CS 152 Computer Architecture and Engineering

Lecture 24: The Future and Closing Remarks

John Wawrzynek

http://inst.eecs.berkeley.edu/~cs152
Future Application Drivers

- Pervasive Speech
- Robotics
- Social Networks
- Self-Driving Cars
- Augmented Reality & VR
- BIG DATA
- Personalized Medicine
Compute Energy “Iron Law”

Performance = Power * Energy Efficiency

\(\text{Performance} = \frac{\text{Tasks/Second}}{\text{Tasks/Joule}} \times \frac{\text{Power} (\text{Joules/Second})}{\text{Energy Efficiency}}\)

- When power is constrained, need better energy efficiency for more performance
- Where performance is constrained (real-time), want better energy efficiency to lower power

*Improving energy efficiency is critical goal for all future systems and workloads*
Good News: Moore’s Law Continues

“Cramming more components onto integrated circuits”, Gordon E. Moore, Electronics, 1965
Bad News: Dennard (Voltage) Scaling Over

Moore, ISSCC Keynote, 2003

Dennard Scaling

Post-Dennard Scaling

Data courtesy S. Borkar/Intel 2011
1st Impact of End of Scaling: End of Sequential Processor Era

Data partially collected by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond
Parallelism: “A one-time gain” (limited, at least)

Use more, slower cores for better energy efficiency.

Either

- simpler cores, or
- run cores at lower Vdd/frequency

- Even simpler general-purpose microarchitectures?
  - Limited by smallest sensible core
- Even Lower Vdd/Frequency?
  - Limited by Vdd/Vt scaling, errors
- Now what?
2nd Impact of End of Scaling: “Dark Silicon”
Cannot switch all transistors at full frequency!

No savior device technology on horizon.

Future energy-efficiency innovations must be above transistor level.
The Problem With Programmability

- [Cong, 2014]. This is power breakdown for a superscalar out-of-order processor, typically configured
Accelerators

- Specialized hardware that is fixed function
  - One application
  - Or one function common to many applications (e.g., FFT)

```c
typedef unsigned long U32;

U32 cyclic_mac(U32 *p1, U32 *p2)
{
    U32 sum = 0;
    int i;
    for(i = 0; i < BUF_SIZE*4; ++i)
    {
        sum += *p1++ * *p2++;
        if(i % BUF_SIZE) == (BUF_SIZE - 1))
        {
            p1 = BUF_SIZE;
        }
    }
    return sum;
}
```

```assembly
; Enabling modulo addressing for r0
lb0 0x1, moduen
; Setting modulo factor for r0
lb0 64, modi

; Loop prologue
mpy (r0).dw+1, (r1).dw+1
mpypa (r0).dw+1, (r1).dw+1, a0
rep 127
; Loop body
mac (r0).dw+1, (r1).dw+1, a0
ret(ds1, t)
; Disabling modulo addressing for r0
lb0 0x0, moduen
```

![Logic circuit diagram](image)
**The End of General-Purpose Processors?**

- Most computing happens in specialized, heterogeneous processors
  - Can be 100-1000X more efficient than general-purpose processor
- Challenges:
  - Hardware design costs
  - Software development costs

NVIDIA Tegra2
Related Courses

- **CS61C**: Basic computer organization, first look at pipelines + caches
- **CS152/251A**: Computer Architecture, First look at parallel architectures
- **EECS151/251A**: Digital Design
- **EECS251B**: Digital Design (chip design)

**Strong Prerequisite**

- **EECS251B**

**Prerequisites**

- **EECS251B**
- **CS162**: Operating Systems and System Programming
- **CS168**: Introduction to the Internet: Architecture and Protocols
- **CS161**: Computer Security, **EE 122**: Introduction to Communication Networks

**Computer Architecture**

- **EE 130**: Integrated-Circuit Devices
- **EE C149**: Introduction to Embedded Systems
- **EE 140**: Analog Integrated Circuits
End of CS152/252A

- Make sure to do class evaluation
- Thanks for taking the class!
- Best of luck on the final exam.