Energy efficiency in 5G networks

Aarne Mämmelä, aarne.mammela@vtt.fi
VTT Technical Research Centre of Finland
IFIP Networking 2015, Toulouse, France, 20.5.2015
Outline

- Introduction
- KPIs and basic resources
- Trends in mobile data traffic
- Trends in power consumption
- Trends of energy efficiency
- Fundamental limits
- Solutions are trade-offs
- Summary
- Bibliography
Introduction
Introduction

- **5G systems** will be deployed in about 2020 (Andrews 2014, Boccardi 2014).
- Technology **evolution** and fundamental **limits** in digital electronics (Keyes 2005, Zhirnov 2013, Markov 2014) will have a direct effect on **all layers**.
- A trend is towards **small cells** (Cox 2008, Kishiyama 2013, Liu 2014) where **circuit power** (computation) will become important in addition to transmission power (communication) leading to **computation-communication trade-off** (Lettieri 1999).
- The speaker’s background is in the **physical layer algorithms**.
Importance of energy (ITU-R 2015)

- Energy is **expensive for operators and users** and its production has **environmental effects**.
- **Battery capacity** is increasing only 1.5x/decade or 4 %/year.
- **Mobile traffic** is increasing exponentially (up to 1000x/decade) and power consumption should be adapted to the traffic load (Blume 2010).
- **Link throughput** requirements are increasing up to 10 Gbit/s.
- **Moore’s law** (energy efficiency 100x/decade) is slowing down and will have thermal noise death by 2020-2030 (now 10x/decade).
- **Cooling efficiency** will not improve significantly, and active cooling is using energy.

Global mobile traffic (Cisco 2015)
Revolutionary technologies for 5G (Andrews 2014, Boccardi 2014)

1. Small cells (increased area spectral efficiency) (Cooper's prediction, Chandrasekhar 2008)
   - Heterogeneous networks, microcells < 1 km, picocells < 100 m, femtocells < 10 m, energy harvesting networks
   - Software-defined network, separate data and control planes (phantom cell, hyper-cellular network, macro-assisted small cell)
   - Cloud radio access networks (centralized baseband), distributed antenna systems
   - Cooperative communication (coordinated multipoint, relaying)

2. Massive MIMO (increased area spectral efficiency)
   - Number of antennas much larger than the number of devices

3. Millimeter waves (larger bandwidth)
   - 30-300 GHz, wavelength 1-10 mm

4. Smart devices and device-centric connectivity
   - Distributed services
   - Device-to-device (D2D) connectivity
   - Local caching
   - Advanced interference rejection

5. Machine-to-machine (M2M) communications
   - Critical (e.g. vehicles) and massive (e.g. sensors) deployments
   - Minimal data rate with very high link reliability virtually all the time
   - Very low latency and real-time operation
   - Long battery life > 10-15 years
Revolutionary technologies for 5G

- Critical infrastructures
- Device-to-device
- Machine-to-machine
- Energy harvesting
- Recycling
- Cloud-access network
- Data plane
- Control plane
- Software-defined network (hyper-cellular network)
- Distributed antenna system
- Small cells, millimeter waves

Aarne Mämmelä, aarne.mammela@vtt.fi
KPIs and basic resources
Key performance indicators (KPIs)

- Peak data rate, user experienced data rate
- Traffic volume density (sum bit rate/km$^2$)
- Number of connections/km$^2$
- End-to-end latency (ms)
- **Battery life** (3 days for smart phones, up to 10-15 years for machine-to-machine systems, NGMN 2015)
- Mobility (km/h)
- Reliability
Basic resources

- **Materials** (silicon, copper, etc.)
- **Energy** (power) – computation-communication trade-off
- **Time** (delay) – diversity (redundancy), multiplexing
- **Frequency** (bandwidth) – diversity (redundancy), multiplexing
- **Space** (area, volume) – diversity (redundancy), multiplexing
- **Capital** (cost) – human work
Trends in mobile data traffic
Global mobile traffic (Cisco 2015)

- Annual growth is 57%, implying 100-fold growth in ten years (some companies expect 1000-fold growth).
- Majority of the traffic will be video traffic (Ericsson 2014)
Bit rates in wireless systems (Fettweis 2012)

- Bit rates at the same link distance have been increasing 100x/decade
- We use the year 2015 as a reference point (after this some saturation expected for constant power consumption)
  - 100 Mbit/s at 100 m
  - 10 Gbit/s at 10 m
  - 100 Gbit/s at 1 m
The bit rates vs. link distance (Cox 2008)

- Bandwidth is decreased with M-ary modulation when increasing M
Trends in power consumption
Trend in power consumption (TWh) (Ericsson 2013, used with permission of Ericsson)

- Data centers will dominate the power consumption in the network due to cloud access networks
- PC, personal computer, CPE, customer premises equipment
Power consumption of various LTE BS types (%) (EARTH D2.3)

- Baseband dominates in small cells because of computation-communication trade-off
  - In small cells the transmission power is reduced and circuit power starts to dominate
- Increasing carrier frequency will increase path loss and transmission power
- Increasing number of antennas will increase antenna gain and decrease transmission power but increase computation (circuit) power
Power consumption in an LTE smart phone (%) (Wang 2014)

- Percentage of power amplifier is reduced at short links (max. power used here)
- Power control also reduces the average power amplifier power
- BT, Bluetooth, GPS, Global Positioning System
Trends in energy efficiency
Saturation of technology has started in about 2000 (Keyes 2005, ITRS 2012, Frantz 2012, Koomey 2011). Landauer limit is a noise limit for the switching energy identical to Shannon limit.
Power consumption vs. computing power
Power consumption vs. bit rate

- **One 16-bit multiplication with hardware** using the technology of the 2020’s
  - 3000 logic gates = $3 \times 10^5 \times$ Landauer limit = 1 fJ
Fundamental limits
Fundamental limits (Markov 2014)

- Conservation of energy – cooling problems
- Entropy law – available energy reduced or kept constant, new energy needed
- Absolute zero – power density of thermal noise depends on absolute temperature (Zhirnov 2003)
- Upper velocity limit – propagation delays, delays of wires on a chip (Ho 2001)
- Uncertainty principle – lowest line width 4 nm (Zhirnov 2013)
- Channel capacity – maximum link spectral efficiency, Shannon limit for received energy/bit
- Energy limit for computation (circuit power) – Landauer limit for energy/irreversible logic operation
- Unprovable theorems – Gödel incompleteness theorem
- Unsolvable problems – Church-Turing conjecture, Turing machine never stops
- Intractable problems – Cook-Levin theorem, exponential complexity, Turing machine stops after an extremely long time
- Unpredictable deterministic systems – chaotic systems
- Maximum speed-up in parallel processing – not linear, Amdahl’s and Gustafson’s laws, energy efficiency in general reduced in parallel processing
- Resolution of optical imaging instruments – Abbe diffraction limit
Systems approach, emphasizing the information and energy flows

- **Nature**
  - Radiation
  - Convection
  - Conduction

- **Usage model**
- **Traffic model**
- **Environment**
- **Use case**

- **Battery**
- **Energy**
  - Materials
  - Time
  - Frequency
  - Space
  - Capital

- **Processing model**
- **Cooling model**

**Energy flow** (energy/(time x area))

**Information flow**

**Performance** (throughput, delay, error rate)
Ultimate computer

- Ultimate computer has an \textbf{energy efficiency} (energy/logic operation) of $100 \times$ Landauer limit.
Solutions are trade-offs
Why trade-offs?

- We are approaching fundamental and practical limits. We must make a trade-off, for example either spectral efficiency or energy efficiency is maximized but not both (Chen 2011, Deng 2013). Valid for uplink (terminals) and downlink (base stations).

- Energy efficiency (bit/J) is reduced at low and high spectral efficiency (bit/s/Hz)
Effect to the whole system must be considered, several methods to be combined (Alagoz 2011, Mittag 2011, Calhoun 2012)
Optimization methods

- **Single-objective (single-criterion) optimization**: either spectral efficiency (bit/s/Hz) or energy efficiency (bits/J) is maximized to obtain better **average performance**
  - Hyper-cellular networks (phantom cells) (Kishiyama 2013, Liu 2014): data plane and control plane are separated (**sleep modes** in small cells used, Blume 2010)

- **Multi-objective (multi-criteria) optimization** (Marler 2004, Chen 2011, Deng 2013)
  - Near fundamental limits approaching **peak performance** needs joint optimization of both spectral efficiency and energy efficiency implying the use of **trade-offs** since the optimum is **not unique**, for example cross-layer design

- Hyper-cellular network (phantom cells)
Trade-offs

- **Performance-energy trade-off** (limited battery capacity)
  - Energy efficiency – spectral efficiency trade-off (Shannon capacity)
  - Energy-space-time trade-off (cooling problems)
  - Communication-computing trade-off (Shannon and Landauer limits)
  - Precision-bandwidth trade-off (limited word length)
  - Analog-digital trade-off (part of precision-bandwidth trade-off)
  - Software-hardware trade-off (performance-flexibility trade-off)
  - Serial processing – parallel processing trade-off (limited clock rate)

- **Performance-delay trade-off** (spatially distributed systems)

- **Communication-control trade-off**
  - Centralized control – distributed control (cognitive radios)
  - Centralized computing – distributed computing (clouds)

- **Order-chaos trade-off** (interacting network elements, transmitter power control)
Trade-offs

Heat (1 W/cm² - 200 W/cm²)

Transmitter

Receiver

Communication (up to 10 Gbit/s)

Digital SW  Digital HW  Analog HW

Communication

Analog HW  Digital HW  Digital SW

Radiation (Shannon limit)

Computing

Energy (Landauer limit)

Computing

Energy (Landauer limit)

Computing

Serial computing

Parallel computing

Aarne Mämmelä, aarne.mammela@vtt.fi
Physical level exists at all layers

<table>
<thead>
<tr>
<th>Service layer</th>
<th>Functional level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural level</td>
</tr>
<tr>
<td></td>
<td>Physical level</td>
</tr>
<tr>
<td>Network layer</td>
<td>Functional level</td>
</tr>
<tr>
<td></td>
<td>Structural level</td>
</tr>
<tr>
<td></td>
<td>Physical level</td>
</tr>
<tr>
<td>Physical layer</td>
<td>Functional level</td>
</tr>
<tr>
<td></td>
<td>Structural level</td>
</tr>
<tr>
<td></td>
<td>Physical level</td>
</tr>
</tbody>
</table>

- Each OSI layer includes **three description levels**
  - **Functional level** (describes what the layer should do)
  - **Structural (architectural) level** (describes what parts are needed and how they are connected)
  - **Physical level** (physical implementation of the parts, inc. processors, memory, etc.)
Summary
Summary

- Future designs will be resource-limited and trade-offs are needed for peak performance.
- Moore's exponential law has been slowing down since about 2000 and it will eventually undergo a thermal noise death in about 2020-2030.
- The switching energy of CMOS electronics will approach the level of 100 x thermal noise spectral density.
- Future networks are expected to carry 1000 times more mobile data in ten years, but the energy efficiency is improving only 10 times in ten years.
- The technology trends and fundamental limits in the digital electronics will have a direct effect on all layers.
- Networks do not only consume energy in transmission in power amplifiers but also in computation (circuit power) in algorithms and protocols.
- Computation energy will be more important in small cells that are expected to be used in places where user density is high.
- Although peak energy efficiency will saturate, the average energy efficiency may still grow (consumed energy should be proportional to the number of transmitted bits using sleep modes).
Bibliography
Bibliography

- E. Björnson et al., “Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?” IEEE Transactions on Wireless Communications, to be published.
Bibliography

- Ericsson Mobility Report, June 2014.
Bibliography

- B. Wang et al., "End-to-end power estimation for heterogeneous cellular LTE SoCs in early design phases," in Proc. PATMOS’14.
TECHNOLOGY FOR BUSINESS