SDR CLOUDS

RESOURCE MANAGEMENT IMPLICATIONS

INDEX

1. Introduction
2. Enabling Technologies
   + Middleware, Virtualization, Resource Control
3. Resource Management Implications
   + Resource Awareness and Modeling
   + Resource Management
4. Simulation Results
5. Conclusions

INTRODUCTION

The Cloud

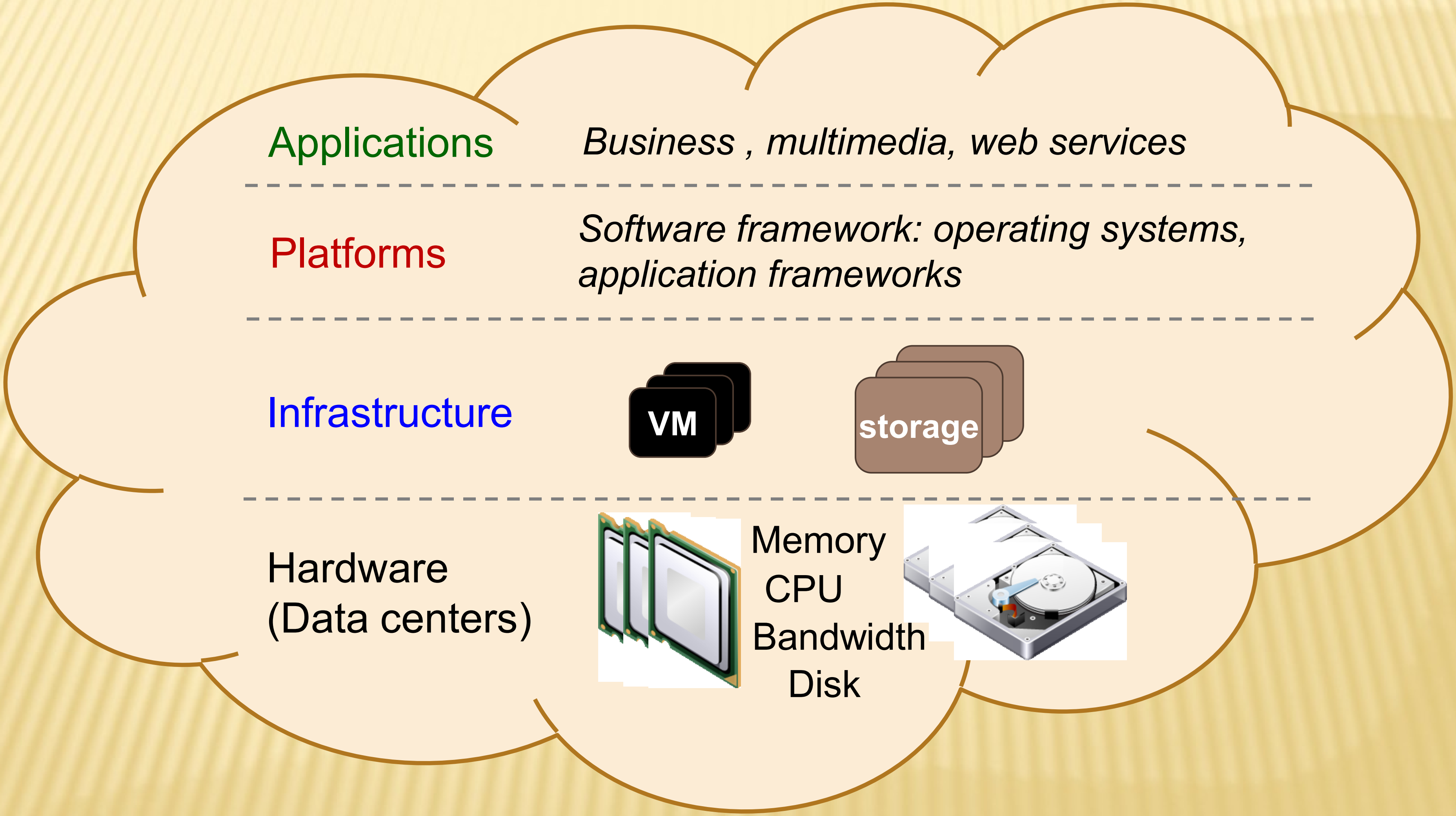
Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimum management effort or service provider interruption.

National Institute of Standards and Technology

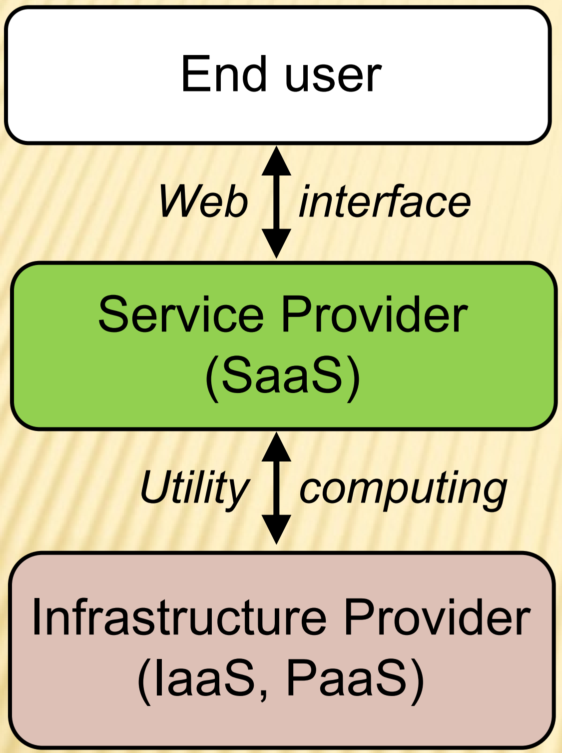
Cloud Computing Characteristics

- Multi-tenancy
- Shared resource pooling
- Geo-distribution and ubiquitous network access
- Service oriented
- Dynamic resource provisioning
- Self-organizing
- Utility-based pricing

Cloud Computing Architecture



Business Models



<u>Software as a Service (SaaS)</u>: providing on-demand applications over the Internet

Google Apps, Facebook, YouTube

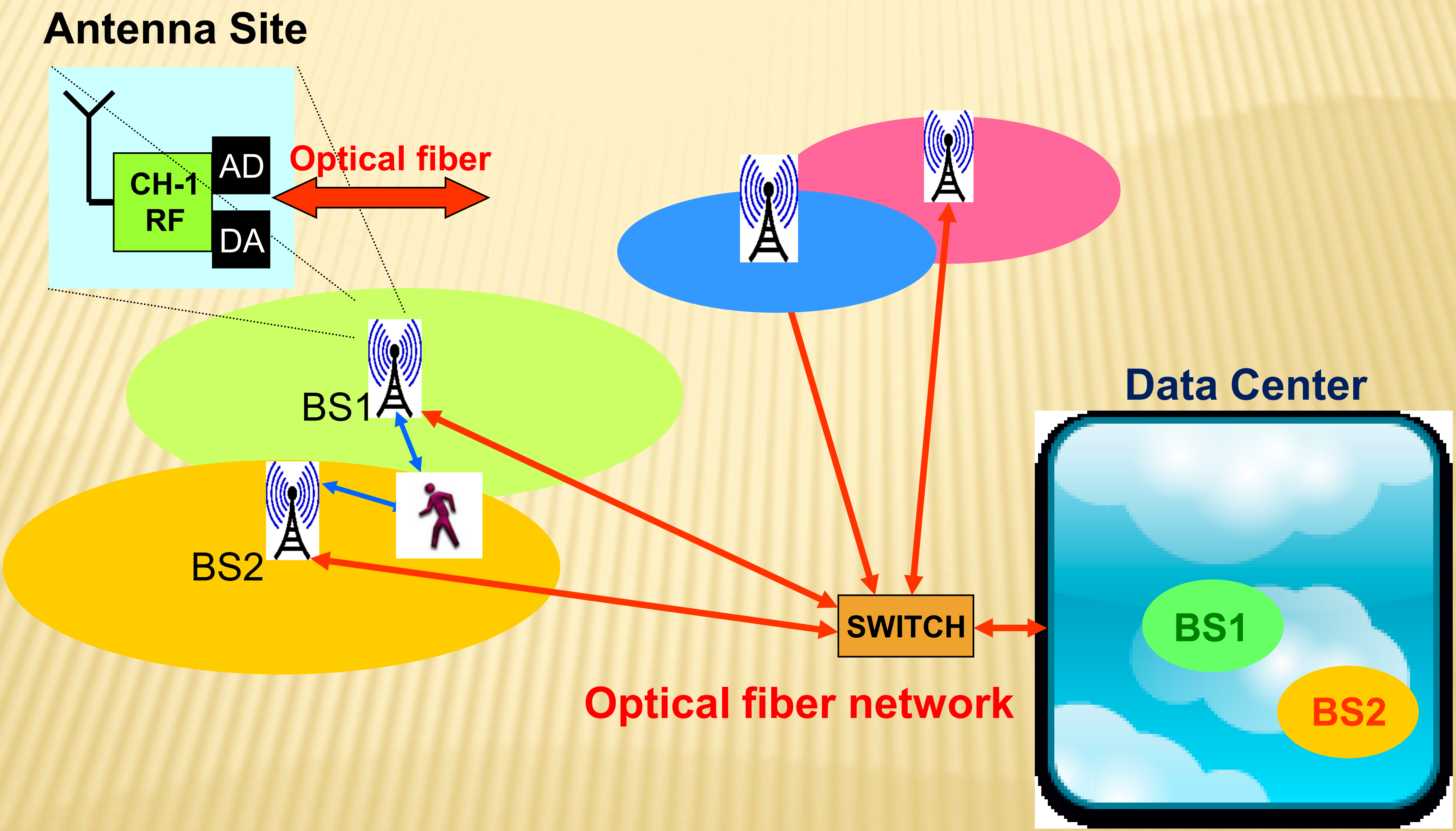
<u>Platform as a Service (PaaS)</u>: providing platform layer resources, e.g., operating system support and software development frameworks

Microsoft Azure, Google AppEngine, Amazon SimpleDB/S3

<u>Infrastructure as a Service</u>: on-demand provisioning of infrastructural resources (VMs)

Amazon EC2, GoGrid, Flexiscale

The SDR Cloud



Feasibility

- SDR clouds need to propagate and process real-time data
- Support high throughput and latency sensitive services. Principal issues:
  - Bandwidth
  - Latency

❖ ***Bandwidth*** limited by analog-to-digital conversion technology

Optical fiber transmission capacity: 10s Gbps (per channel)…10s Tbps (hundreds of channels)

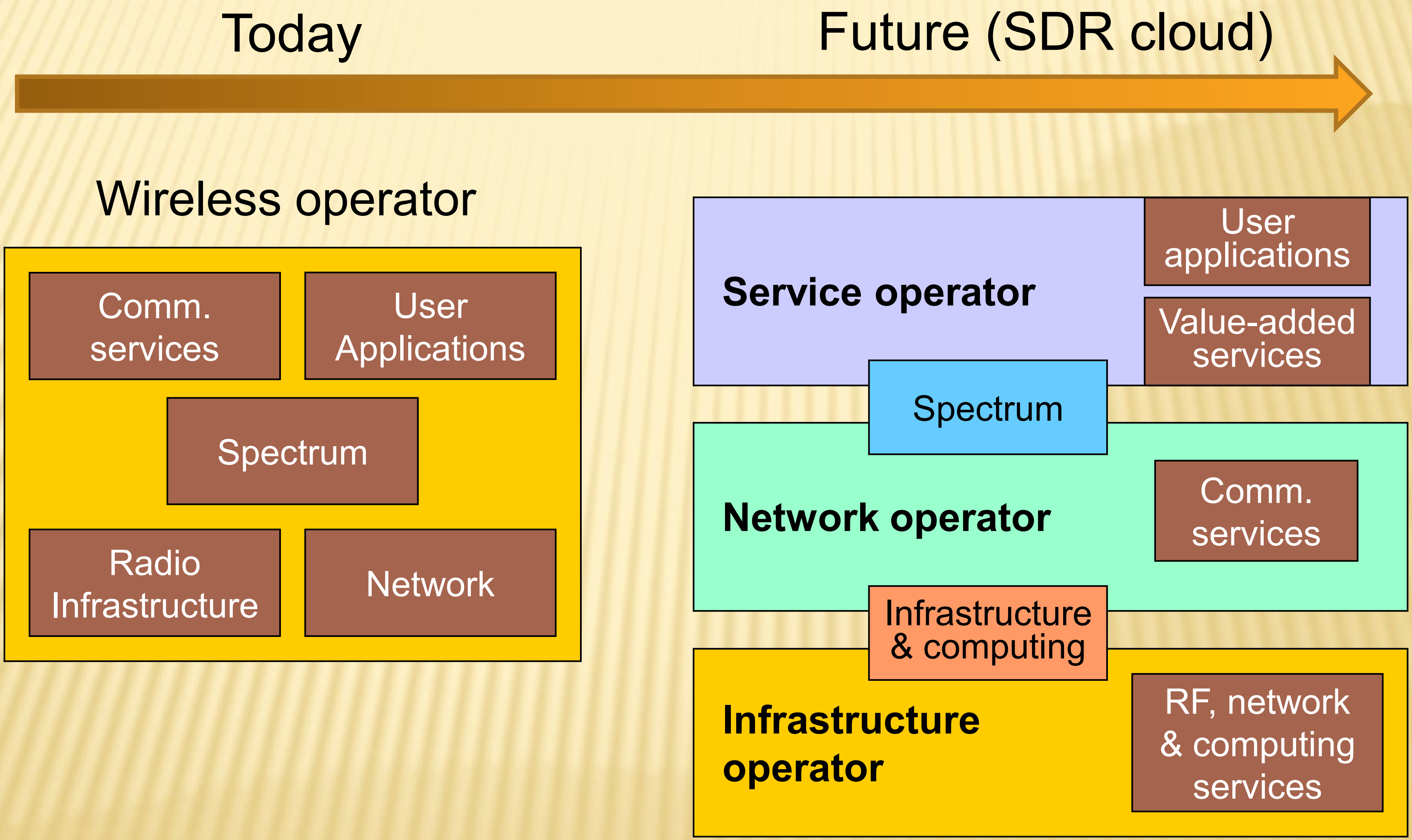
❖ ***Latency*** essentially determined by data path length between antenna site and data center

20 km long optical fiber path → ~0.1 ms

Advantages

- Radio infrastructure sharing (antennas, RF part) → reduced deployment cost
- Higher density of antennas, centralized processing of signals facilitates increasing the spectral efficiency
- Computing resource sharing, fewer over-provisioning, secondary use of idle resources → efficiency, scalability
- Waveform sharing, central repositories
- On-demand resource provisioning and charging
- New markets and market shares → value-added services
- Data centers upgradable with latest technology

Evolution



Today
Future (SDR cloud)
Wireless operator
Comm. services
User Applications
Spectrum
Radio Infrastructure
Network
Service operator
User applications
Value-added services
Spectrum
Network operator
Comm. services
Infrastructure & computing
Infrastructure operator
RF, network & computing services

SDR Cloud Services

- IaaS – VMs, distributed antennas, communication network (optical fiber)
  - Today's radio operators may become infrastructure operators
- PaaS – SDR frameworks/execution environments enabling and controlling distributed real-time execution of waveforms: SCA, ALOE, …
  - Software support tools designed by different R&D teams
- SaaS – available waveforms (SDR applications)
  - Today's radio operator may become SaaS providers, testing and approving waveforms designed by third parties

ENABLING TECHNOLOGIES

Middleware

- Middleware facilitates modular application design and distributed synchronized execution
- Provides communication services to components or processes running in different computers
- Synchronization necessary
  - between processors
  - between the data center and data converters

Virtualization

- Virtualization enables resource sharing
- SDR clouds may implement minimum level of virtualization
  - SDR applications compiled for the specific processor architecture (or for several architectures, if necessary)
  - virtualized or abstract computing resources: e.g., processor time, communication bandwidth, and system memory

- Resources shared between different clients/waveforms
- Mechanisms needed to ensure that
  - each client gets the required amount of resources (allocation)
  - no client can use more than the allocated resources (control)

Resource Control

- Resource control ensures that processes do not access more than the assigned amount of resources
- A high-resolution resource control necessary to instantly identify any runtime resource violation and impede that one waveform blocks the real-time execution of others
- Resolutions orders of 0.1 ms without excessive overhead
- Grid or cloud computing do not provide this accuracy

RESOURCE MANAGEMENT IMPLICATIONS

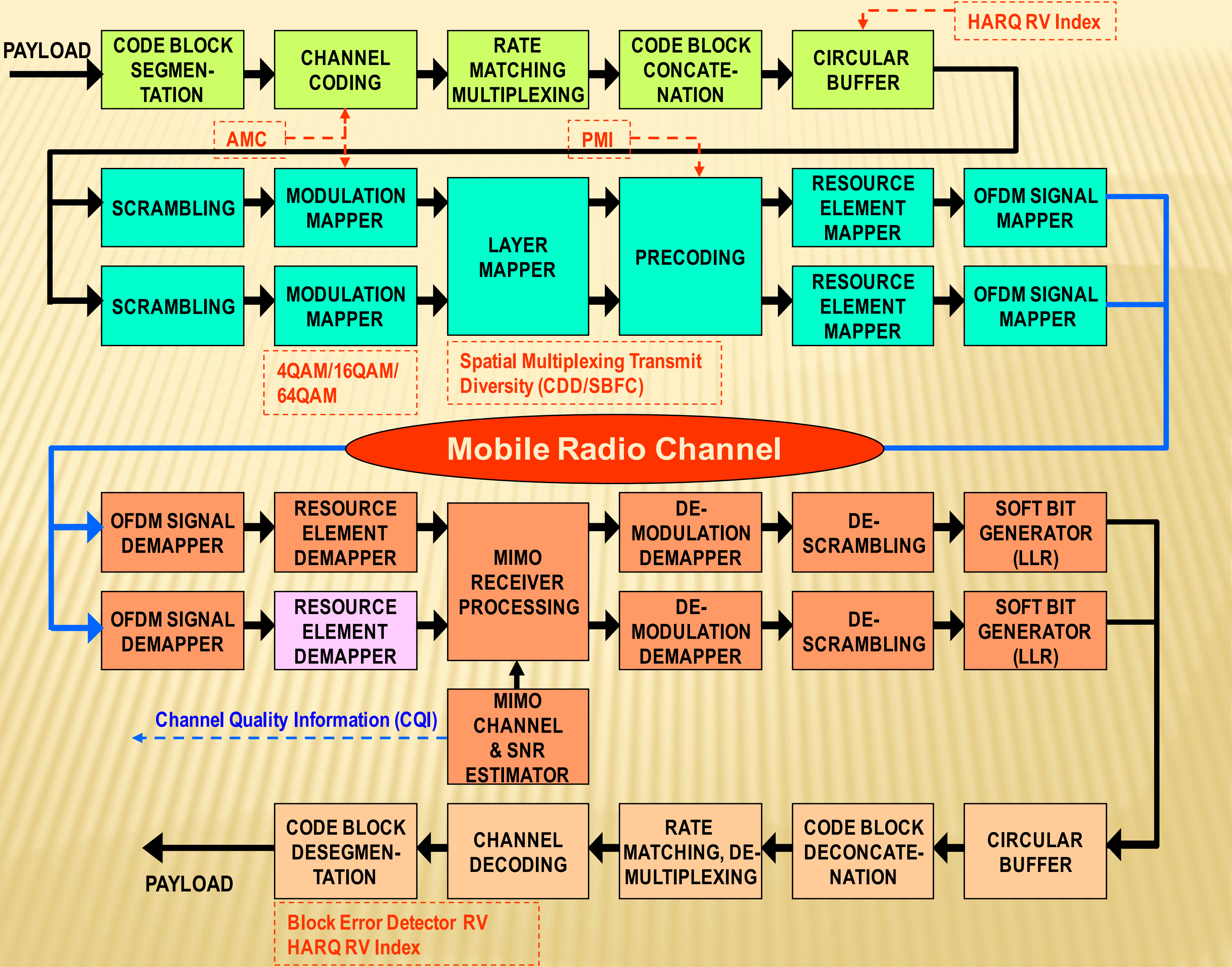
Resource Management Context

- Wireless subscribers demand different types of comm. services throughout a day
- User penetrate different geographical regions
- Initiating a user session involves allocating computing resource for physical layer digital signal processing
- Only a few (10s) milliseconds available for establishing data route from antenna to data center and allocating computing resources for waveform processing
- 1000s of processors available in the data center for serving 1000s of waveforms at a time

Motivation

- Ad-hoc SDR cloud solutions are not reasonable
- Platform-independent SDR provides highest flexibility:
  - Deployment on different hardware (data centers)
  - Accelerates waveform design and innovation
  - Dynamic provisioning of new and personalized services
  - …

➡ Computing resource awareness
and dynamic, real-time allocation


PAYLOAD
CODE BLOCK SEGMEN-TATION
CHANNEL CODING
RATE MATCHING MULTIPLEXING
CODE BLOCK CONCATE-NATION
CIRCULAR BUFFER
HARQ RV Index
AMC
PMI
SCRAMBLING
MODULATION MAPPER
LAYER MAPPER
PRECODING
RESOURCE ELEMENT MAPPER
OFDM SIGNAL MAPPER
SCRAMBLING
MODULATION MAPPER
RESOURCE ELEMENT MAPPER
OFDM SIGNAL MAPPER
4QAM/16QAM/64QAM
Spatial Multiplexing Transmit Diversity (CDD/SBFC)
Mobile Radio Channel
OFDM SIGNAL DEMAPPER
RESOURCE ELEMENT DEMAPPER
MIMO RECEIVER PROCESSING
DE-MODULATION DEMAPPER
DE-SCRAMBLING
SOFT BIT GENERATOR (LLR)
OFDM SIGNAL DEMAPPER
RESOURCE ELEMENT DEMAPPER
DE-MODULATION DEMAPPER
DE-SCRAMBLING
SOFT BIT GENERATOR (LLR)
MIMO CHANNEL & SNR ESTIMATOR
Channel Quality Information (CQI)
CODE BLOCK DESEGMEN-TATION
CHANNEL DECODING
RATE MATCHING, DE-MULTIPLEXING
CODE BLOCK DECONCATE-NATION
CIRCULAR BUFFER
PAYLOAD
Block Error Detector RV
HARQ RV Index

Resource Awareness and Modeling

- SDR applications run at highest priority and should not interrupted
- Deterministic execution times, SNR dependent (e.g. iterative decoders)
- SDR applications need to be certified => correct and deterministic execution behavior (SNR-dependent)
- Measure execution time or resource consumption offline, e.g. with random input data (time → MOPS)
- Create corresponding models (waveform computing requirements)

Resource Management

- Objective: Ensure real-time execution of waveforms under service-dependent end-to-end latency constraints
- Continuous allocation and reallocation of resources
- Stringent timing constraints
- Resource allocation (mapping and scheduling) very complex

➡ Hierarchical resource management

High-level resource management

- Data centers can be grouped in clusters
- It is often more efficient to “move” the computation to the data, rather than moving large data amounts
- The high-level resource management assigns clusters to radio operators, radio cells, user groups, or …
- This management is dynamic, but slowly varying
- It may take into account communications statistics for facilitating secondary usage of idle clusters

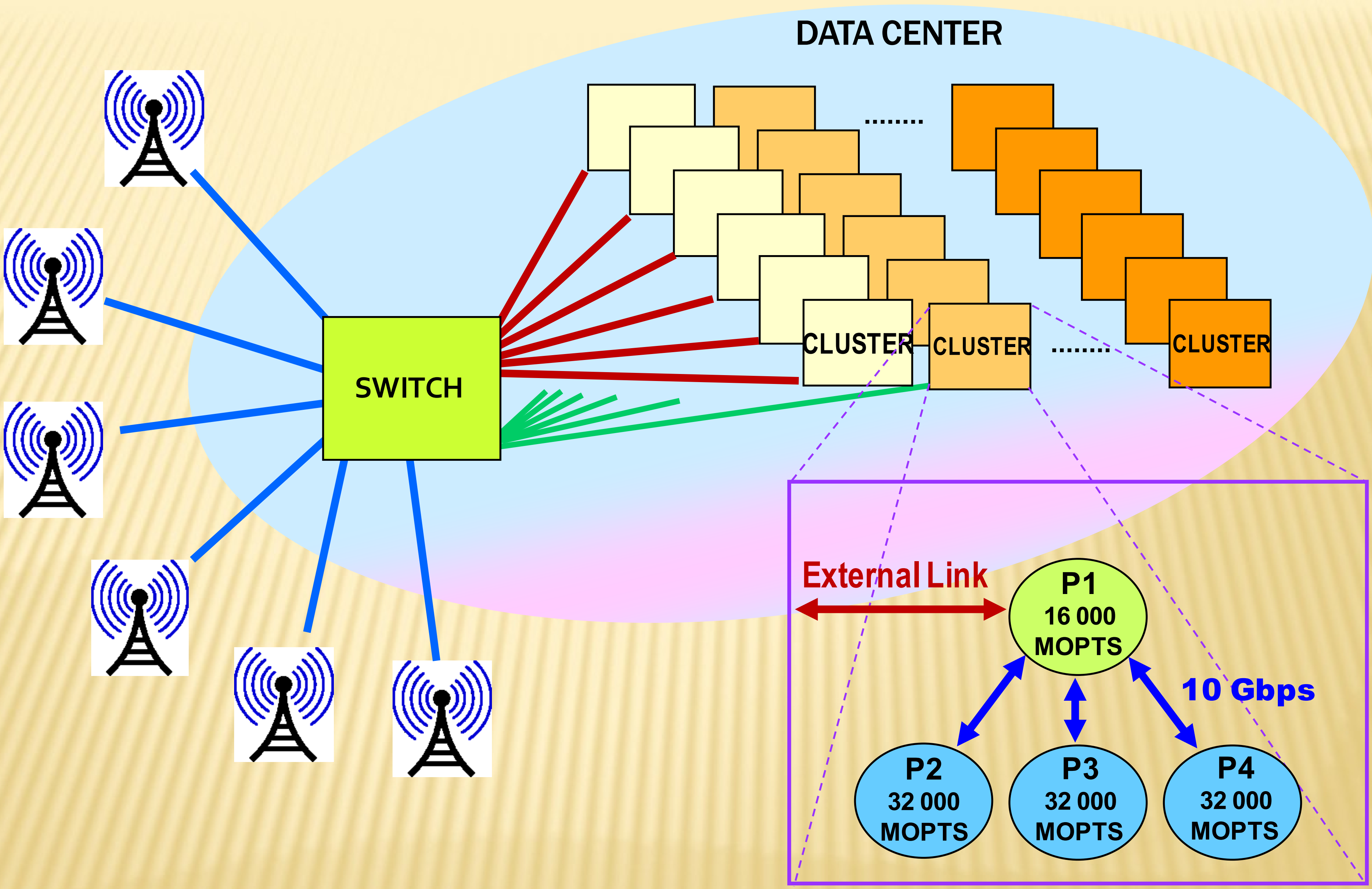
Low-Level Resource Management

- Real-time allocation of individual computing resources (CPUs, memory, bandwidth, …): mapping of computing requirements to computing resources
- The goal is to find sufficient resources within a cluster or tightly-coupled group of clusters (previously assigned) in real-time (ms)
- Waveform modules can then be loaded to processors for immediately processing incoming and outgoing signals
- Highly dynamic: resources allocated during session establishment and freed when session terminates

SIMULATION RESULTS

Scenario

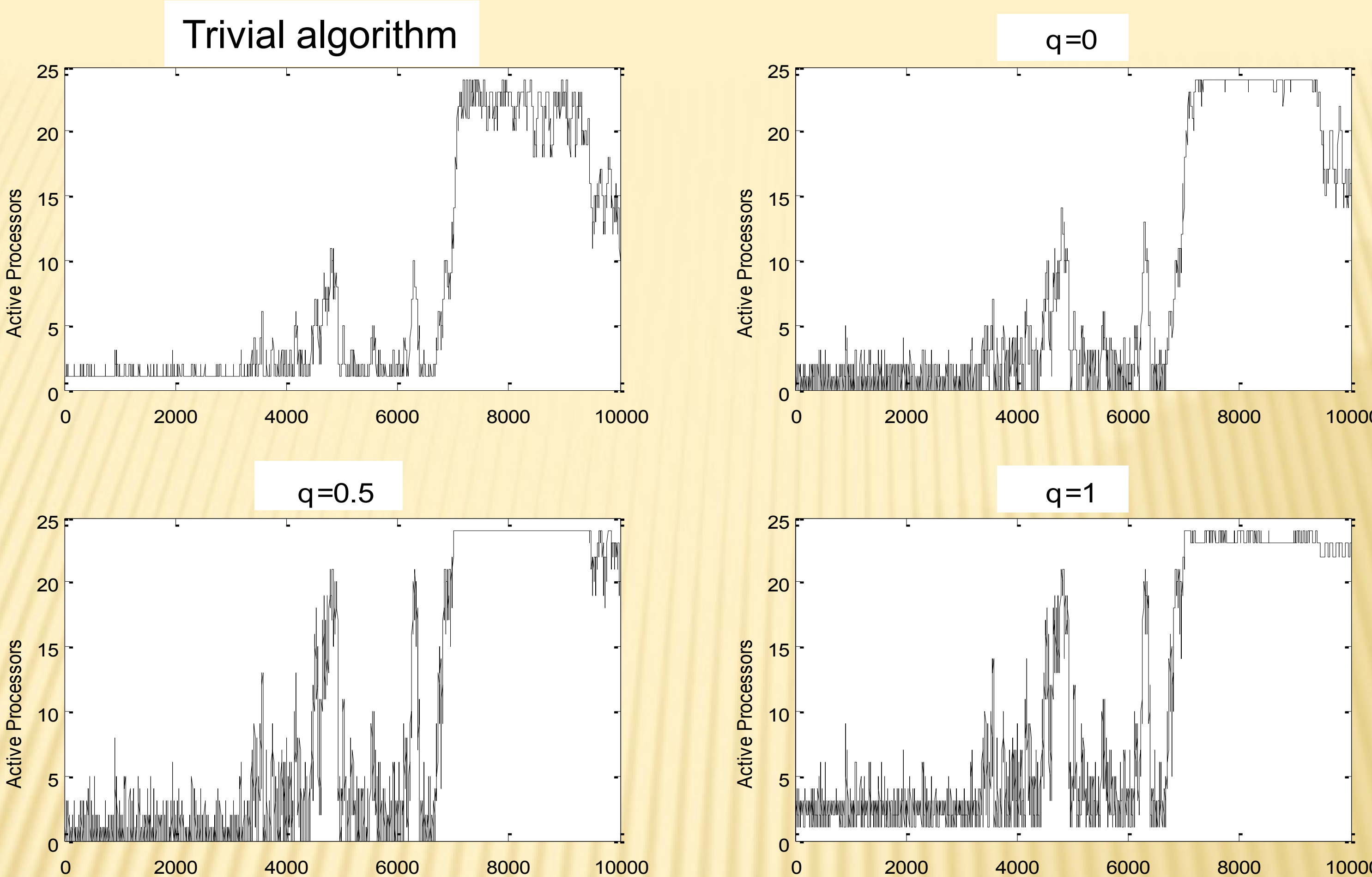
- Radio operator wants to deliver 3G access in certain area
- Receiver digital signal processing chain requires ~8150 MOPS (chip- & bit-rate processing model of UMTS receiver)
- 3G service area covered by a set of antennas
- An analog-to-digital converter at each antenna samples the signal with 16 bits per sample at a rate of 65 MHz
- Samples are sent to the datacenter switch at ~1 Gbps



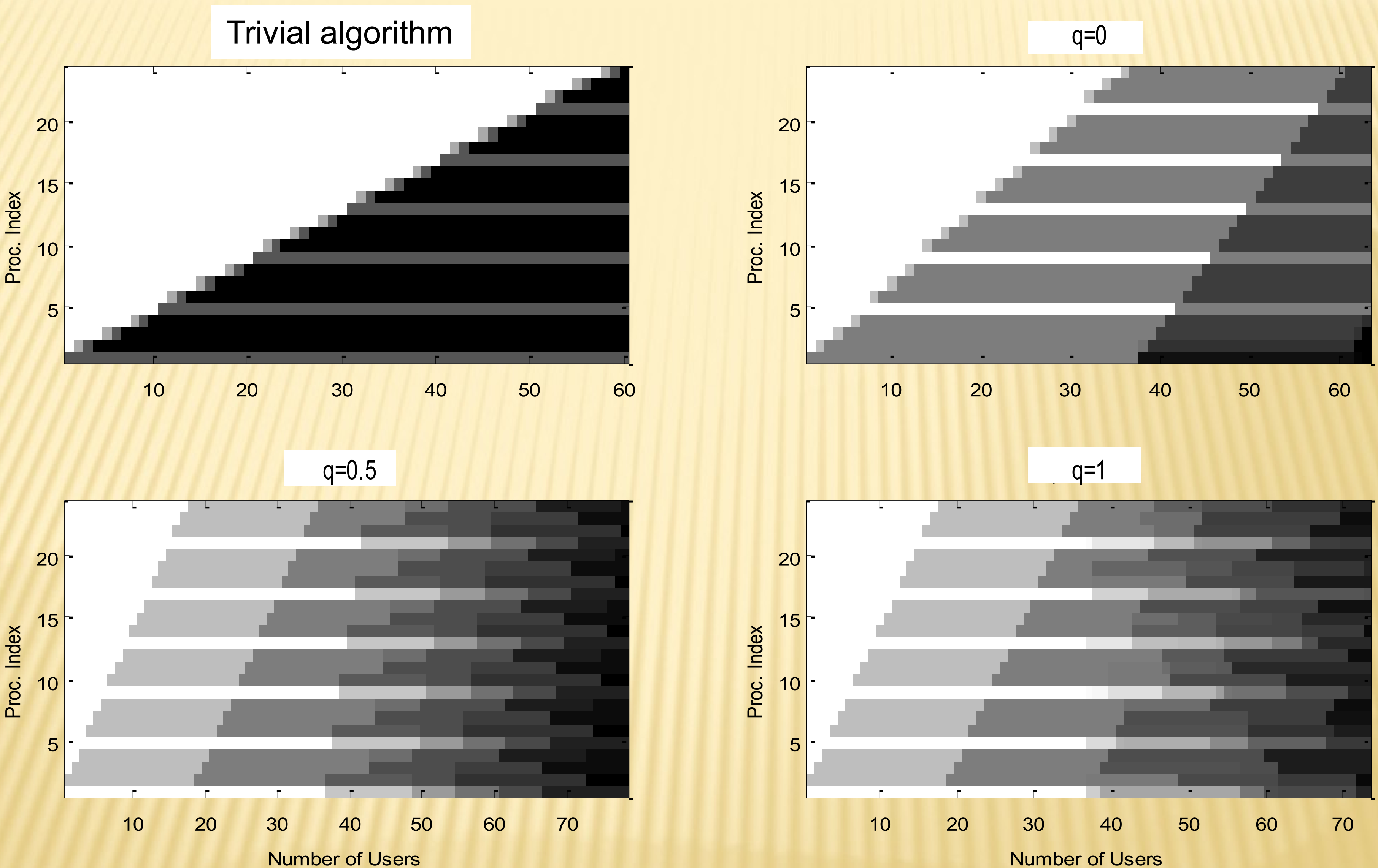
6 clusters, 672,000 MOPTS total processing capacity (max. 82 users)

Simulation

- Markov-chain user arrival and serving process (M/M/1)
- System load changed from low to medium and then to unstable
- Computing resource allocation algorithms:
  - Trivial algorithm: fills processors one after another
  - t_1-mapping: dynamic programming algorithm
    - cost function balances the processing load (q) and minimizes interprocessor data flows (1–q)



	Trivial algorithm	*t_1-mapping*		
		***q* = 0**	***q* = 0.5**	***q* = 1**
Max users	61	64	**79**	74
Avg. load	70 %	74 %	**94 %**	87 %


Trivial algorithm
q=0
q=0.5
q=1
Proc. Index
Number of Users
5
10
15
20
10
20
30
40
50
60
70

CONCLUSIONS

- SDR clouds: merge SDR with cloud computing
- Scalable solution for wireless communications
- Technological constraints
- Computing resource management implications
  - Real-time computing resource allocation for very-large scale systems
  - End-to-end system latency control and management