Baseband Pooling in 4G Cellular Networks

Thomas Werthmann and Heidrun Grob-Lipski
thomas.werthmann@ikr.uni-stuttgart.de,
heidrun.grob-lipski@alcatel-lucent.com

Alcatel·Lucent

Universität Stuttgart
Institute of Communication Networks
and Computer Engineering (IKR)
Prof. Dr.-Ing. Andreas Kirstädter

Outline

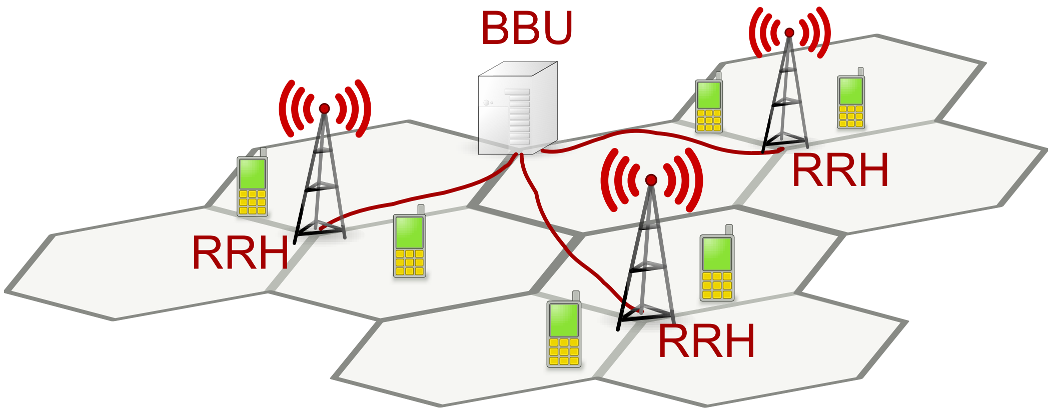
Baseband Unit Pooling

Idealized Evaluation of Sum Processing Effort

Worst-Case and Optimized Placement of UE Stacks

Optimized Placement of Whole Cells

Baseband Unit Pooling



LTE Cloud-RAN Architecture

- Remote Radio Head (RRH)
  - located at previous eNodeB locations
  - consists of antenna, power amplifier, and AD/DA converters
- Baseband Unit (BBU)
  - located at a central office
  - performs baseband and higher layer computation
  - consists of ASICs, FPGAs, DSPs, general-purpose processors

Why Pool BBU Resources?

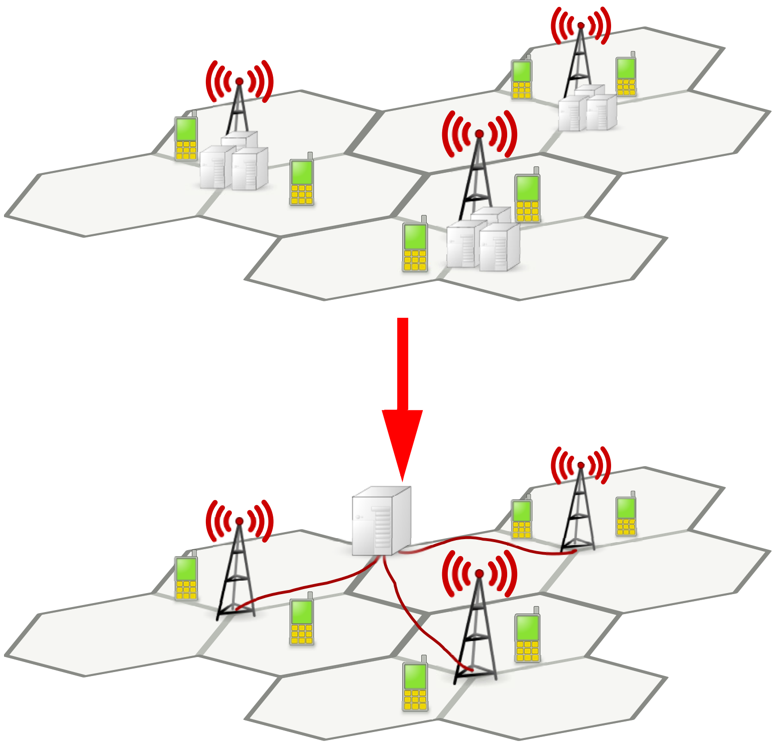
- Improved support of LTE-advanced features, e.g. easier implementation of CoMP mechanisms
- More efficient maintenance (OPEX)
- **Hardware pooling gains (CAPEX)**

Hardware Pooling Gains

BBU Compute Resource Dimensioning

- Separate BBU for each sectors
  - should be able to achieve LTE peak capacity
  - → processing resources dimensioned for peak load
  - → overdimensioned
    (most of the time less resources required)

- BBU Pool serving multiple sectors
  - accept that the system capacity
    can be limited by processing resources
  - → processing resources dimensioned
    according to demand probability
  - → a pooling gain can be realized



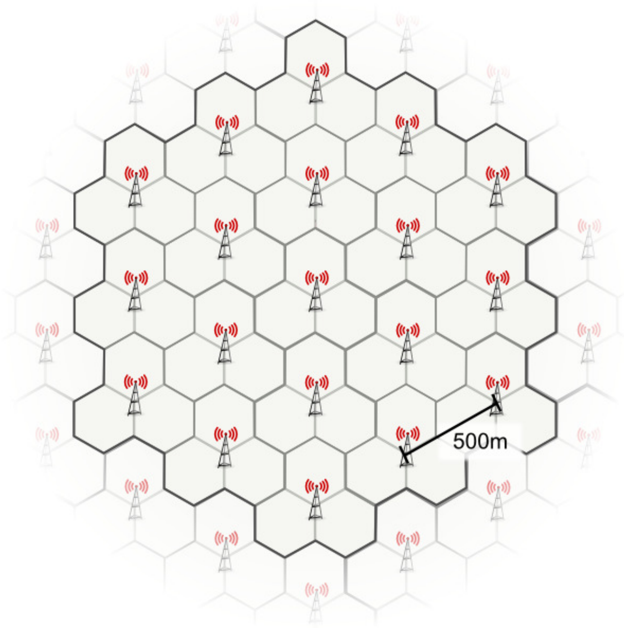
Research Questions

- How large is this pooling gain?
- How to organize the resources in the central BBU pool?

Evaluation Scenario (1/3)

LTE Network Model

- 3GPP compliant LTE Rel. 8 model
- 57 macro cells
- Uniform user distribution (in paper also non-uniform)
- No mobility, but new coordinates for each download



Further details: [Werthmann 2013]

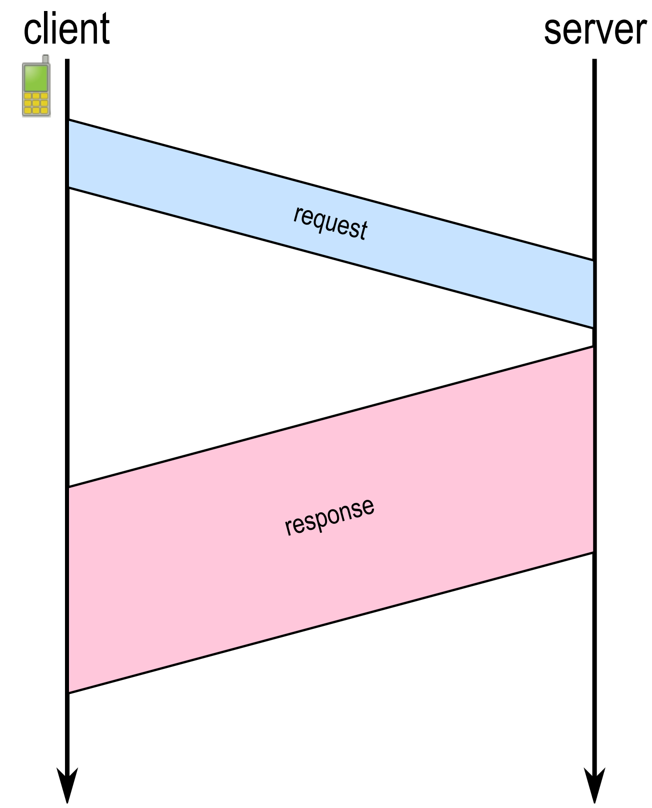
Evaluation Scenario (2/3)

LTE Network Model

- 3GPP compliant LTE Rel. 8 model
- 57 macro cells
- Uniform user distribution (in paper also non-uniform)
- No mobility, but new coordinates for each download

Traffic Model

- Downlink web traffic modelled on application layer
- Object sizes according to fixed network measurement [Hernandez-Campos 2004]



Further details: [Werthmann 2013]

Evaluation Scenario (3/3)

LTE Network Model

- 3GPP compliant LTE Rel. 8 model
- 57 macro cells
- Uniform user distribution (in paper also non-uniform)
- No mobility, but new coordinates for each download

Traffic Model

- Downlink web traffic modelled on application layer
- Object sizes according to fixed network measurement [Hernandez-Campos 2004]

$$P = \left(3A + A^2 + \frac{MCL}{3}\right) \cdot \frac{R}{10}$$

with

- P processing effort
- A number antennas
- M modulation bits
- C FEC code rate
- L num. of MIMO layers
- R number of PRBs

Processing Effort Model

- Processing effort of Phy layer for each allocated physical resource block (PRB)
- Scales with occupied PRBs, modulation and coding scheme (MCS), and MIMO mode [Desset 2012]

Further details: [Werthmann 2013]

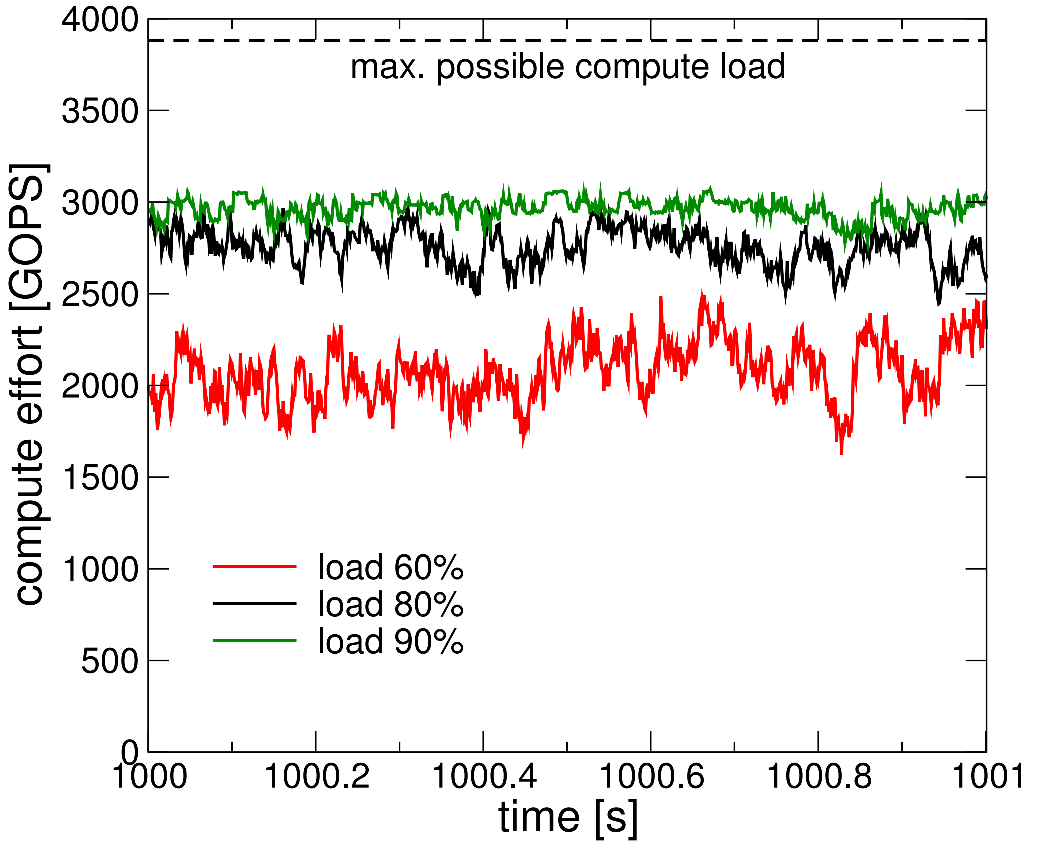
Sum Processing Effort (1/2)

Study

Sum of all required processing resources for each TTI

Results

- Fluctuations caused by traffic model, scheduling, and channel quality
- Even at high system load, the max. possible compute load is not reached
- Larger gains can be achieved with non-uniform user distributions (not shown here)



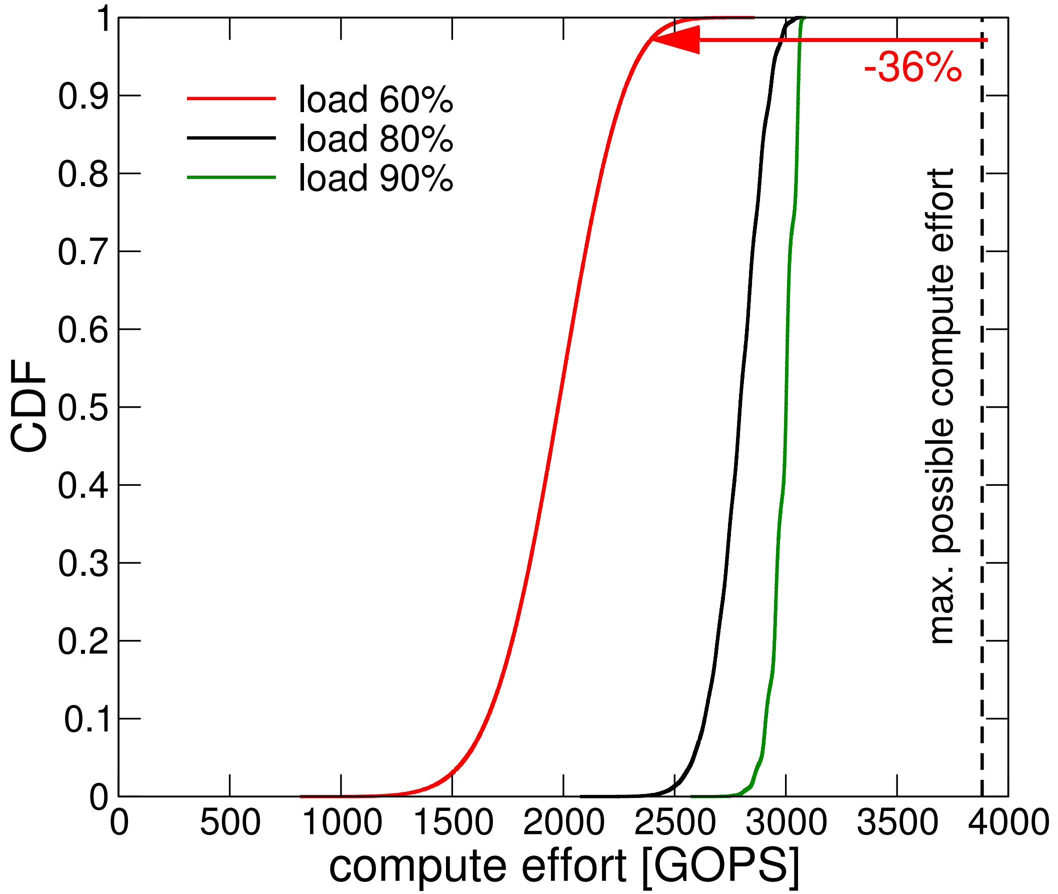
Sum Processing Effort (2/2)

Study

Sum of all required processing resources for each TTI

Results

- Fluctuations caused by traffic model, scheduling, and channel quality
- Even at high system load, the max. possible compute load is not reached
- Larger gains can be achieved with non-uniform user distributions (not shown here)
- All following evaluations are simulated with 60% load
- Pooling gain derived from 99%tile



Placement of Processing Load on Processors

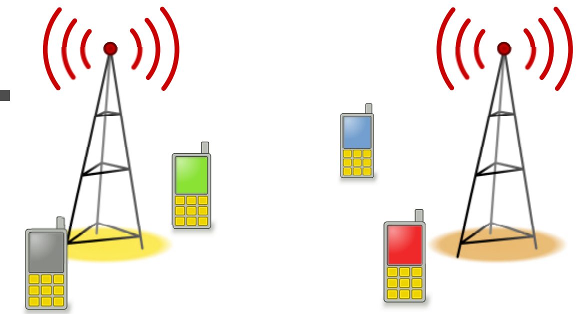
Previous Evaluation [Werthmann 2013]

- Just evaluated sum of processing requirements
- Corresponds to one large homogeneous processor

Following Evaluations

- Placement on small discrete processors
- Comparison of different placement strategies

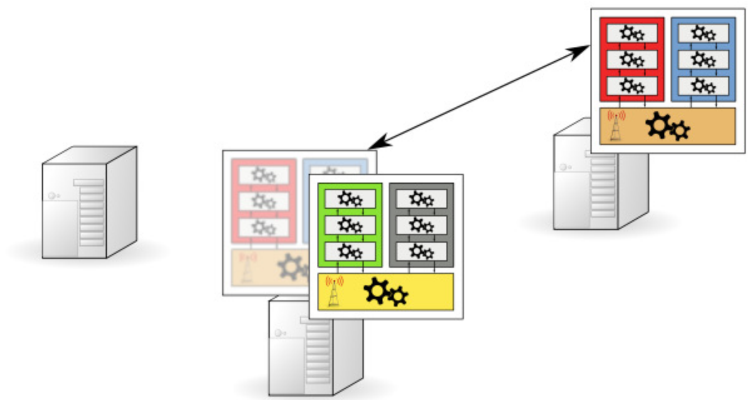
Level of Virtualization



Whole Cells

Combined virtualization of UE Stacks and cell functions

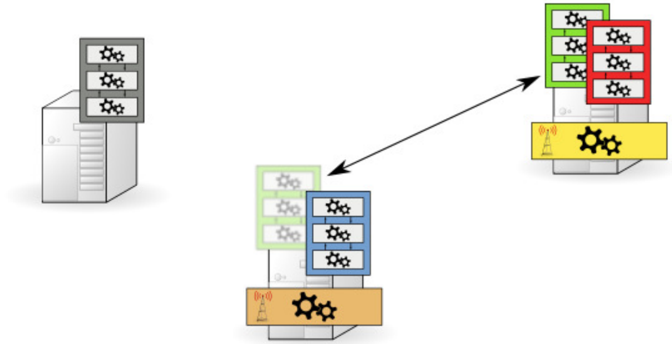
$\rightarrow$ Low granularity



UE Stacks

Separate virtualization of UE Stacks and cell functions

$\rightarrow$ High granularity



Note: Compute effort for cell functions and communication effort not evaluated here

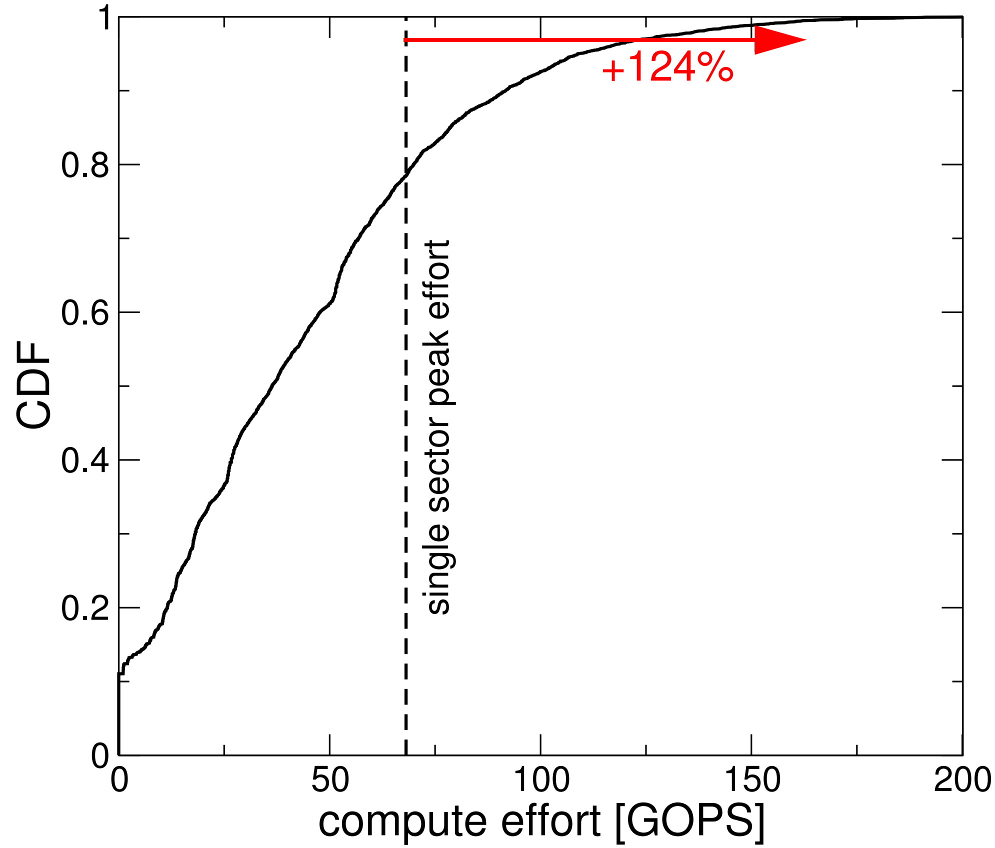
Worst-Case Placement of UE Stacks

Study

- Aggregated load of randomly selected 1/57th of the UEs
  → Same average load as one sector

Results

- No mutual restriction of air interface resources
  → High variation
- Peak processor capacity of a single sector is exceeded in 20% of the TTIs
- 124% Multiplex loss (reference capacity to 99%tile of required capacity)



Worst-Case placement strategy → lower bound of multiplexing gain

Optimized Placement of UE Stacks (1/3)

Assumptions

- Processors of predefined size
- Idealized communication between processors
- Instantaneous movement of UEs between compute units

Optimization problem

Every TTI, place all active UEs on compute units,
so that the number of used compute units is minimized

$\rightarrow$ NP-hard binpacking problem

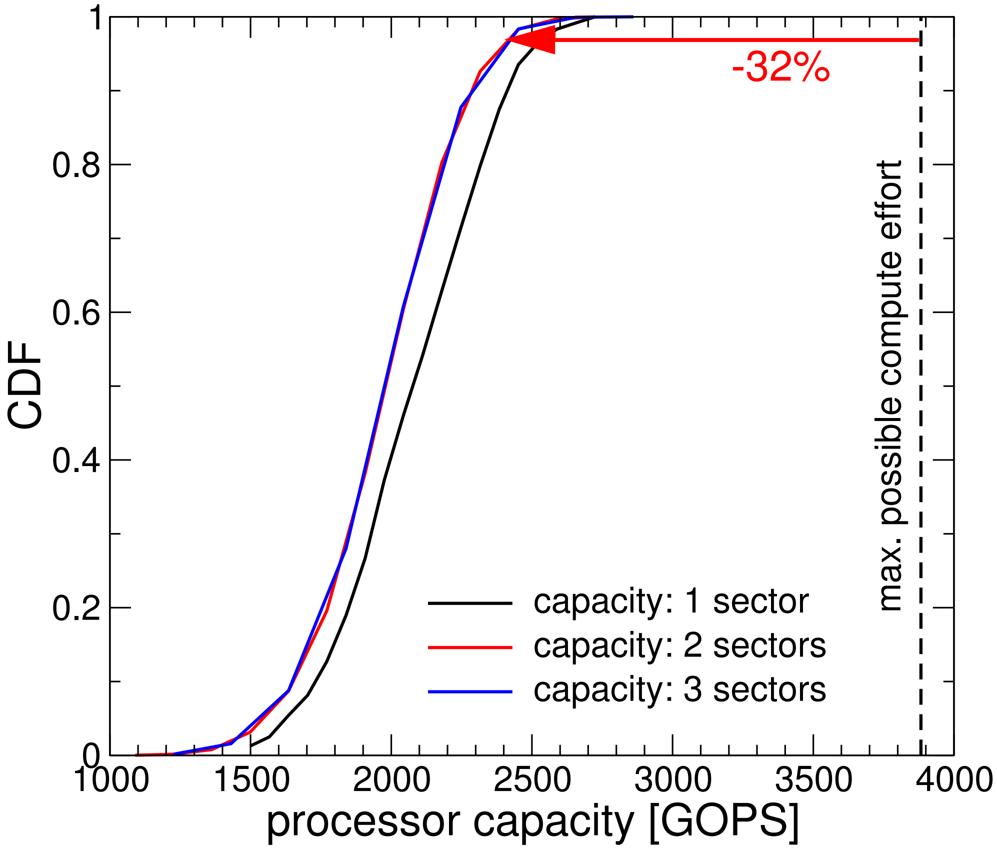
Optimized Placement of UE Stacks (2/3)

Study

- Processor capacities are multiples of single sector peak effort (68 GOPS)
- Placement of all **active UEs** on the processors
- Output: Number of required processors per TTI

Results

- Multiplexing gain instead of loss
- Gain close to previous idealized studies (there: 36%)
- Small increase of the multiplexing gain for larger processors
- Lower gain for small processors caused by off-cuts ("Verschnitt")



Ideal placement strategy → upper bound of multiplexing gain

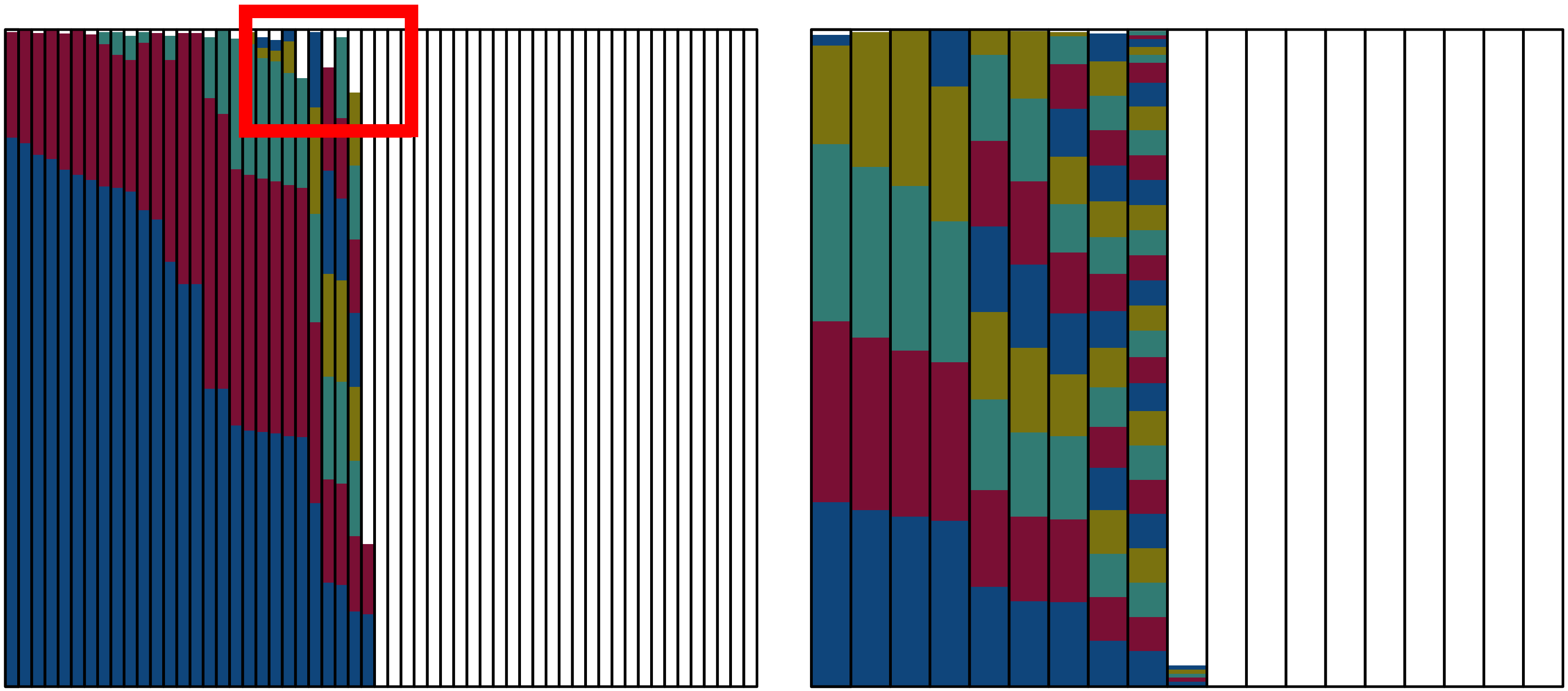
Optimized Placement of UE Stacks (3/3)

Example

Placement of UE-Stacks for a single TTI on processors of different size

57 processors with single-sector capacity

19 processors with three-sector capacity



→ Larger off-cuts occur with smaller processors

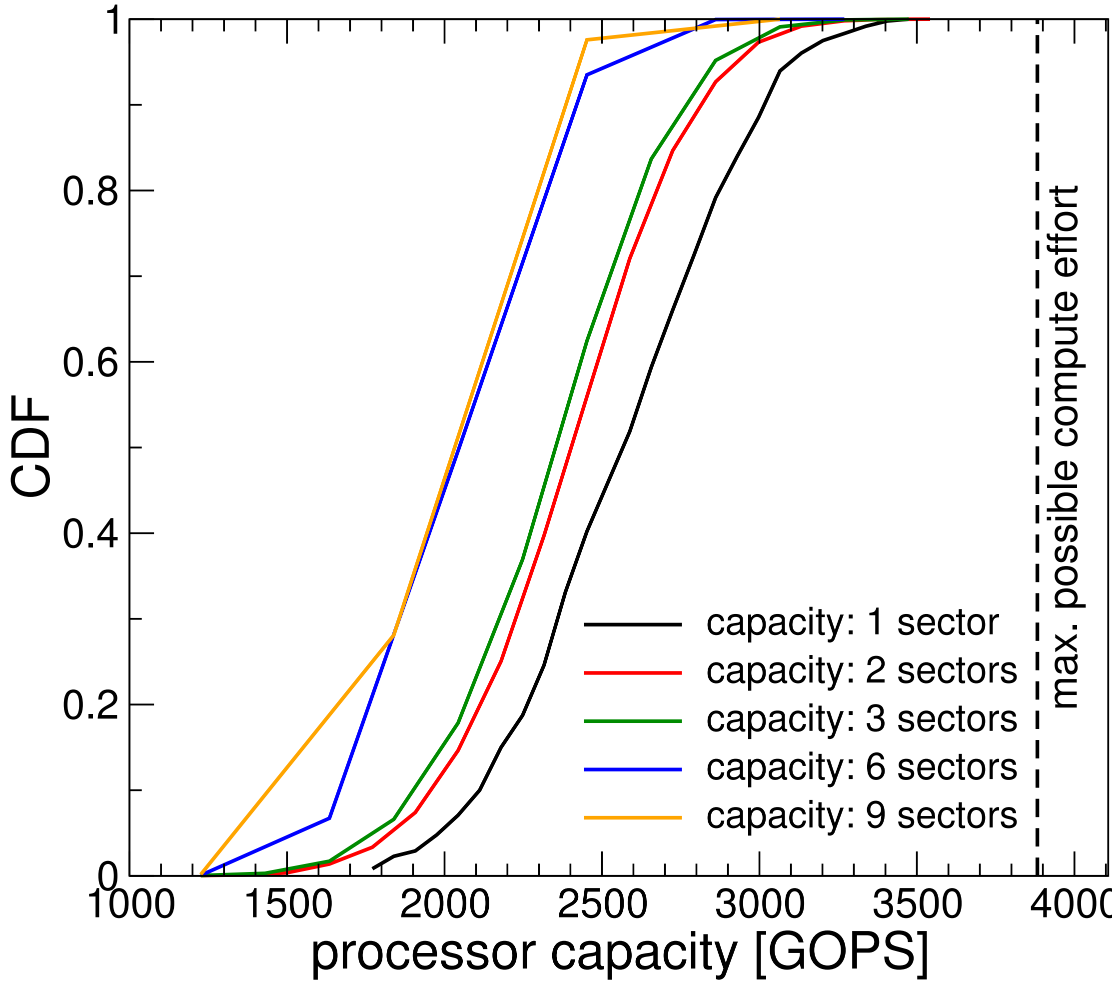
Optimized Placement of Whole Cells

Study

- Predefined processor capacities
- Placement of all **active cells** on the processors (sum effort for all UEs of a cell)
- Output: Number of required processors (per TTI)

Results (compared to UE virt.)

- Lower granularity
  → lower multiplexing gain for small processors
- Similar multiplexing gain reached only with processors with a capacity >= 6



→ **Virtualization of whole cells requires larger processors to avoid off-cuts**

Conclusion and Next Steps

Conclusion

Evaluated theoretical savings in hardware required for physical layer computation

- Significant difference between random placement (124% loss) and optimized placement (32 % gain)
- Optimized placement achieves multiplexing gain close to that achieved in idealized evaluations

→ Placement strategy has to be selected carefully

Evaluated packing off-cut for two levels of UE virtualization

- Virtualizing UE stacks is complex, but provides more flexibility
- Hardware savings are similar for large processors

→ Relation of processor size and job size has to be considered to avoid off-cut

Next Steps

- Evaluate realizable placement strategies
- Consider costs introduced by the additional complexity

Literature

[Hernandez-Campos 2004]
Felix Hernandez-Campos, J.S. Marronb, Gennady Samorodnitsky, F.D. Smith: Variable heavy tails in Internet traffic, Performance Evaluation 2004

[Desset 2012] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M.J. Gonzalez, H. Klessig, I. Godor, M. Olsson, M.A. Imran, A. Ambrosy, O. Blume: Flexible power modeling of LTE base stations, Wireless Communications and Networking Conference (WCNC) 2012

[Werthmann 2013]
Werthmann, Grob-Lipski, Proebster: Multiplexing Gains Achieved in Pools of Baseband Computation Units in 4G Cellular Networks. PIMRC 2013, London