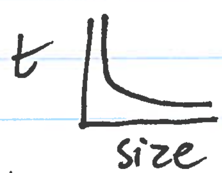
Energy Optimization

- power is usually a hard constraint: exceeding will damage circuit
  e.g. TDP ~~to~~ thermal design power = power consumption under max theoretical load
- optimize for energy: $E = Pt$

Caches

- power consumption $\propto$ size^2
- ~~exe~~ execution time decreases non-linearly with size



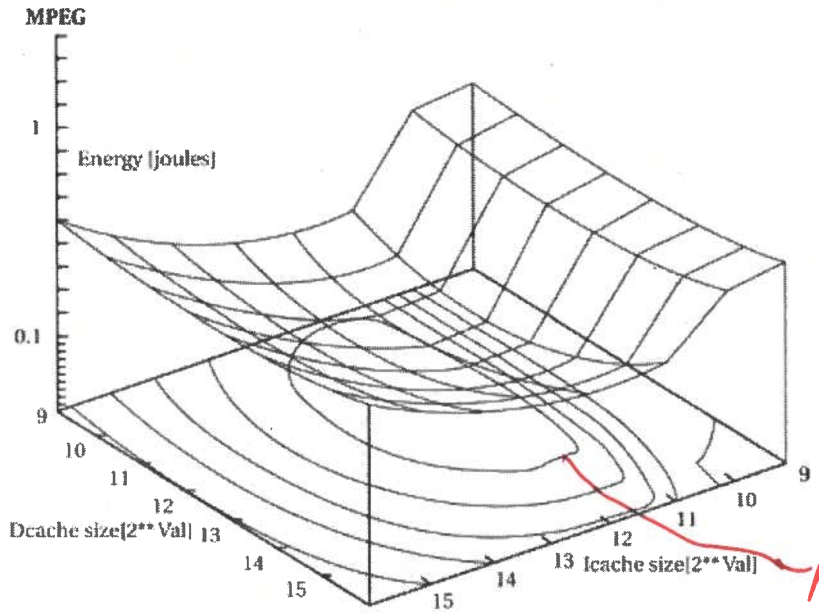
- see Handouts > energy MPEG energy vs cache size diagram

Dynamic Voltage and Frequency Scaling

- most effective when dynamic power consumption dominates
- $P \propto V^2$
- logic gate delay $\propto V \Rightarrow f \propto V$
- running at a lower voltage increases energy efficiency
- generally OSs monitor workload and adjust V and f
- we focus on picking V and f for an embedded application
- execution time, $t = \frac{K}{f}$ — total PE cycles to execute application
- goal: find best V and f such that $P < P_{max}$ and $t < T$
  └ max execution time

2

MPEG: energy consumption vs cache size

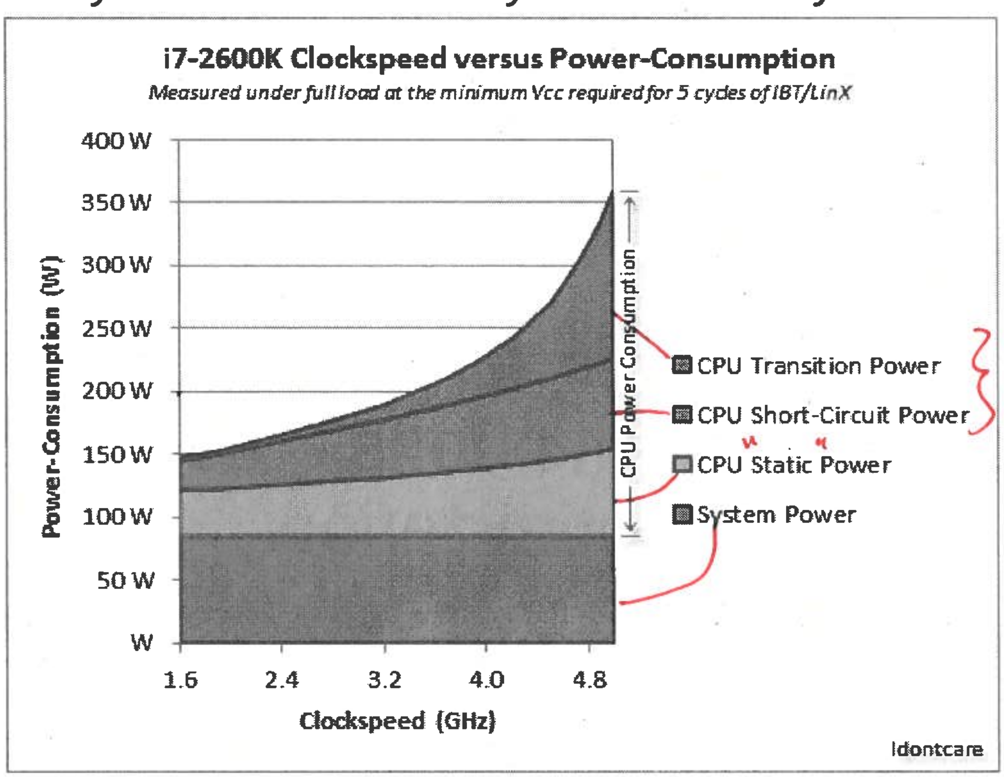


points with same energy consumption

R Pellizzoni (publication unknown)

$$w = R + 46.9V + 14.4f + 3.7V^{4.9}f$$

system static dynamic-sc dynamic-st



dynamic

R Pellizzoni (publication unknown)

CMOS power model

- dynamic power consumption
  a) signal transition: $W_{ST} = \frac{1}{2} C \times SW \times V^2 \times f$

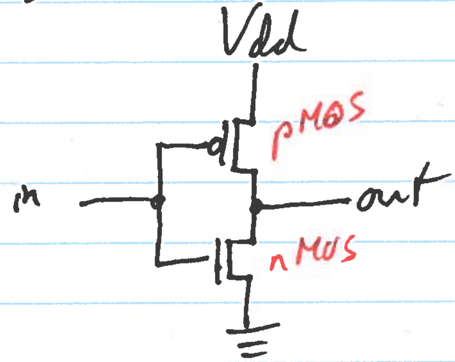
  due to output capacitance

  SW ← switching rate: rate of change of logic gate outputs

  b) short circuit: $W_{SC} = E_{SC} \times SW \times f$

  due to non-zero rise/fall times

  CMOS Inverter



- static power consumption
  leakage current - becomes more important as device sizes decrease
  e.g. subthreshold leakage through transistor
  independent of f

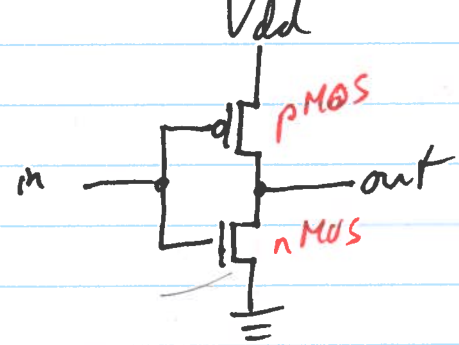
- system power
  loses from other components such as power supply, I/O
  (generally static)

  - see clockspeed vs power diagram

CMOS power model

- dynamic power consumption
  - a) signal transition: $W_{ST} = \frac{1}{2} C \times SW \times V^2 \times f$
    - SW = switching rate: rate of change of logic gate outputs
    - due to output capacitance
  - b) short circuit: $W_{SC} = E_{SC} \times SW \times f$
    - due to non-zero rise/fall times

CMOS Inverter



- static power consumption
  - leakage current - becomes more important as device sizes decrease
    - e.g. subthreshold leakage through transistor
  - independent of f

- system power
  - loses from other components such as power supply, I/O (generally static)

  - see clockspeed vs power diagram (page 2)

Example: Battery Powered CMOS Device

- assume $W_{ST} = W_{leakage}$ and W_{SC} and W_{sys} are negligible
  dynamic static
- assume $W_{leakage} \propto V$
- $E_{task} = (W_{ST} + W_{leakage}) \times t = 2wt$
- option 1: scale frequency only (V remains unchanged)

$$f' = f/2$$

$$W_{ST}' = W_{ST}/2, \quad W_{leakage} \text{ unchanged}, \quad t' = 2t$$

$$E'_{task} = (W/2 + W) \times 2t = 3wt \quad > 2wt$$

- option 2: scale voltage (requires scaling frequency)

$$V'' = V/2, \quad f'' = f/2$$

$$W_{ST} = \tfrac{1}{2} C \times SW \times V^2 \times f$$

$$W''_{ST} = \tfrac{1}{2} C \times SW \times \left(\tfrac{V}{2}\right)^2 \times \tfrac{f}{2} = W_{ST}/8$$

$$W''_{leakage} = W_{leakage}/2, \quad t'' = 2t$$

$$E''_{task} = (W/8 + W/2) \times 2t = 1.25wt \quad < 2wt$$

- need to scale both V and f to save energy

DVFS for PE + memory

- execution time, $t = K/f_c + M/f_m$
  - K: active core (PE) cycles (mem. standby)
  - M: active memory cycles (core standby)

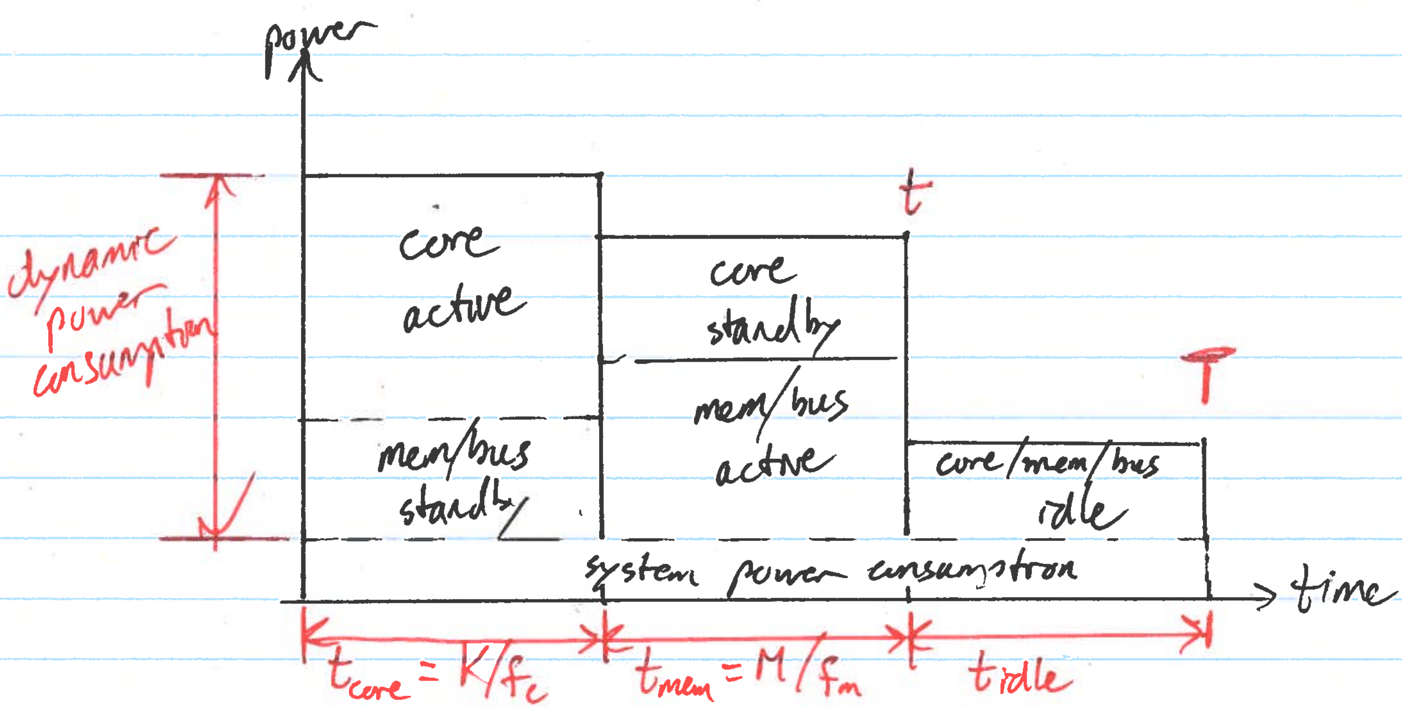
  f_c = core frequency, f_m = memory frequency

- assume that bus and memory share a clock and are active/standby at the same time

- core (PE) states
  - Active – processing instructions
  - Standby – waiting for memory
  - Idle – low power mode (independent of V and f)
- use same kind of model for memory + bus
- ignore W_{sc} (dynamic-short circuit) and $W_{leakage}$ (static) power consumption

Task Set Power Model

- task set super period, T, (= LCM of periods)
- constraint $t < T$



power mode continued

E_{xy}: $x = c$ (core) or m (memory)
$y = a$ (active) or s (standby)
* the same voltage is used for all components

$$E_{taskset} = E_{ca/ms} + E_{cs/ma} + E_{idle}$$

$$= \left(\overbrace{C_{ca} \times V^2 \times f_c}^{W_{ST}} + \overbrace{C_{ms} \times V^2 \times f_m}^{W_{ST}} + \overbrace{R}^{W_{sys}}\right) \times \frac{K}{f_c} \quad (E_{ca/ms})$$

$$+ \left(C_{cs} \times V^2 \times f_c + C_{ma} \times V^2 \times f_m + R\right) \times \frac{M}{f_m} \quad (E_{cs/ma})$$

$$+ \underbrace{(I + R)}_{\text{idle power}}\underbrace{(T - t)}_{t_{idle}}$$

- see ARM926EJ-S data + graph (page 3)