UCN'16

Workshop on Future challenges in User-Centric Networks

Antibes Juan-les-Pins, France, June 14, 2016

EURECOM

Sophia Antipolis

When RAN meets Cloud and SDN

Navid Nikaein

Communication System, Eurecom

FLEx

FIRE LTE TESTBEDS FOR OPEN EXPERIMENTATION

Coherent

MCN

Mobile Cloud Networking

5G PPP The 5G Infrastructure Public Private Partnership

European Commission

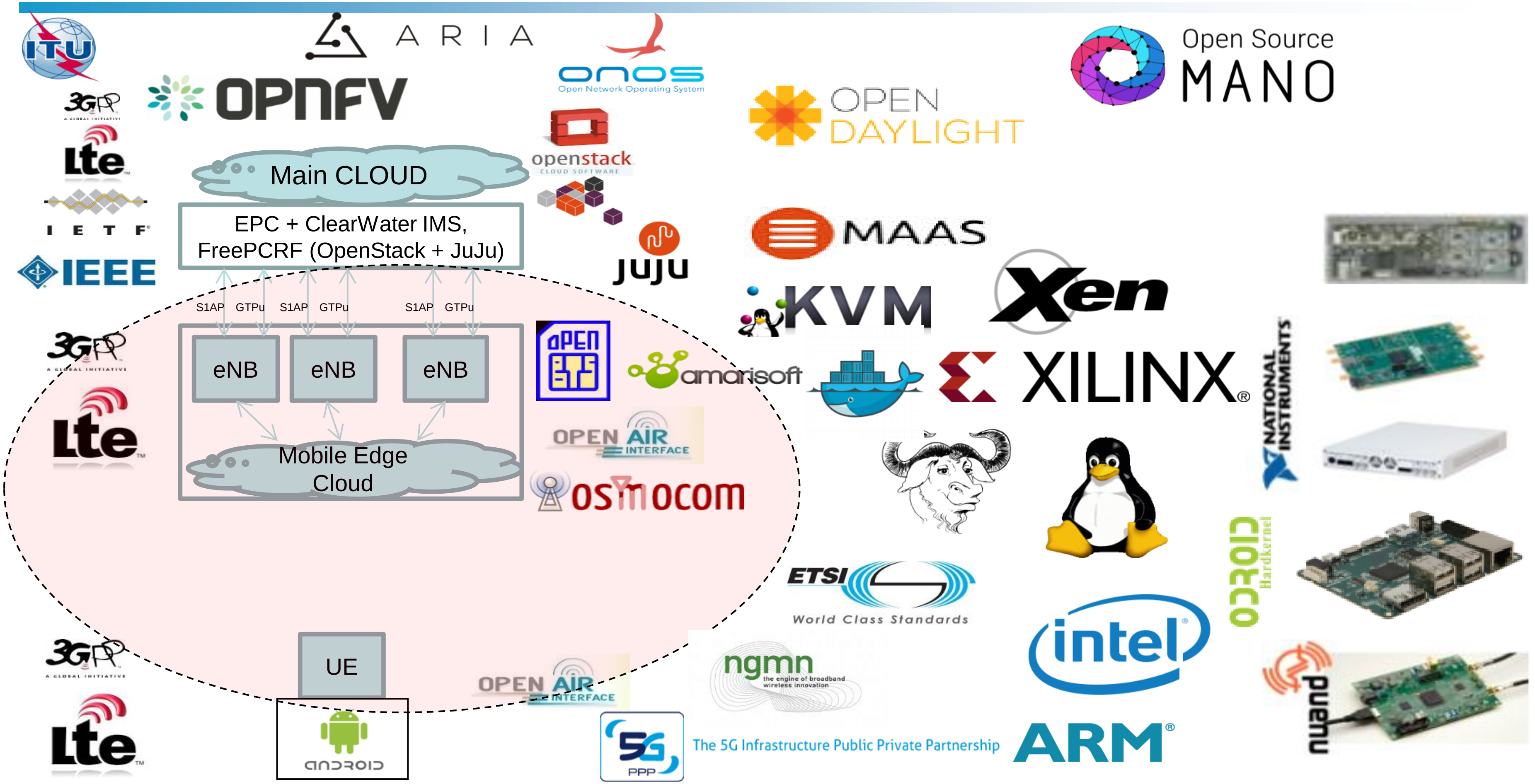
OPEN AIR INTERFACE

16th June, 2016

Economics of mobile are changing

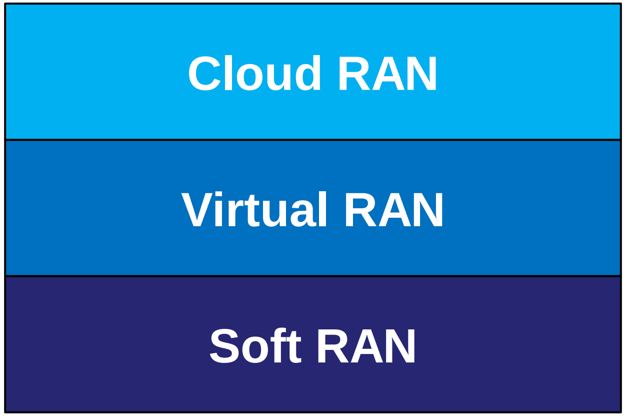
- **Softwarization and Commoditization**
  - Software implementation of network functions on top of GPP with no or little dependency on a dedicated hardware
    - Full GPP vs. accelerated vs. system-on-chip
  - Programmable RF
- **Virtualization and Cloudification**
  - Execution of network functions on top of virtualized computing, storage, and networking resources controlled by a cloud OS.
  - Share I/O resources among multiple guests
- **Emergence of rich ecosystem and opensource for telecom**
  - NFV, SDN and MEC
  - Open APIs and standardized I/F

Ecosystems and Activities



Soft RAN

- **4G Feasible on General Purpose Processors (x86)**
- **An eNB is approximately 1-2 x86 cores on Gen 3 Xeon silicon**
  - Perhaps more power efficient solutions from TI, Freescale or Qualcomm
  - But: lose commodity software environment and common HW platform to high-layer protocols and cloud



Soft-RAN Processing Budget

eNB Rx stats (1subframe)

- OFDM demod : 109.695927 us
- ULSCH demod: 198.603526 us
- ULSCH Decoding : 624.602407 us

➔ 931 us (<1 core)

eNB Tx stats (1 subframe)

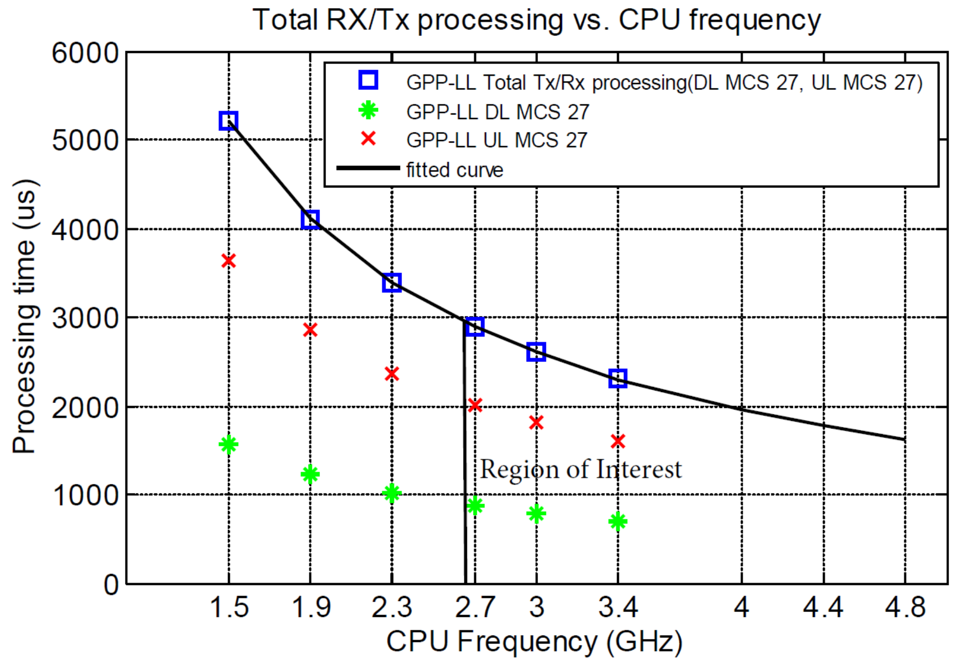
- OFDM mod : 108.308182 us
- DLSCH mod : 176.487999 us
- DLSCH scrambling : 123.744984 us
- DLSCH encoding : 323.395231 us

➔ 730 us (< 1core)

- **Efficient base band unit is challenging**
- **With AVX2 (256-bit SIMD), turbo decoding and FFT processing will be exactly twice as fast**
  - <1 core per eNB
  - .4 core per eNB without turbo en/decoder ← can this be exploited efficiently with HW acceleration? (Solution adopted in China Mobile CRAN project, offload of TC on Altera FPGA)
- **Configuration**
  - gcc 4.7.3, x86-64 (3 GHz Xeon E5-2690),
  - 20 MHz bandwidth (UL mcs16 – 16QAM, DL mcs 27 – 64QAM, transmission mode 1 - SISO)
  - 1000 frames, AWGN channel

Soft-RAN
Processing Budget for Peak Rate

- **Processing time reduces with the increase of CPU Freq.**
- **min CPU Freq is 2.7GHz**
  - HARQ deadline
- **$T_{subframe} = \alpha / x$,**
  - $\alpha = 8000$
  - x is the CPU freq GHZ



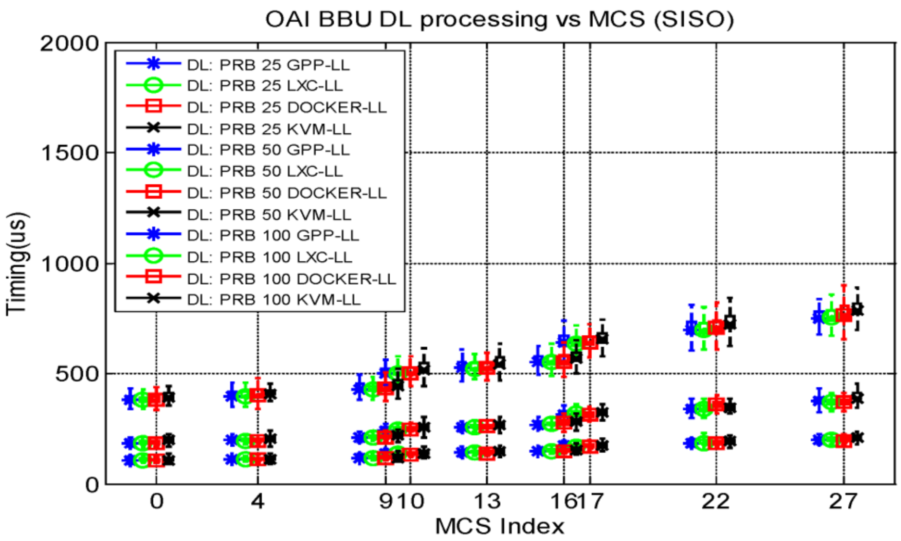
- **Note: FDD LTE HARQ requires a round trip time (RTT) of 8ms**
  - $Tx + RX \leq Tharq/2 - (acquisation + transport + offset) \approx 3ms$
  - ~2ms RX and 1ms TX (can't be fully parallelized)

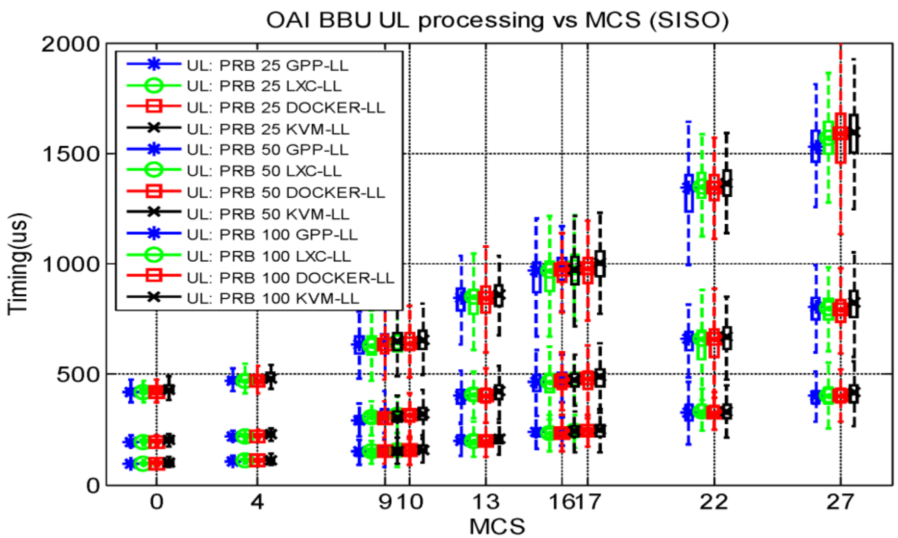
Soft RAN Considerations

- **Key Consideration to meet the deadlines (SF, protocol)**
  - Real-time OS (linux with deadline scheduler) and optimized BIOS
    - Problem: OS scheduler latency (kernel is not pre-emptible)
  - Real-time data acquisition to PC
  - SIMD optimized integer DSP (SSE4, AVX2)
  - Parallelism (SMP)
  - x86-64
    - more efficient for Turbo decoding because of the number of available registers is doubled
- **Remove bottlenecks with**
  - hardware accelerators or hybrid CPUs
    - Turbo decoders (easily offloaded to FPGA-based accelerators), FFT, PDCP (de)enryption
  - GPUs or Xeon PHY-type devices
    - Perhaps interesting for Turbo-decoders and encoders than FFT
    - Main issue in both FPGA/GPU offloading
      - High-speed low-latency bus between CPU memory and external processing units

Virtual-RAN Processing Budget for Peak Rate

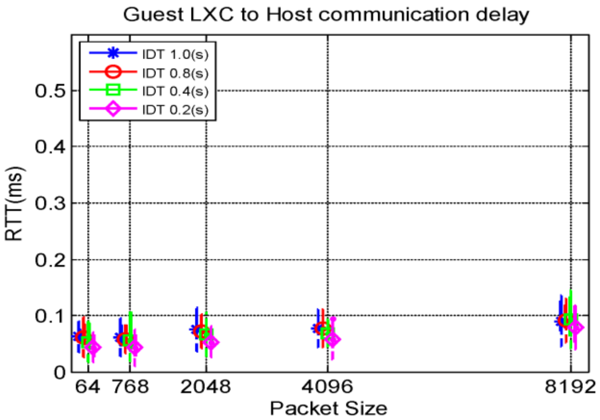
- **DL and UL BBU processing load for various MCS, PRB, and virtualization flavor**
  - Comparable BBU Processing time

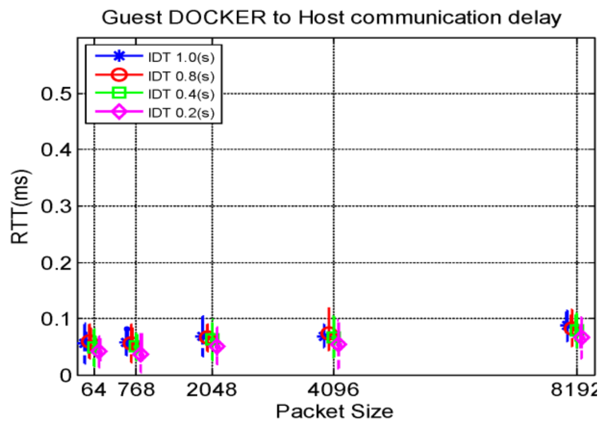


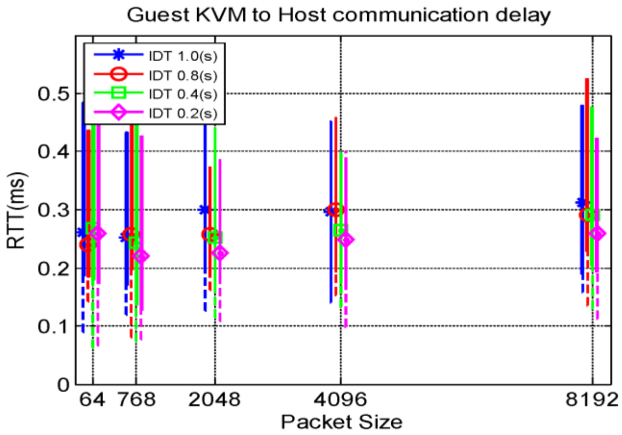


Virtual-RAN
Additional Consideration

- **I/O access delay**
  - RF, ETH, and HW accelerator
  - RF Passthrough vs Hardware virtualization (and sharing)
  - Delay and jitter requirement on the fronthaul network
- **Limitation of the guest-only network data rate**



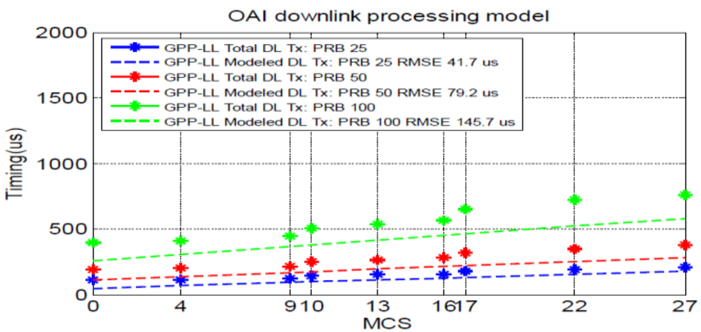


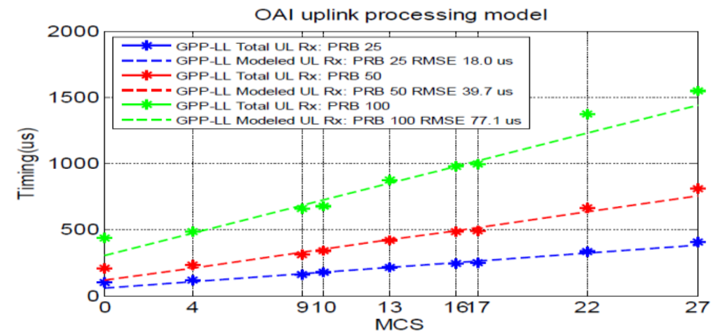


EURECOM

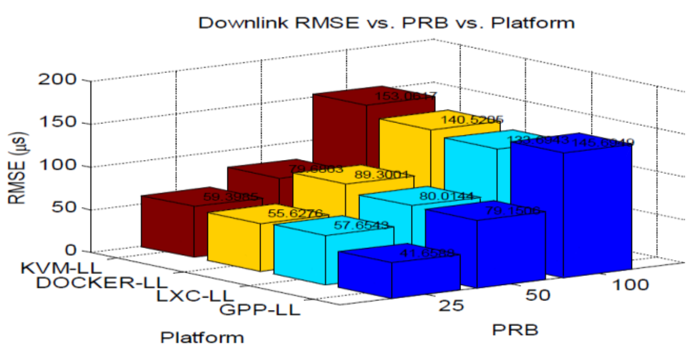
Modelling Processing Budget

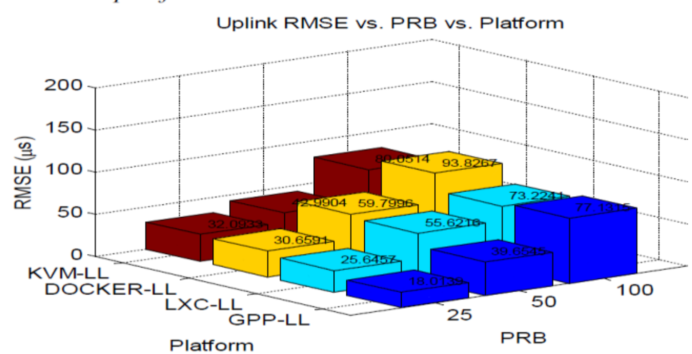
- $T_{subframe} = c[x] + p[w] + u_s(x,y) + u_r(x), where$
- $u_s(x,y) = a(x).y + b(x), and\ x \in PRB, y \in MCS,\ w \in VE$





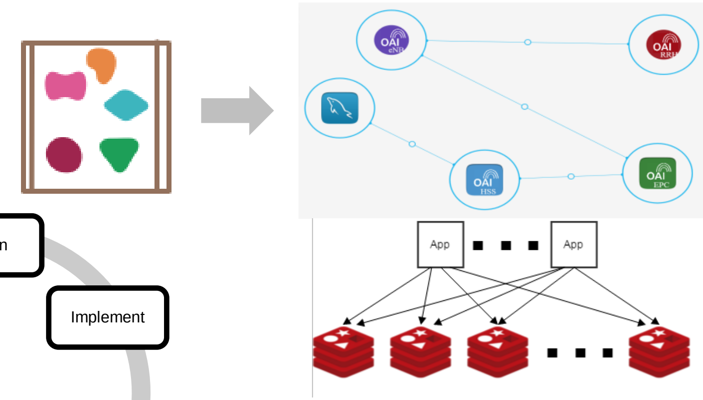
(a) *Fitted curves for GPP-LL platform*

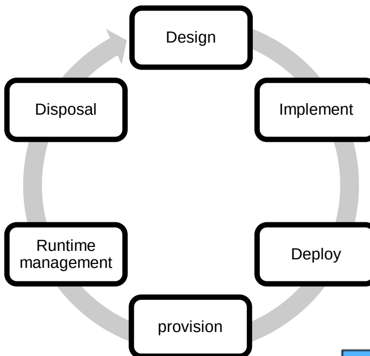


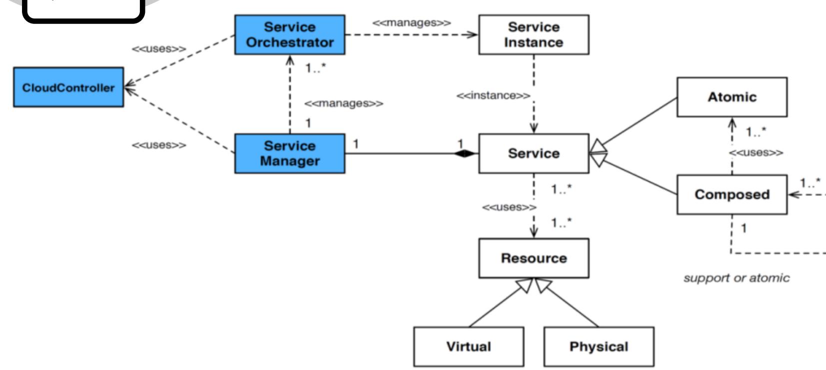


Cloud-Native RAN

- **Mircoservice Architecture along with NFV**
  - Flexible Functional split
  - Move form monolitic to a composed and metered service
  - Stateless, composable, reusable
- **Scalability**
  - Scale in and out, pay-as-you-go
- **Reliability**
  - Redundancy and stateless
- **Multitenancy**
  - Share the resources
  - (spectrum, radio, and infrastructure)
- **Placement**
  - Optimize the cost and performance
  - Supported Hardware, in particular for RAN
- **Realtime edge services**
  - Direct access to the radio information



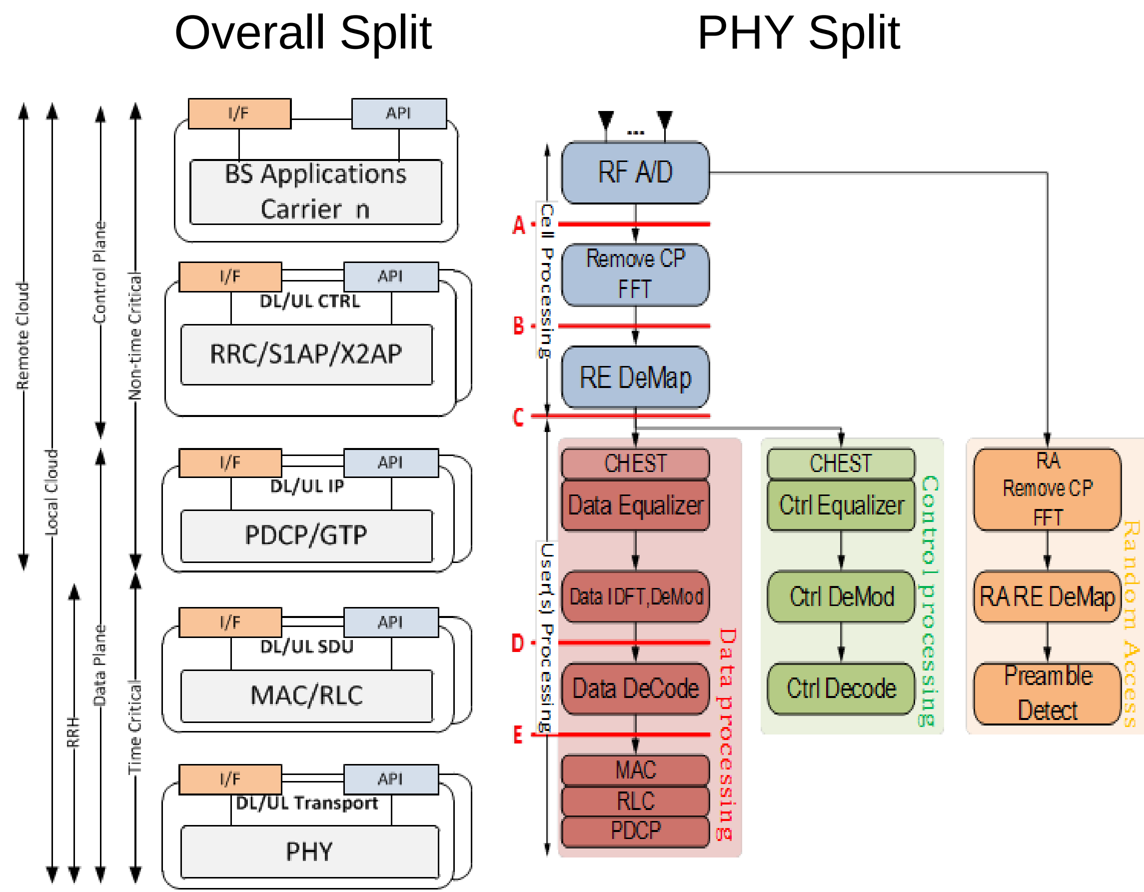




EURECOM

Cloud-native RAN
Where to split?

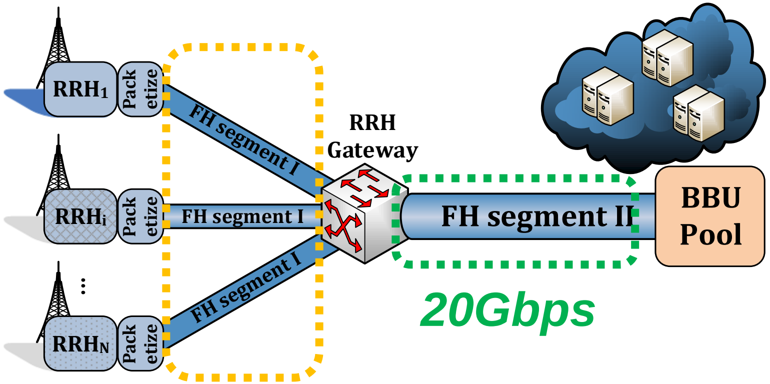
- **RRC and MME Placement**
- **PDCP as a convergent layer**
- **PHY$_{user}$ as a variable**
  - W and W/O MAC/RLC
- **Allow split across RRH, local, and remote cloud**
- **I/F**
  - Orchestration logic
- **API**
  - Controller logic



Cloud-native RAN
Where to split?

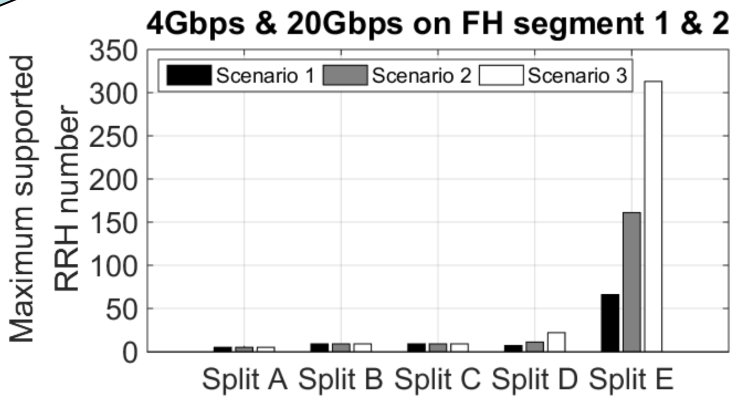
- **Derive maximum supported RRHs based on achievable peak-rate**

Scenario	1	2	3
Bandwidth	20 MHz		
Oversampling Ratio	1		
Rx Antennas	4		
Cyclic prefix length	Normal		
MIMO	4 Layer		
PUCCH RB	4		
SRS BW Config	7		
SRS SF Config	9		
Control Overhead	4.3%		
RA Config	0		
RA Overhead	0.3%		
Modulation	64 QAM	16 QAM	QPSK
TBS index	26	16	9
Time sample bitwidth	16		
Frequency sample bitwidth	16		
LLR bitwidth	8		



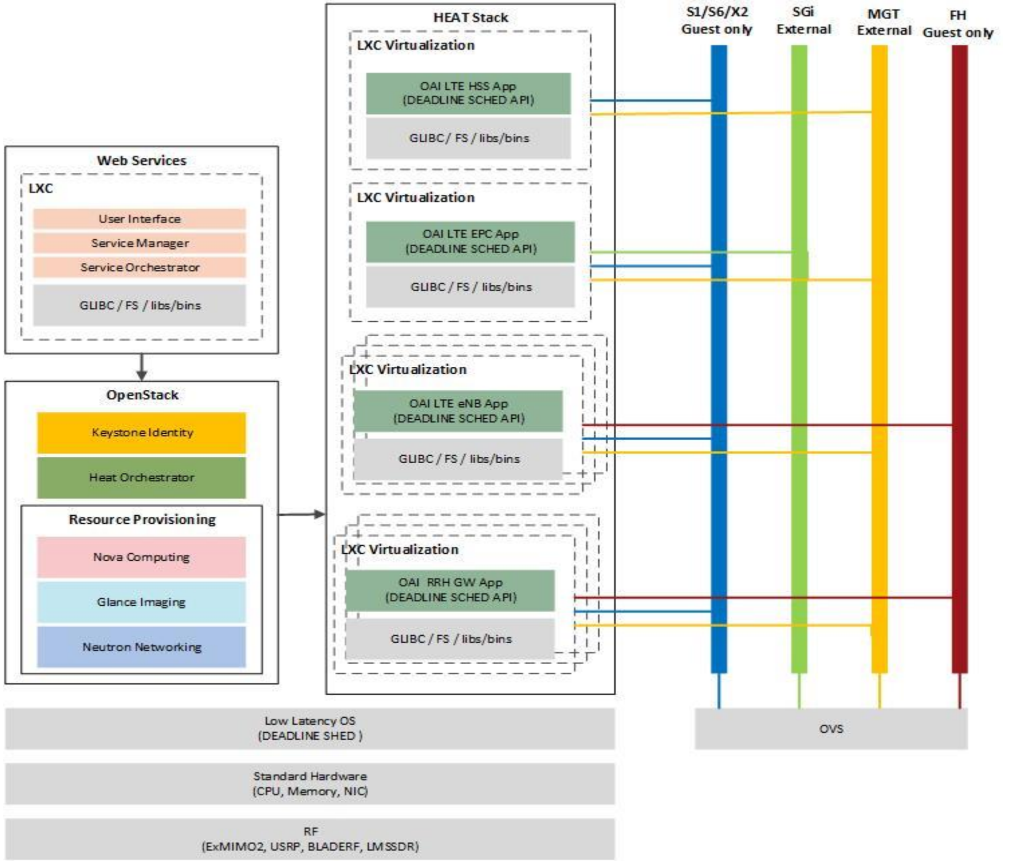
Based on achievable peak-rate on all RRHs

Scenario	1	2	3
Split A	5		
Split B	8		
Split C	9		
Split D	7	11	22
Split E	66	161	313

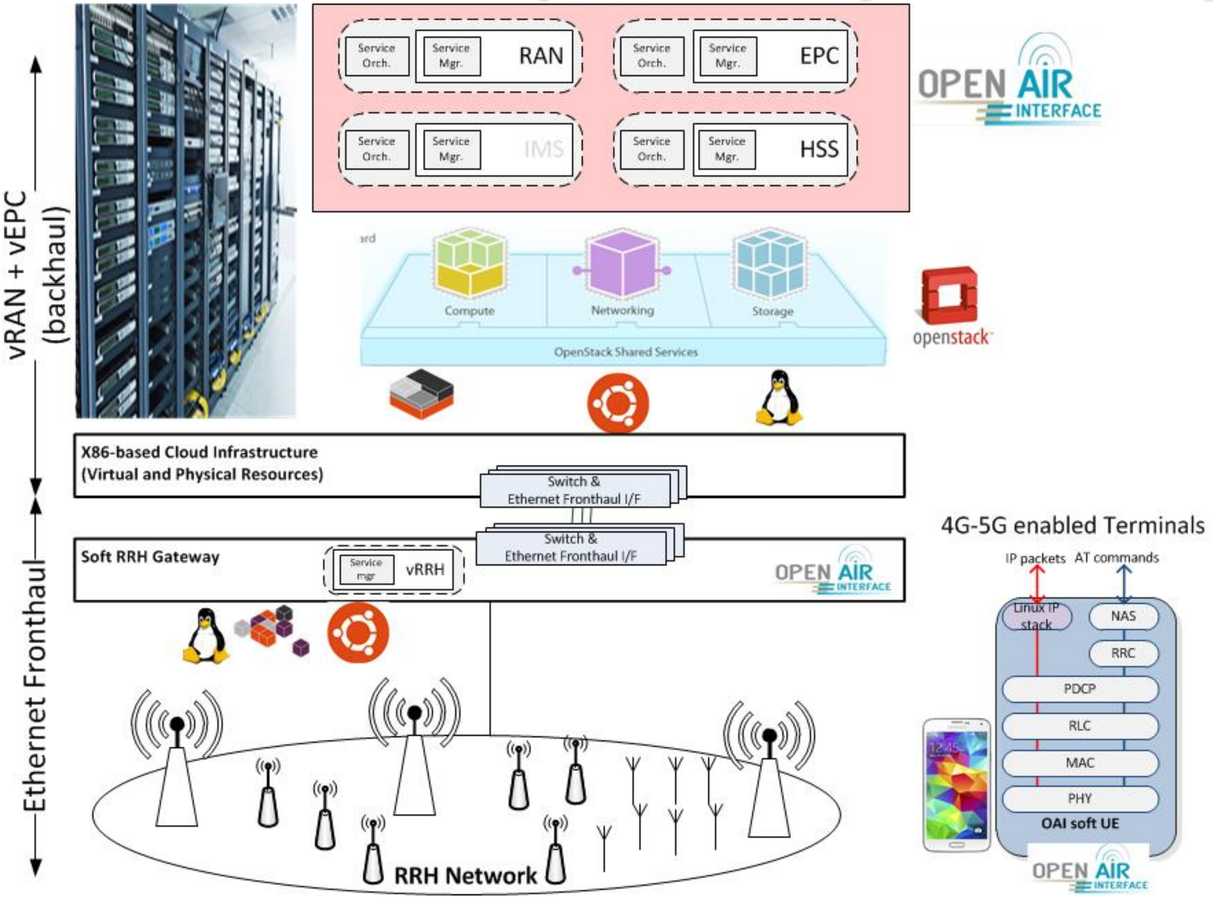


Cloud-native RAN Experiment Setup

- **Three components**
  - web service
  - OpenStack
  - Heat stack
- **Heat Template describes the virtual network deployment**
- **Linux Container**
- **Open vSwitch**
- **Low latency kernel**
- **RF frontend HW**



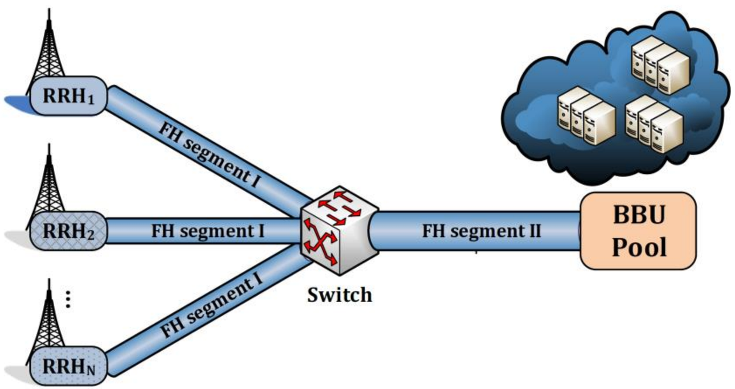
Cloud-native RAN
C-RAN Testbed on Sophia Antipolis Campus

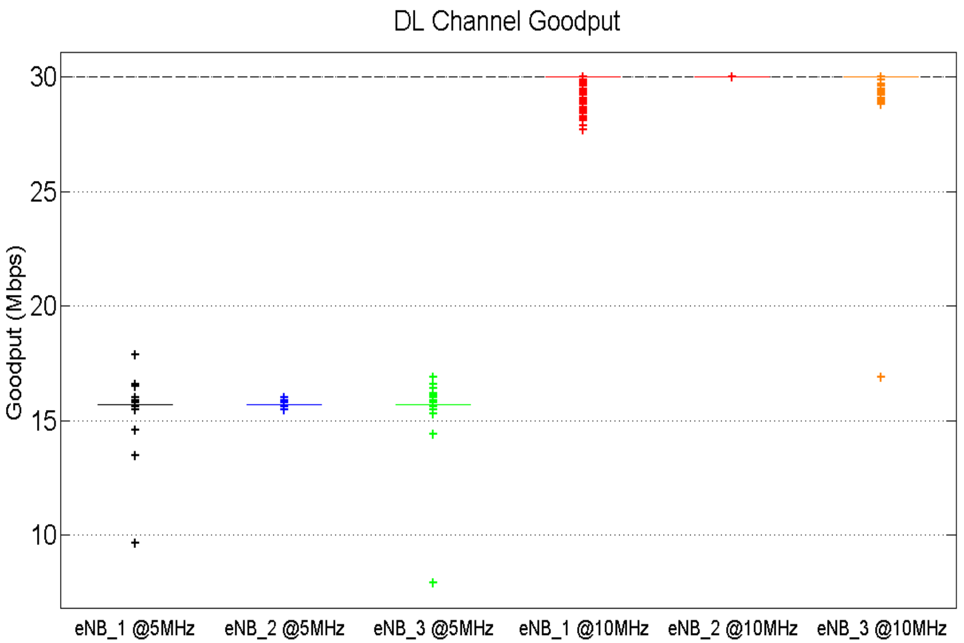


EURECOM

Cloud RAN
DL Performance

- **Three setting (FDD, SISO, with USRP B210 RF, Eth fronthaul network)**
  - eNB_1: No RRH
  - eNB_2: Local RRH
  - eNB_3: Remote RRH





EURECOM

Conclusions

- **4G/4G+ feasible on General Purpose Processors (x86) and Virtualization environment**
  - Exploit hybrid CPUs
- **Gap between virtualization and cloudification**
  - Exploit the microservice and NFV principles
- **Realtime network programmability is feasible at TTI level**
  - Exploit MEC principles for the data-plane programmability
- **Gap between static and cognitive management, self-adaptive, and learning methods**
  - Exploit machine learning and data mining techniques