Dark Silicon

Dan Butler

2/27/13

GreenDroid & Intro to "dark silicon"

Claim:

On 3W of power, only 1% of a typical mobile chip can switch at full frequency

Authors describe trends that led to this situation

Outline their approach to utilizing "dark silicon"

GreenDroid & Intro to "dark silicon"

- Idea of GreenDroid is to use more transistors (which are plentiful) but less power (which is scarce)

- 100 "highly specialized energy-reducing cores"

- Claim "11 times less energy" at same or better performance

GreenDroid & Intro to "dark silicon"

Transistor property	Classical	Leakage-limited
Δ Quantity	S^2	S^2
Δ Frequency	S	S
Δ Capacitance	$1/S$	$1/S$
$\Delta\ V_{dd}^2$	$1/S^2$	1
$\Rightarrow \Delta$ Power $= \Delta\ QFCV^2$	1	S^2
$\Rightarrow \Delta$ Utilization $=$ 1/Power	1	$1/S^2$

Table 1. *Classical vs. leakage-limited scaling. In contrast to the classical regime proposed by Denard, under the leakage-limited regime, the total chip utilization for a fixed power budget drops by a factor of S^2 with each process generation.*

- S is "scaling factor" between feature size (S = 45/32 = 1.4x)

- Leakage breaks the voltage scaling law

- Whole-chip power consumption increases with smaller features

GreenDroid & Intro to "dark silicon"

- Per-transistor consumption is constant, but the transistors are smaller

- Can't you just space out the components more to avoid the leakage problems?

- GreenDroid seems to point towards hardware specialization?
  - Put many specialized circuits on a chip?

End of Multicore Scaling

- People have assumed that multi-core is the way to overcome the end of Denard Scaling

- Paper argues the attainable speed-ups from multicore will not support expected perf increases

- Paper argues dark silicon will be pervasive

End of Multicore Scaling

- Device scaling model
  - How area, freq, power reqs will evolve through 2024
- Core scaling model:
  - Attainable power/perf and area/perf "Pareto frontier" (optimal design space) for single core
- Multicore scaling model:
  - Attainable area, power, perf for different CPU- and GPU-like multicore configurations
- Cartesian products of these design spaces

End of Multicore Scaling

Conclude:

- Over next 5 generations, only 7.9x speedup (instead of 32x)

- Dark silicon will increase to 20% then 50% of chip area

End of Multicore Scaling

- In 8/12 benchmarks, there isn't enough parallelism to sustain Moore's law even when power is allowed to climb up to 500W
- In 4/12 benchmarks, there is sufficient parallelism, but only at overly high power consumption (not attainable, resulting in dark silicon)

End of Multicore Scaling

- "...few applications can efficiently use a 100-core or 1000-core chip..."
  - Is this fundamentally true? I thought the theoretical parallelization limit of most things was very high?

- Why are their estimates of % dark silicon so different from GreenDroid paper?