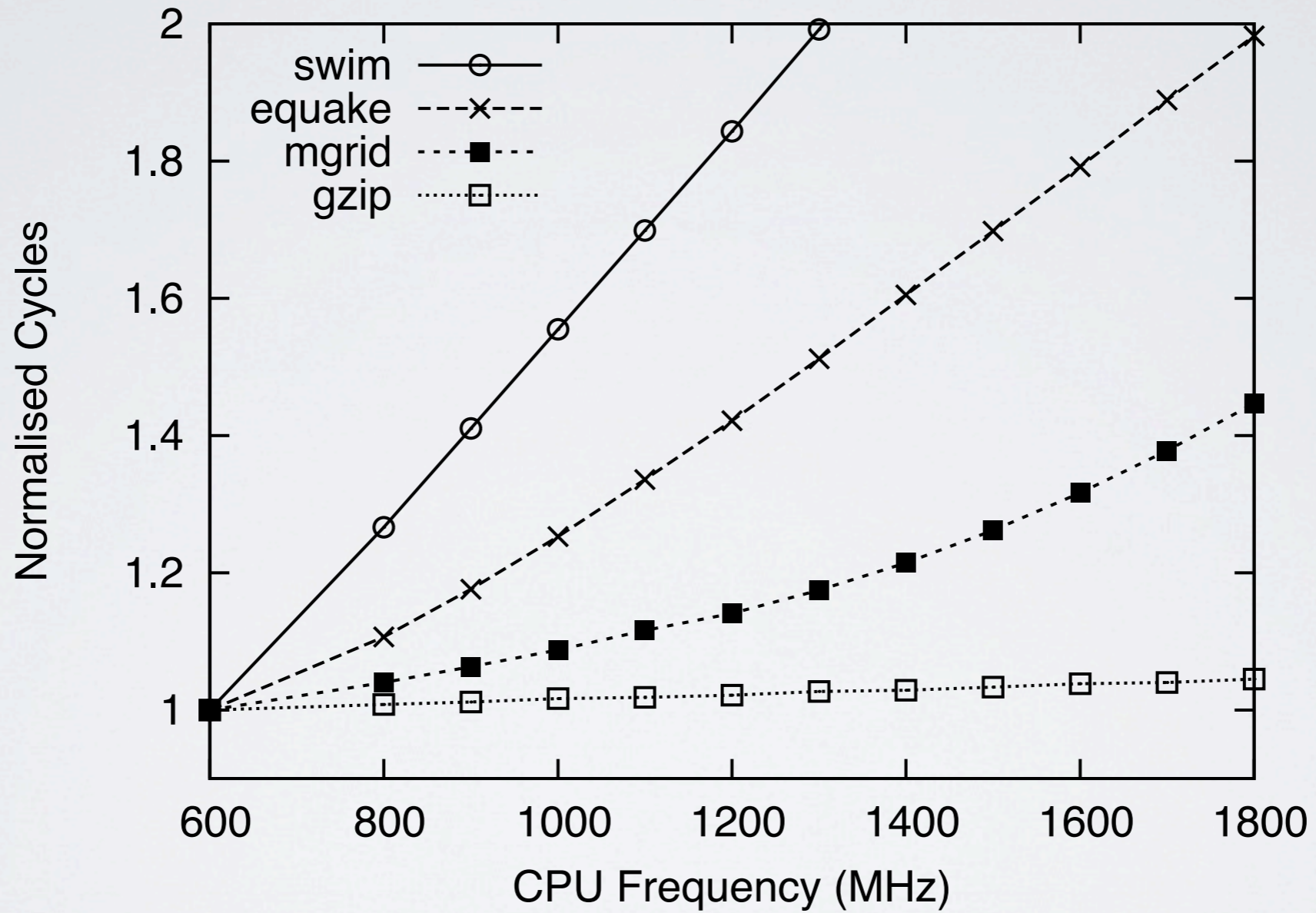
SIMPLISTIC POLICIES



REALITY CHECK



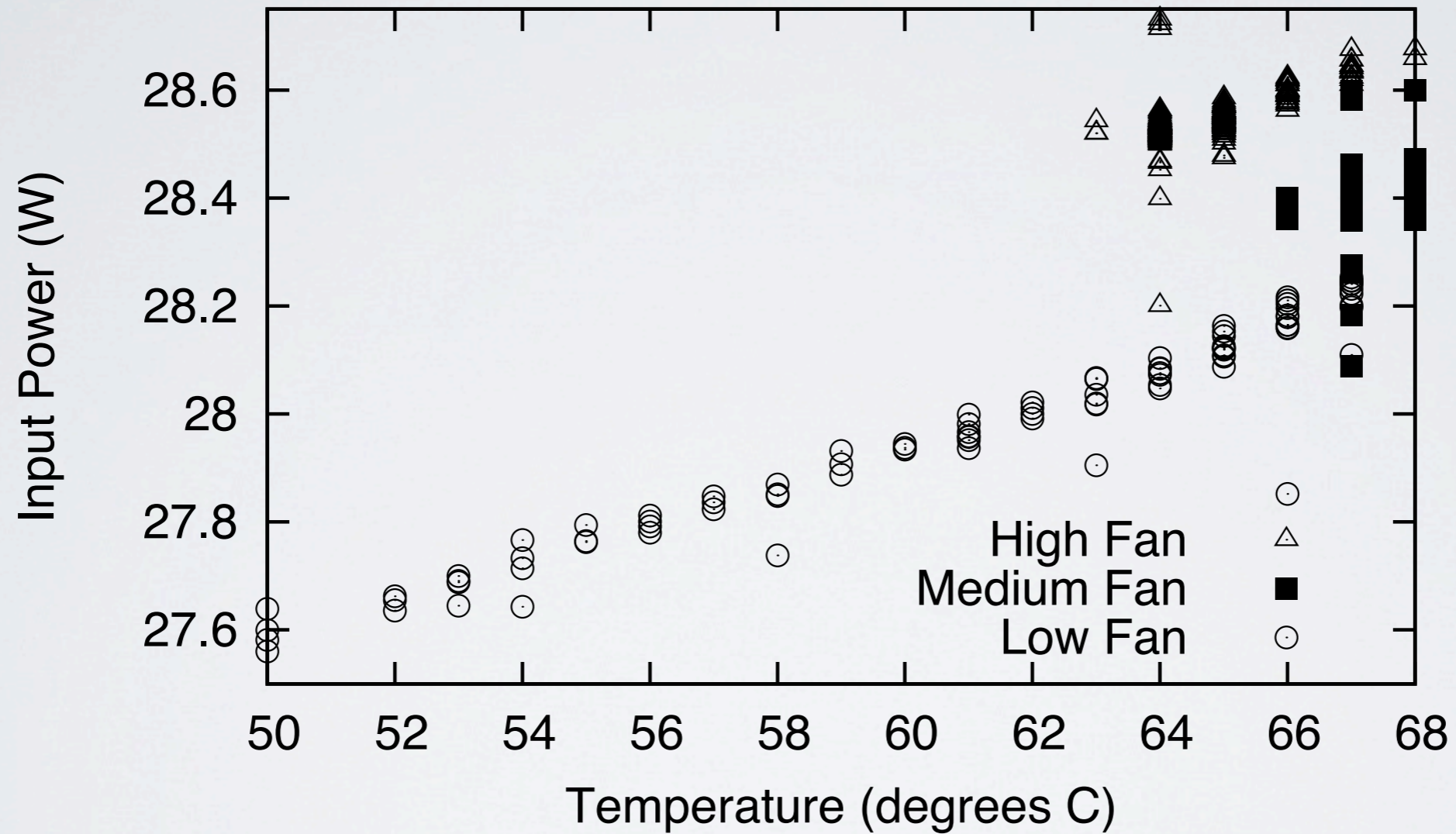
CYCLE COUNT



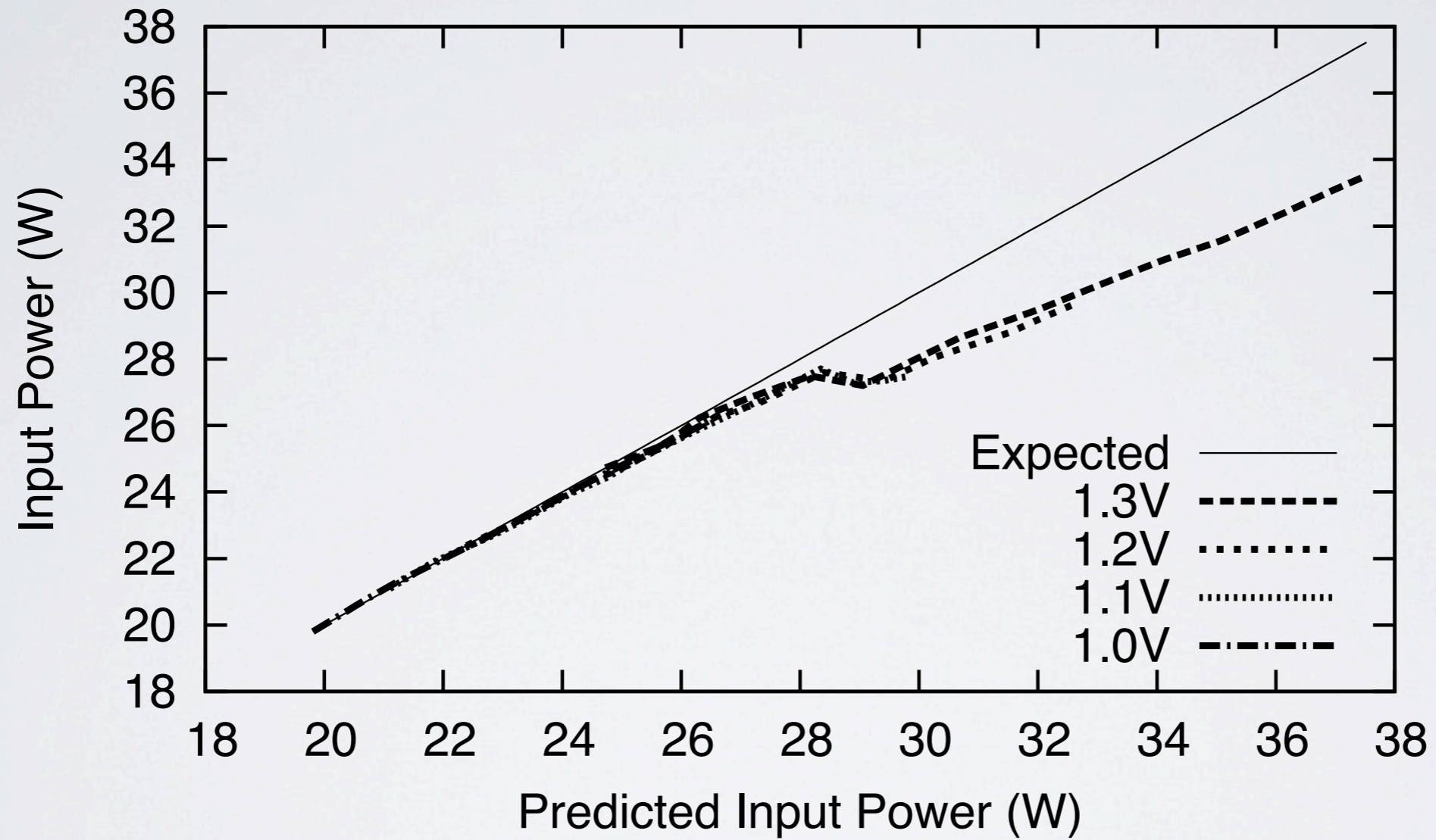
OVERHEAD

- switching frequency and power incurs CPU downtime
- Pentium-M
 - frequency change: $10\mu\text{s}$
 - voltage change asynchronous
- Opteron
 - frequency and voltage change: 2ms ($140\mu\text{s}$ out of spec)

TEMPERATURE



POWER SUPPLY



TIME MODEL

$$T = \frac{C_{cpu}}{f_{cpu}} + \frac{C_{bus}}{f_{bus}} + \frac{C_{mem}}{f_{mem}} + \frac{C_{io}}{f_{io}} + \dots$$

$$C_{bus} = \alpha_1 PMC_1 + \alpha_2 PMC_2 + \dots$$

$$C_{mem} = \beta_1 PMC_1 + \beta_2 PMC_2 + \dots$$

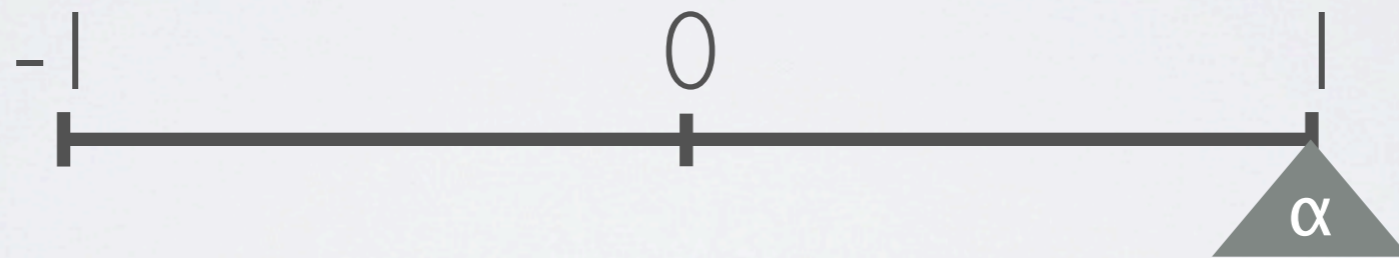
ENERGY MODEL

$$E_{dyn} \propto cyc \times V^2$$

$$\begin{aligned} E = & V_{cpu}^2 (\gamma_1 f_{cpu} + \gamma_2 f_{bus} + \gamma_3 f_{mem}) \Delta t + \\ & V_{cpu}^2 (\alpha_0 PMC_0 + \dots + \alpha_m PMC_m) + \\ & \gamma_4 f_{mem} \Delta t + \beta_0 PMC_0 + \dots + \beta_m PMC_m + \\ & P_{static} \Delta t \end{aligned}$$

UNIFIED POLICY

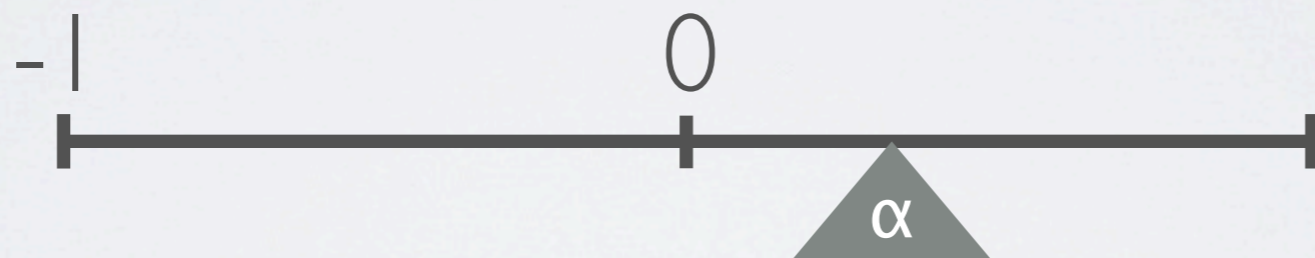
$$\eta = P^{1-\alpha} T^{1+\alpha}$$



forces highest performance

UNIFIED POLICY

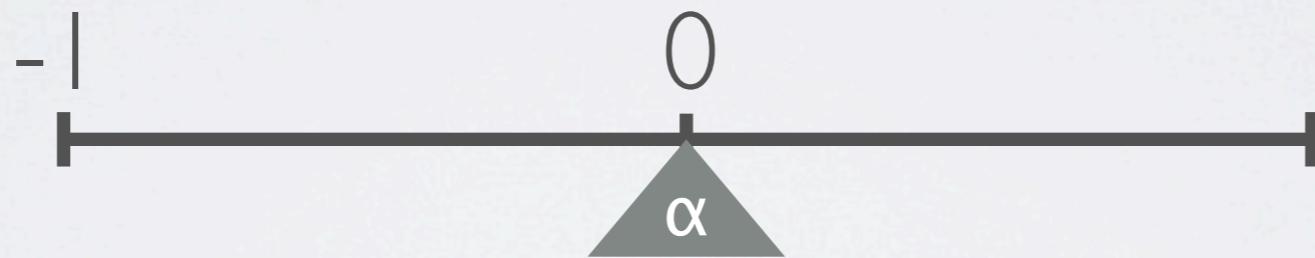
$$\eta = P^{1-\alpha} T^{1+\alpha}$$



minimises energy-delay product

UNIFIED POLICY

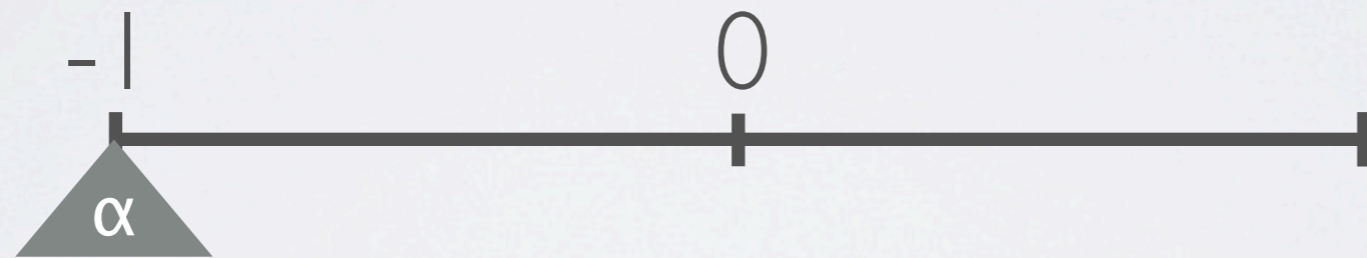
$$\eta = P^{1-\alpha} T^{1+\alpha}$$



minimises energy

UNIFIED POLICY

$$\eta = P^{1-\alpha} T^{1+\alpha}$$



minimises power consumption

IMPLEMENTATION

- recent Linux kernel (2.6.24.4)
- per-process collection of relevant statistics
- policy-decision when process blocks or preempts
 - use data from previous time slice to predict optimal setting
 - assumes temporal locality
- uses logarithmic tables to simplify calculation (no float)

EVALUATION

The Laptop

- Dell Latitude D600
- Pentium-M 0.8 – 1.8 GHz
- 0.98 – 1.34 V
- three sleep states
- measured at battery

The Server

- AMD Opteron 246
- 0.8 – 2 GHz
- 0.9 – 1.5 V
- high switching overhead
- measured at wall socket

CHARACTERISATION

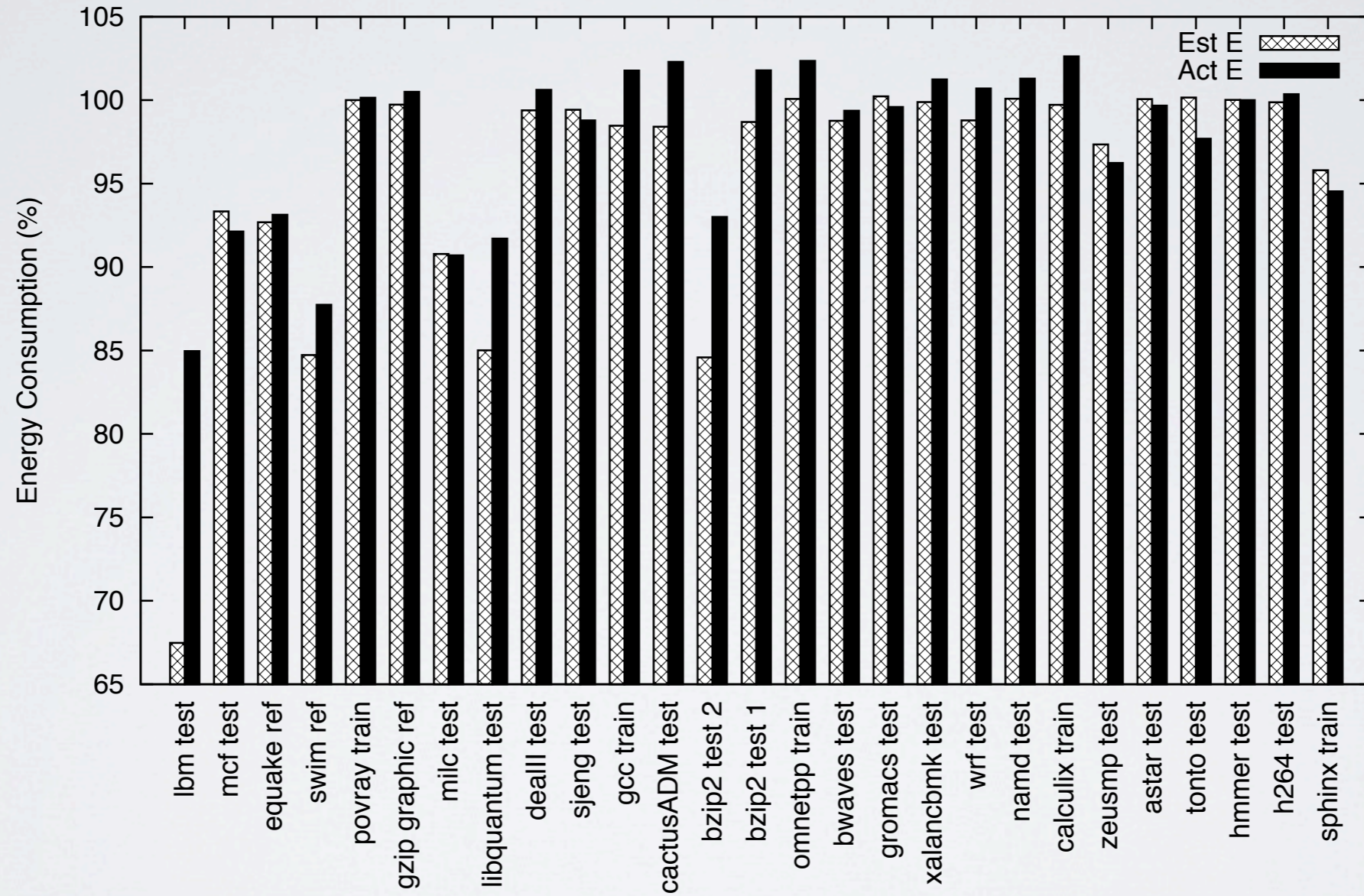
The Laptop

- number of completed burst transactions
- number of lines removed from L2 cache
- correlation 0.98 / 0.96

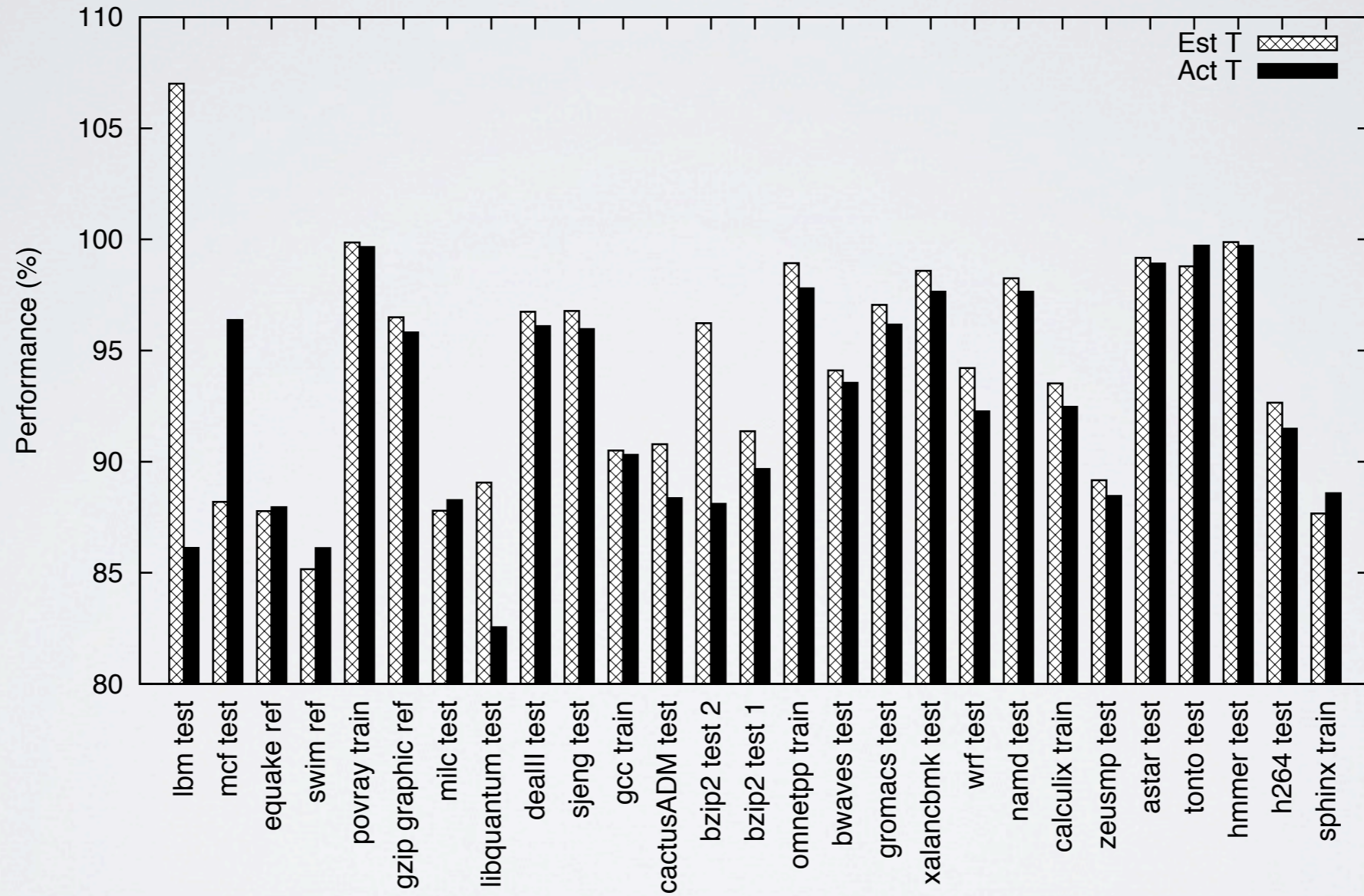
The Server

- quadword write transfers
- L2 cache misses
- dispatch stalls due to reorder buffer being full
- DRAM accesses due to page conflicts
- correlation 0.98 / 0.98

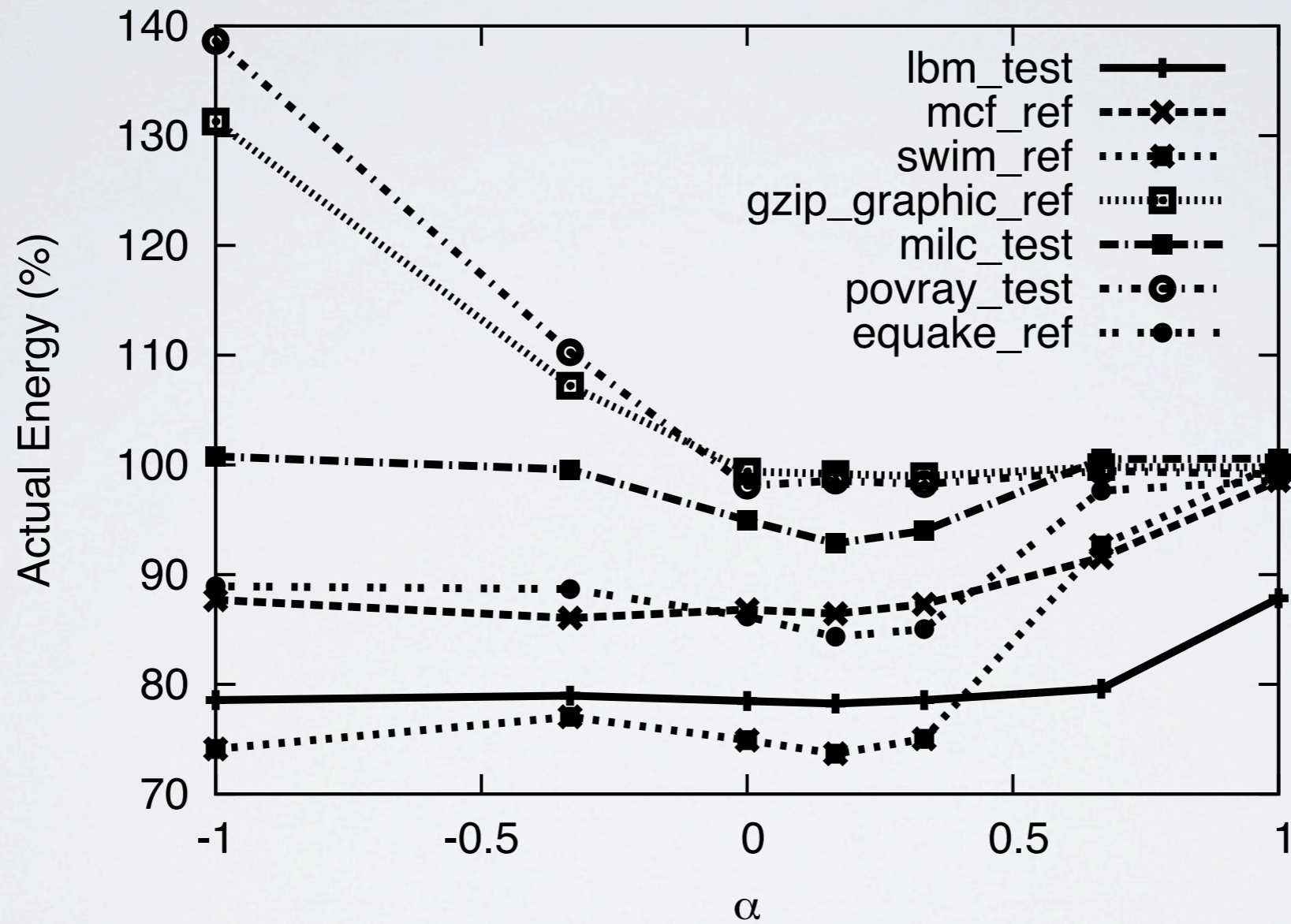
MODEL ACCURACY



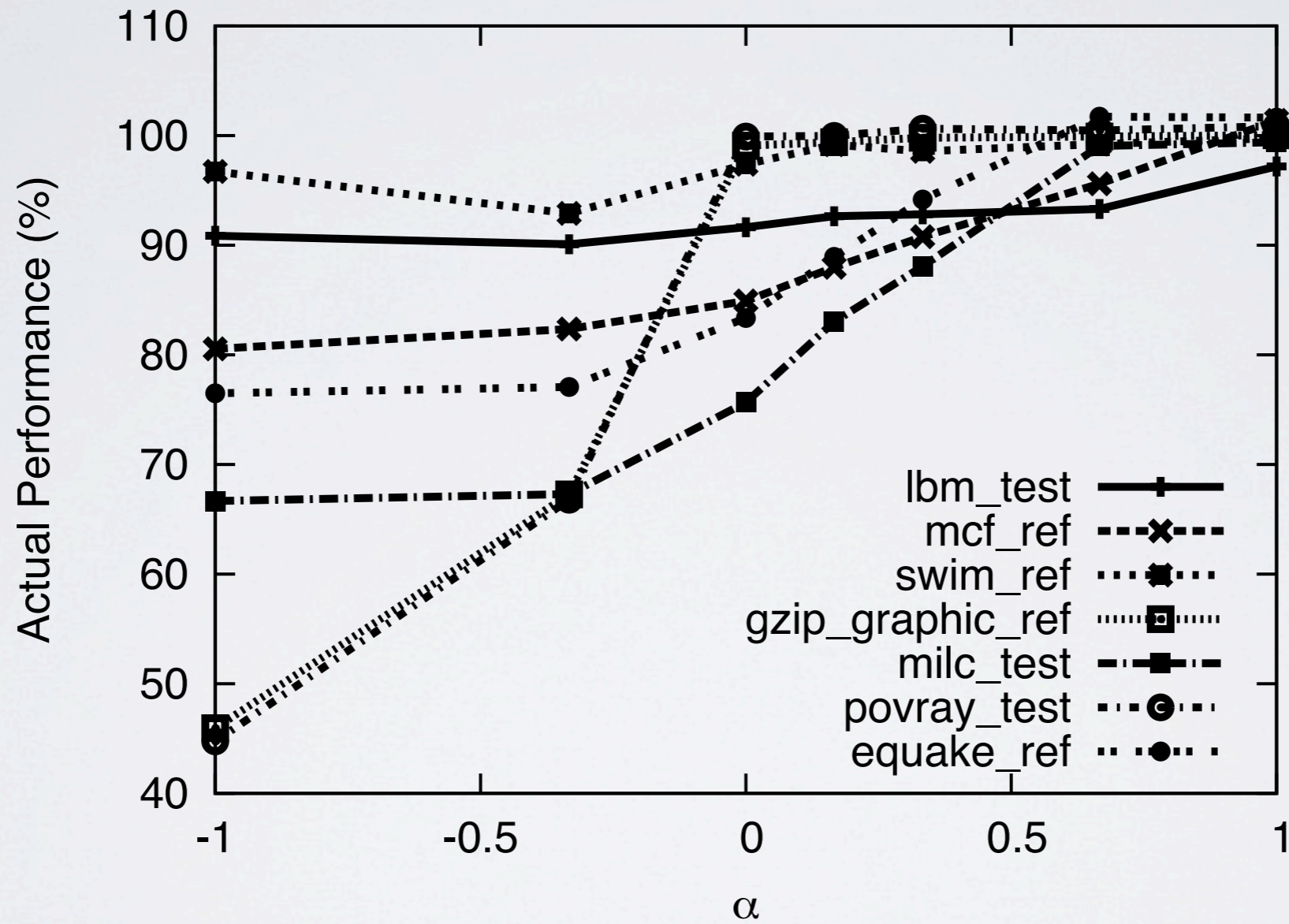
MODEL ACCURACY



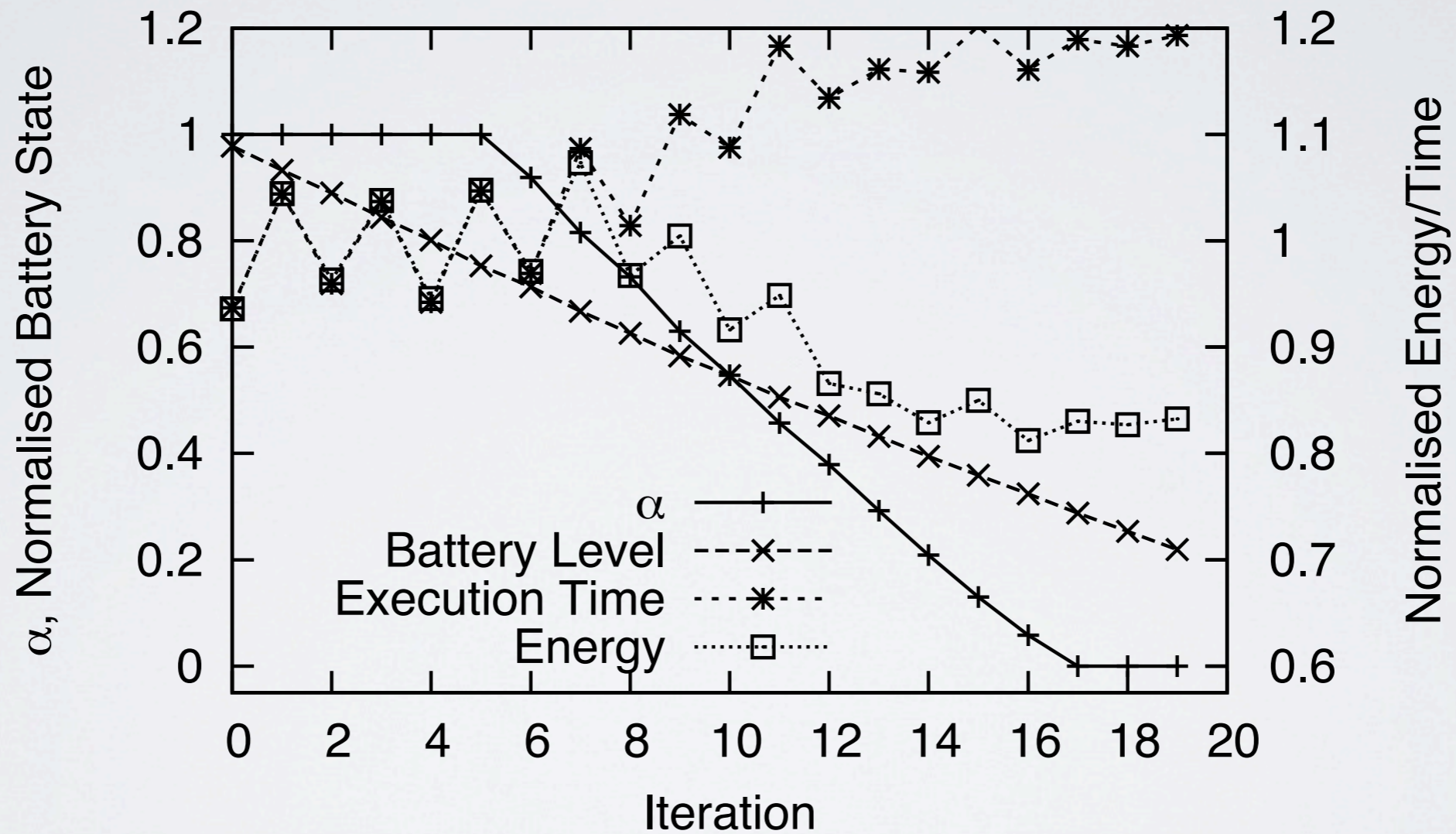
UNIFIED POLICY



UNIFIED POLICY



BATTERY-AWARE POLICY



DISCUSSION

- practicality
- cooperation from vendors, built-in power measurement
- energy management by hardware or software
- hints from applications
- is this too fine-grained
- dumb component shutdown (both software and hardware)