Power Limitation Impact: Computing vs. Wireless

Gerd Ascheid

MPSoC'16

ICE | RWTH AACHEN UNIVERSITY

Abstract

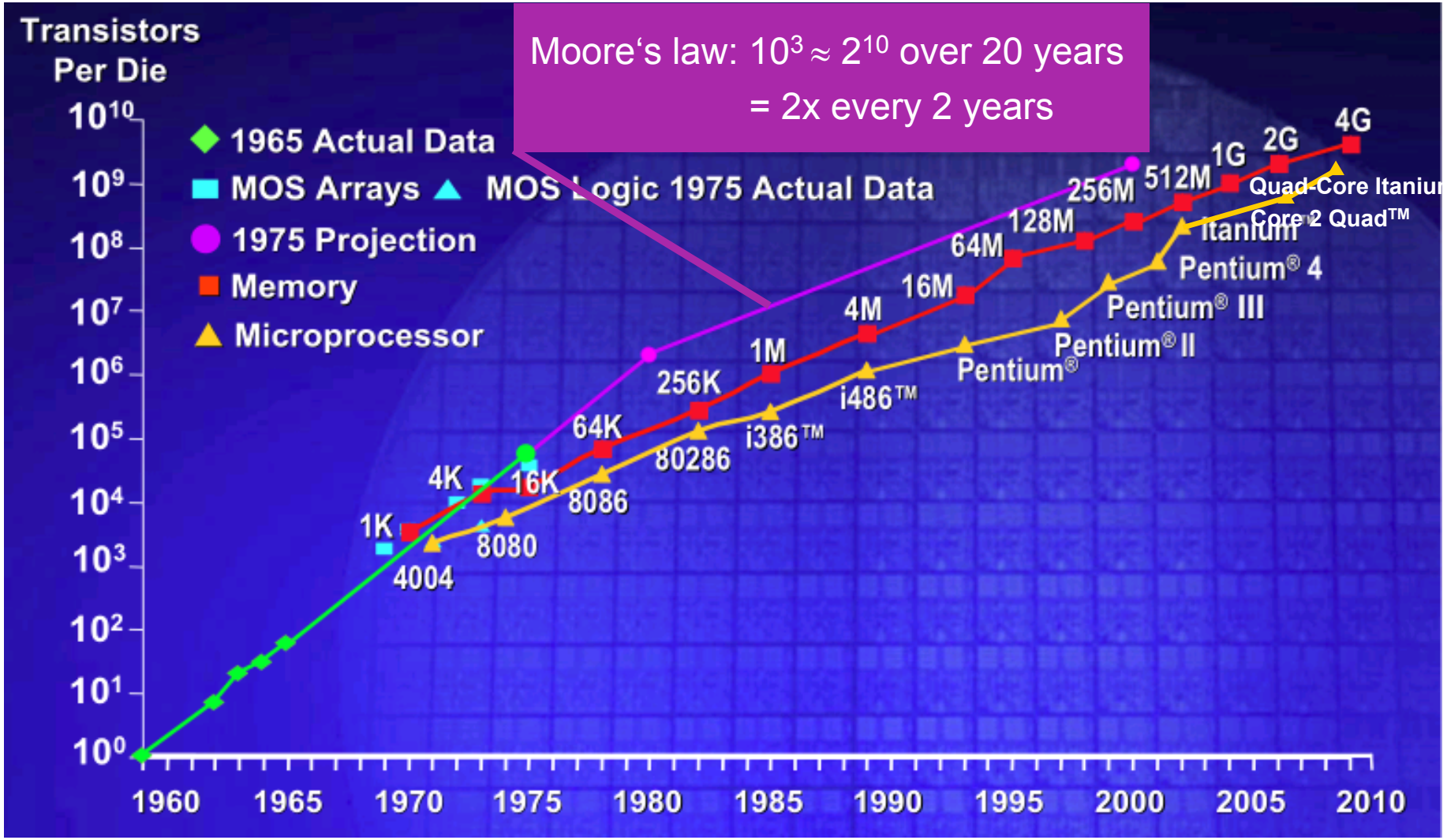
A recent NVIDIA study suggests that power reduction by 7nm CMOS technology is not sufficient to enable exascale computing. Rather, advances in design – from architecture to circuit – have to contribute substantially to power efficiency increase. For portable devices, where forced cooling of integrated circuits is not viable, the resulting power limit also has an impact on the achievable data rate and communication performance. In particular, for ultra high data rates the power limitations will have to be considered during communication system design. The assumptions of the study and how they impact communication system architectures will be discussed. A power estimation methodology for design space exploration will be presented.

Outline

CMOS future and challenges

A new wireless system design paradigm?

Silicon Technology Roadmap



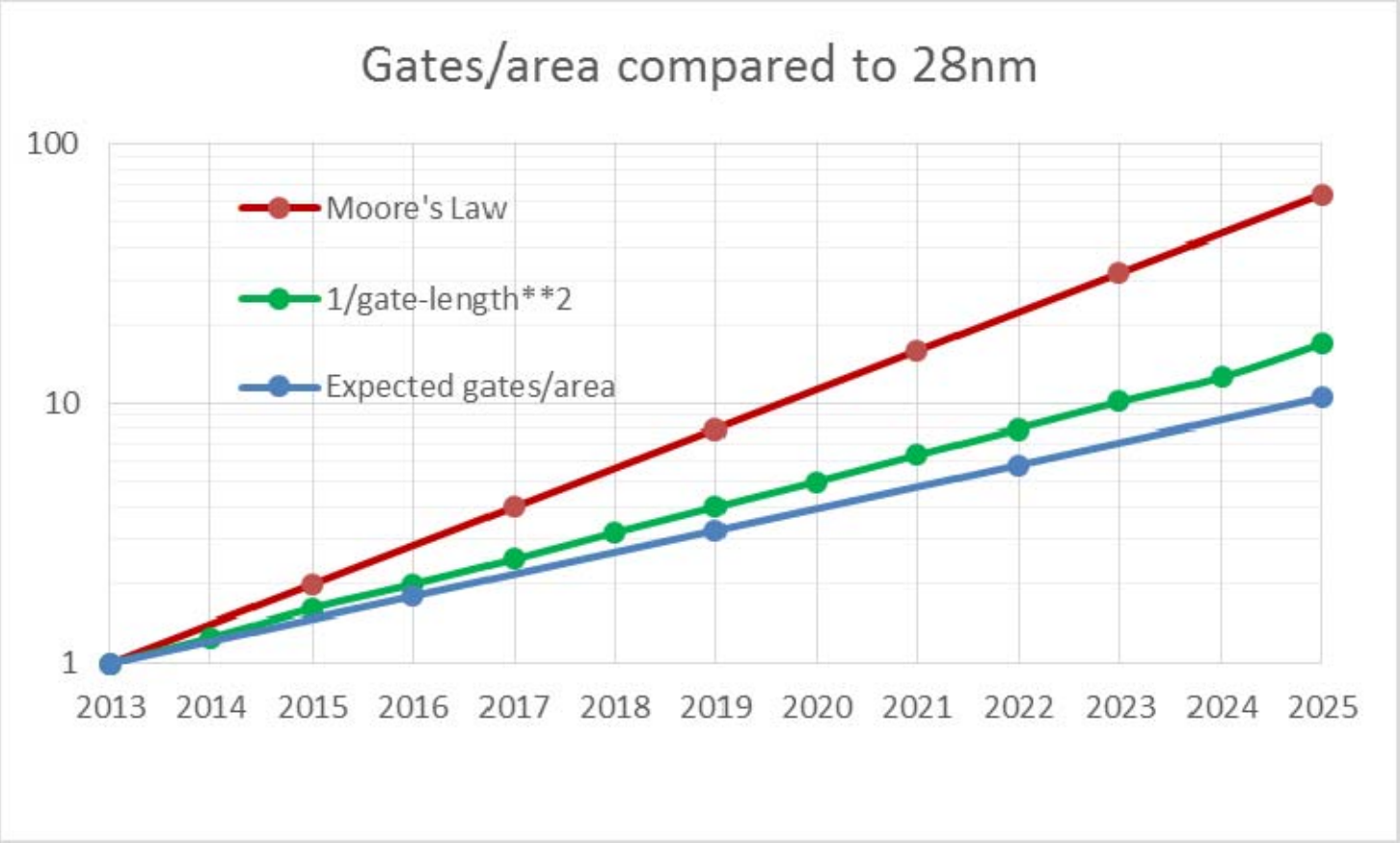
Source: thenextwavefutures.wordpress.com

Silicon technology progress is a key enabler of progress in wireless communication systems

Source: thenextwavefutures.wordpress.com

Moore‘s Law In The Next 10 Years

- Density growth has slowed down significantly (2x only every 3.6 years)



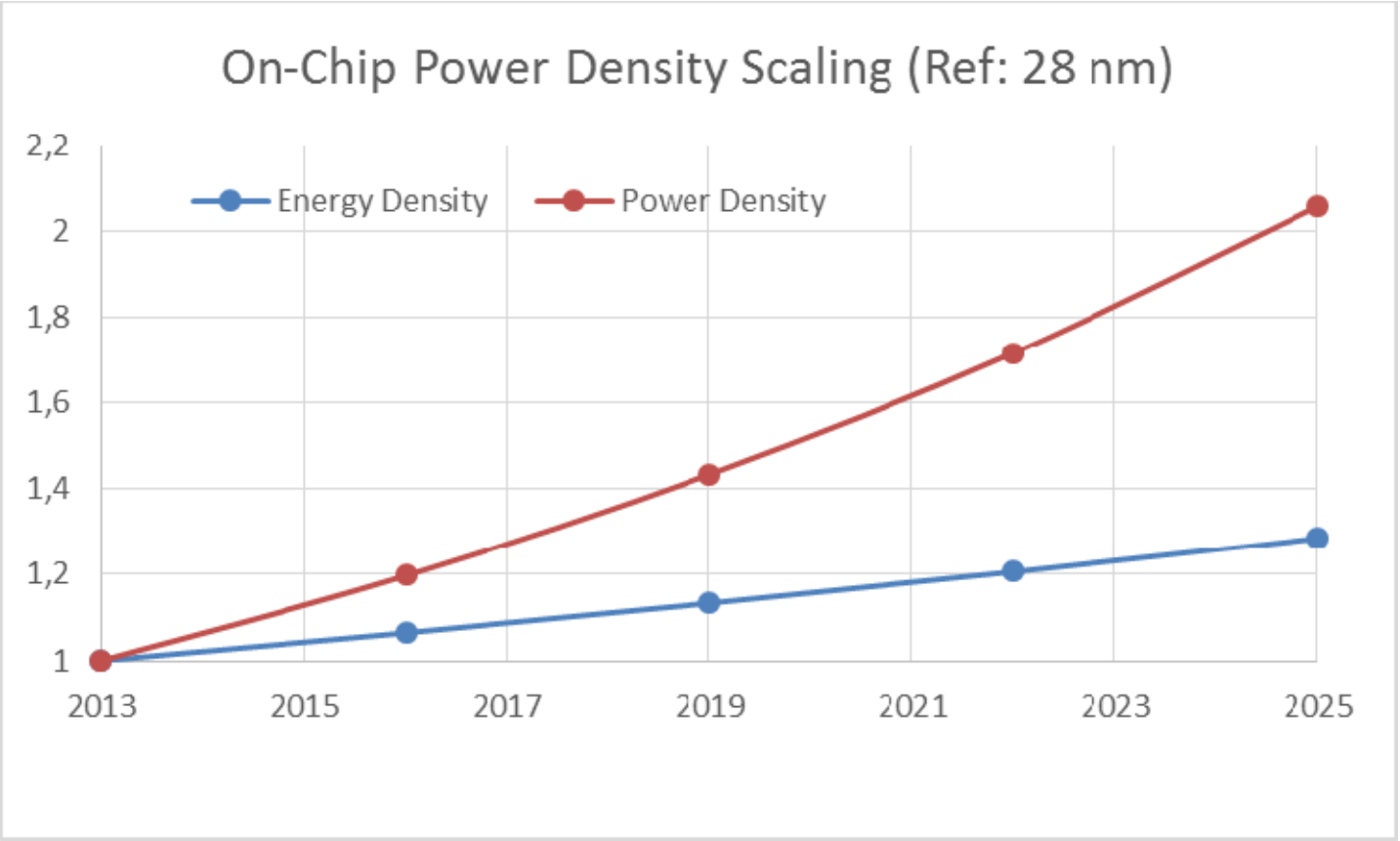
Sources:
1.) ITRS Report 2013
2.) Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

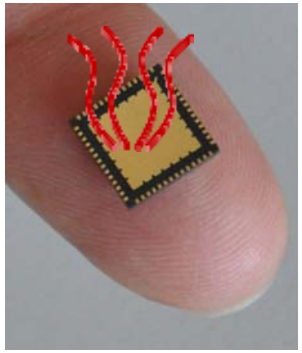
Conclusion #1

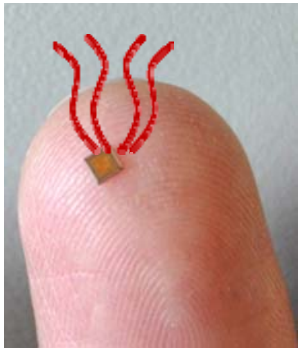
Not more than 10x complexity
on same chip size by 2025

On-Chip (Dynamic) Power Density

- Thermal energy and thermal power (Energy*Clock_Frequency)



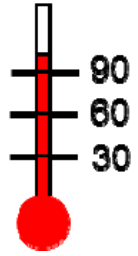




Sources:
1.) ITRS Report 2013
2.) Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

Conclusion #2

Energy efficient design required to avoid doubling of power density (→ heat radiation)



Power is not just an issue for battery operation but, more importantly, for heat dissipation

Mobile Communication Timeline

- Mobile communication has been a key domain for (MP)SoC
- Cellular communication system generations roughly occur per decade (with intermediate improvements)

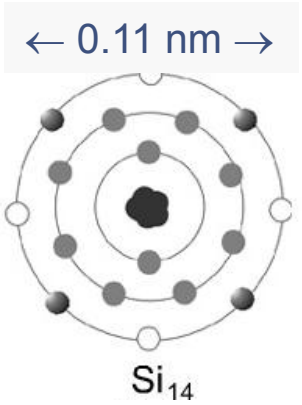
- 2G (GSM) 1990
- 3G (UMTS) 2000
- 4G (LTE) 2010
- 5G expected 2020

⇨*2025 is in the middle of 5G lifetime*

Technologies beyond 2025

- Can we expect scaling much below 5 nm?

Silicon crystallizes in a diamond cubic crystal structure with a lattice spacing of 0.54 nm
→ 5 nm ≈ 10 atoms



2016 Technology Update (revisited)

- Is this the end of progress in wireless communication?

  New wireless system design paradigms,
  e.g. IoT devices combining application and (wireless) communication

- Technologies beyond CMOS („more than Moore“)

  – 3D

  Limited scaling

  – New components

  New computing architectures
  ⇨ new wireless system architectures ?

  – Advanced concepts (Quantum computing, Biocomputing, ...)

  when available (if at all)?

Q: How do we get back to exponential performance scaling?

IEEE Rebooting Computing Initiative

Paolo Gargini @ ITRS/RC Summer Meeting 2015

Rebooting Wireless Communication !

Outline

CMOS future and new challenges

A new wireless system design paradigm?

Example

Transmitting 100 Gb/s and beyond*

- Key ICE contributors: Gaojian Wang, Andreas Bytyn
- Project Partners:
  - Renato Negra, HFE, RWTH Aachen
  - Norbert Wehn, EMS, TU Kaiserslautern
- Project funded by German Science Foundation under a special program

DFG Deutsche Forschungsgemeinschaft

Boundary Conditions for Mobile Devices

- Power consumption limit without forced cooling: order of magnitude is 1W
- For a data rate of 100 Gb/s, processing energy per bit is limited to 1W / 100 Gb/s = 10 pJ
- Based on a NVIDIA study at 28nm: 1-2 Ops/nJ
- Technology-based gain at 7 nm compared to 28nm : 8.17
  ⇨ 8-16 Ops/nJ or *at 100 Gb/s 0.08-0.16 Ops/bit*
- Caveats: Above numbers
  - do not include analog and ADC/DAC
  - are for 2025 technology!
- Frequency ranges of interest are around 60 GHz or beyond 100 GHz
  ⇨ several GHz of bandwidth available (but not necessarily 50 GHz+)

Source: Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

Boundary Conditions

- Power consumption limit without forced cooling: order of magnitude is 1W
- For a data rate of 100 Gb/s, processing energy per bit is limited to 1W / 100 Gb/s = 10 pJ

Research questions:

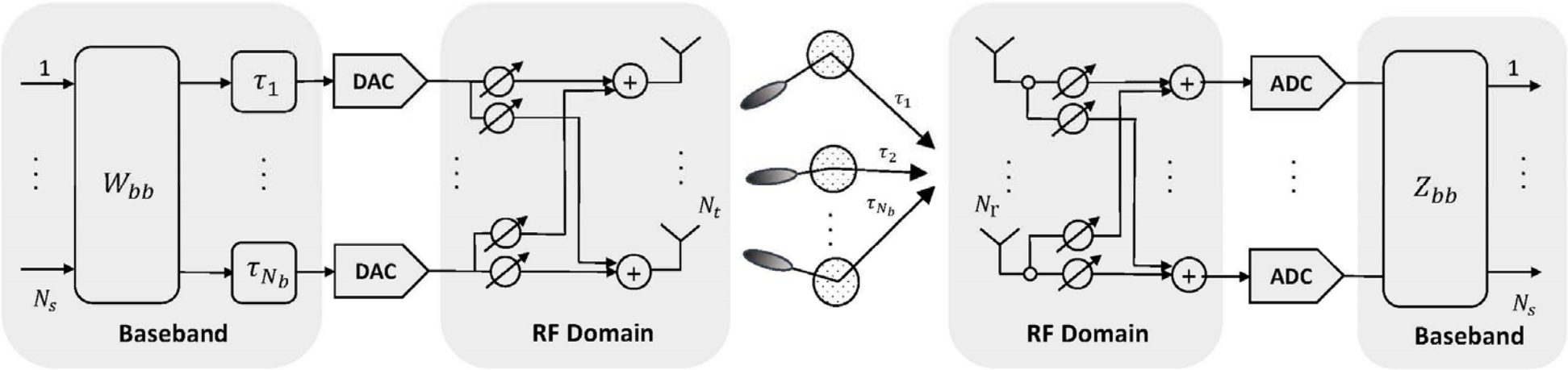
- How many bits/s possible within processing energy constraints?

- Which system architecture (MIMO, modulation, coding, …) enables the maximum information bit rate?

- Frequency ranges of interest are around 60 GHz or beyond 100 GHz
  ⇨ several GHz of bandwidth available (but not necessarily 50 GHz+)

Source: Scaling the Power Wall: A Path to Exascale, multiple authors from NVIDIA, SC14, Int. Conf. for High Performance Computing, Networking, Storage and Analysis

System Architecture Overview



- Consider power in wireless system architecture design phase
- Some conclusions for analyzed system:
  - Do not use OFDM
  - Make channel frequency flat by time-delay compensation
  - Hybrid beamforming to maximize capacity and energy efficiency (analog/digital processing trade off)
  - Optimize number of power-hungry ADCs and DACs

Power Modelling & Energy Budgeting

- Estimate power for all sub-components
  - RF/analogue: analytical model built on state-of-the-art power models
  - Beamforming: approximation based on elementary operations in algorithm
  - Modulation and Coding (M&C): models and simulation of layout netlist