Department of Signal Theory and Communications
UNIVERSITAT POLITÈCNICA DE CATALUNYA

RESOURCE MANAGEMENT STRATEGIES FOR SDR CLOUDS

Vuk Marojevic
Ismael Gomez
Pere Gilabert
Gabriel Montoro
Antoni Gelonch

INDEX

1. Introduction
2. Resource Management Context & Approach
3. Resource Management Strategies
4. Simulations
5. Conclusions

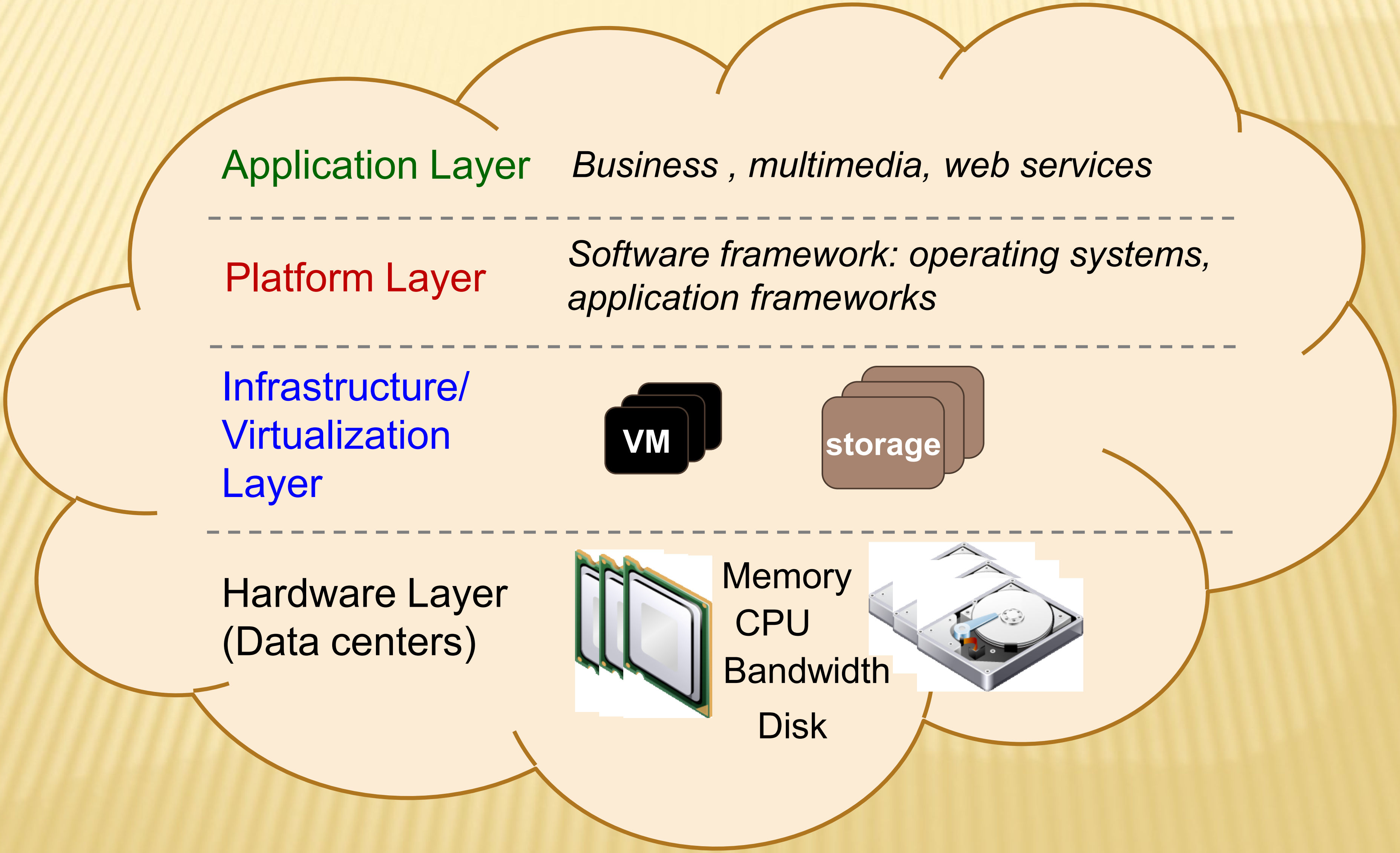
INTRODUCTION

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

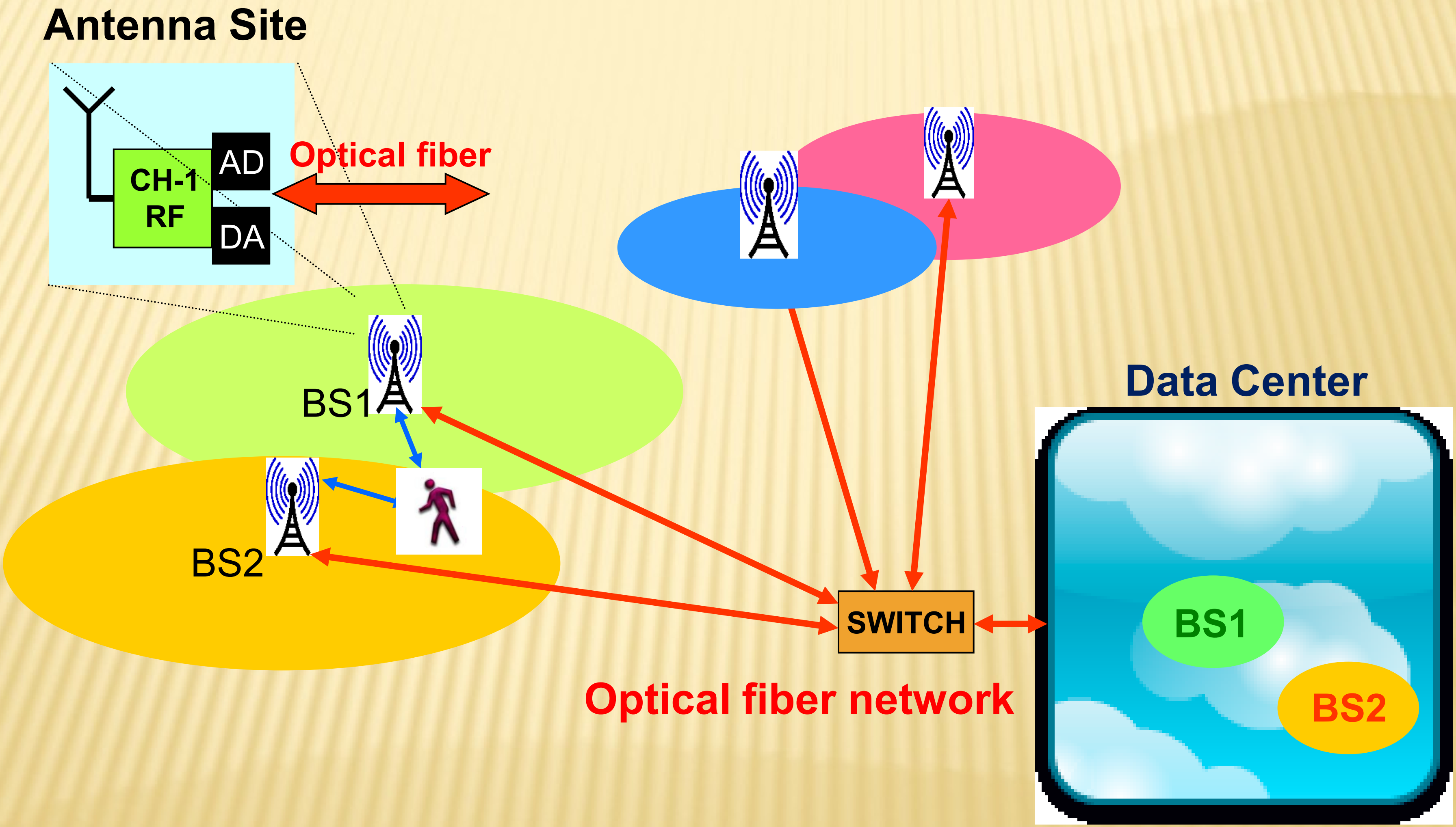
Cloud Computing Architecture



Cloud Computing Characteristics

- Service oriented
- Multi-tenancy
- Ubiquitous network access
- Shared resource pooling
- Dynamic resource provisioning
- Self-organizing
- Utility-based pricing

The SDR Cloud



Advantages

- Radio infrastructure sharing (antennas, RF part) → reduced deployment cost
- 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

RESOURCE MANAGEMENT CONTEXT & APPROACH

Coverage

- Latency-constrained
- Transmission delay over optical fiber
  - Distance, routing path, optical fiber switches
- 20 km data path: approx. 0.1 ms
- Assuming 10 km (6.2 mi) radius → 314 km^2 (120 mi^2) (Barcelona: 100 km^2 , 1.6 M inhabitants)

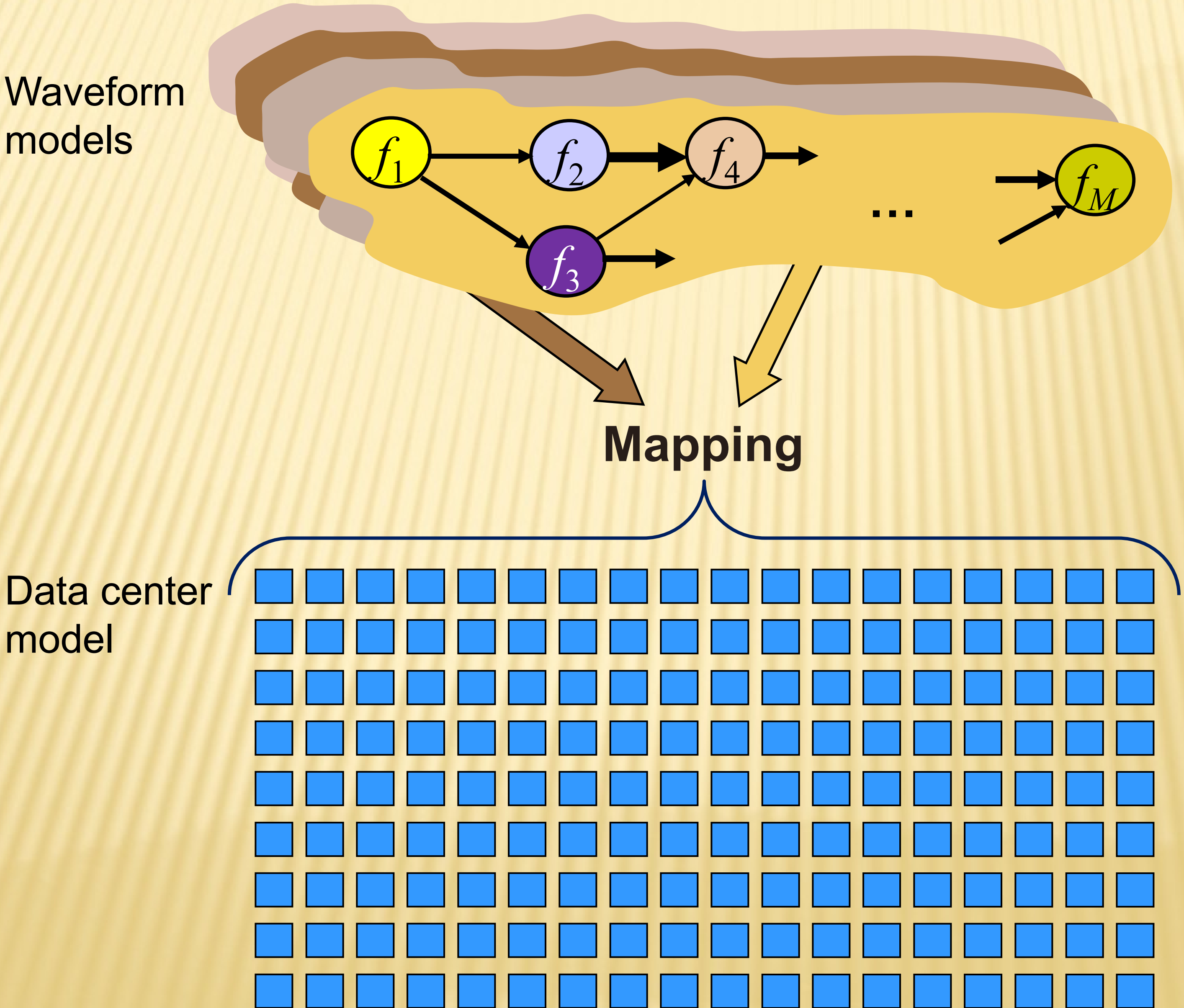
Traffic Implications

- Independently session initiations and terminations
- Several communication sessions per day of different durations
- Users mobility
- More than 20,000 wireless communications sessions at peak (2 % of one million subscribers)
- 10 GOPS for the PHY processing per user → 200,000 GOPS for 20,000 parallel sessions
- 10x, 100x, … for future SDR communications systems

Resource Management requirements

- Dynamic & continuous allocation and reallocation of resources
- Ensure real-time execution of waveforms under service-dependent throughput & latency constraints
- Adapt to the given traffic distribution
- Dispatch huge number of parallel session requests
- Acceptable session establishment times: real-time computing resource allocation
- Serve as many users a possible (high resource occupation)

Resource Management



Mapping Complexity

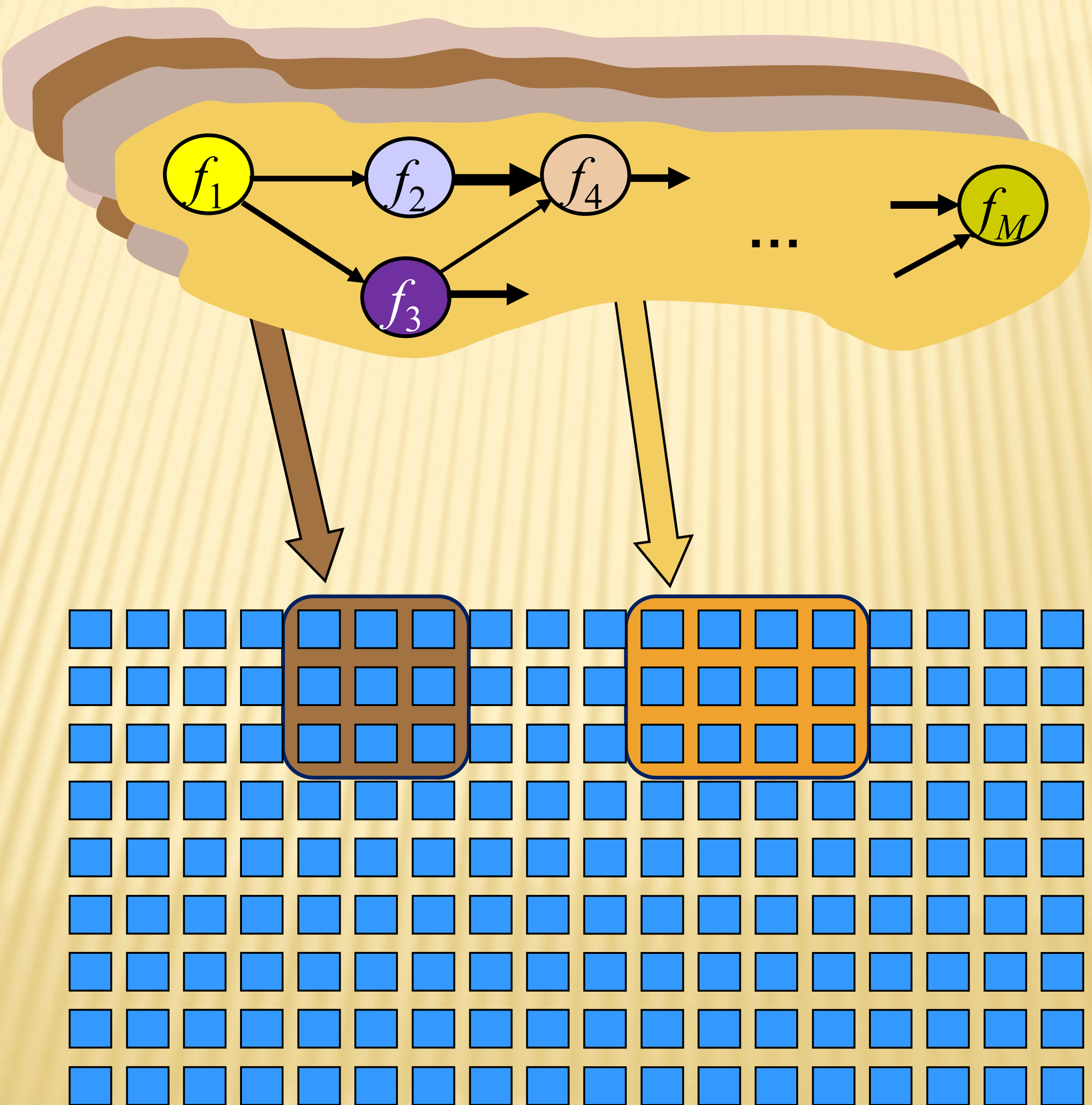
- t_w-mapping algorithm complexity: $O(M \cdot N^{w+1})$
- Execution time in seconds (2.67 GHz i7 Quadcore, M = 24):

N (number of processors)	*w* (window size)		
	1	2	3
20	0.005	0.09	1.57
30	0.025	0.61	16.23
40	0.075	2.43	87.77
50	0.17	7.2	326.4
100	2.9	221.3	-
200	68.6	-	-
300	329.2	-	-

V. Marojevic, X. Revés, A. Gelonch, "A computing resource management framework for software-defined radios," *IEEE Trans. Comput.*, vol. 57, no. 10, pp. 1399-1412, Oct. 2008.

Hierarchical Resource Management

Waveform models



Data center model

High-level resource management

- Divide data center into computing clusters
- High-level resource manager assigns clusters to radio operators, radio cells, services, or …
- Dynamic clustering, slowly varying
- Account for communications statistics for secondary usage of idle clusters

Low-Level Resource Management

- Real-time allocation of computing resources (CPUs, memory, bandwidth, ...) to waveforms
- Waveform modules then loaded to processors for immediately processing incoming/outgoing signals
- Very dynamic: resources allocated during session establishment and freed when session terminates

RESOURCE MANAGEMENT STRATEGIES

Strategy 1 (S1): Operator Clusters

- Clusters assigned to radio operators
- Radio operators may demand a certain number of clusters based on expected traffic loads → pre-allocations
- Dynamic allocation
- Combination resource pre-allocation (minimum resource guaranteed) and dynamic allocation
- Only few radio operators and large service area → combining Strategy 1 with another

Strategy 2 (S2): Cell Clusters

- Clusters assigned to radio cells
- Different cell sizes & time-varying traffic loads
- Pre-allocations vs. dynamic clustering
- Clusters may grow or shrink as required
- S2 may simplify the access to the fiber optical network

Strategy 3 (S3): Service Clusters

- Clusters assigned to different services
- Service-dependent resource optimization goals
- Different services have more or less stringent timing and computing constraints
  - High throughput services (mobile TV) → allocate parallel resources
  - Low latency services (voice, video) → less parallelization, less processing latency
- S3 may be combined with another (S2)

SIMULATIONS

Simulation Setup

- 2 radio operators
- 64 radio cells
- 3G services:
  - 64 kbps (voice), 128 kbps, 384 kbps, and 1024 kbps
  - UMTS receiver digital signal processing chain (chip- & bit-rate processing model, ~7000-10,000 MOPS)
- Data center:
  - 256 Quad-cores (1024 processors)
  - 12 GOPS per core
  - Fully connected

Scenarios

	OP 1	OP 2	64 kbps *voice*	128 kbps *data*	384 kbps *data*	1024 kbps *data*	User distr.
I	50 %	50 %	50 %	20 %	20 %	10 %	Uni-form
II	75 %	25 %	50 %	20 %	20 %	10 %	Uni-form
III	50 %	50 %	25 %	25 %	25 %	25 %	Uni-form
IV	50 %	50 %	50 %	20 %	20 %	10 %	Gaus-sian

Scenario IV: Strategies

- 64 radio cells, divided into 16 zones
- 128 Quad-cores per operator assigned to zones as shown:

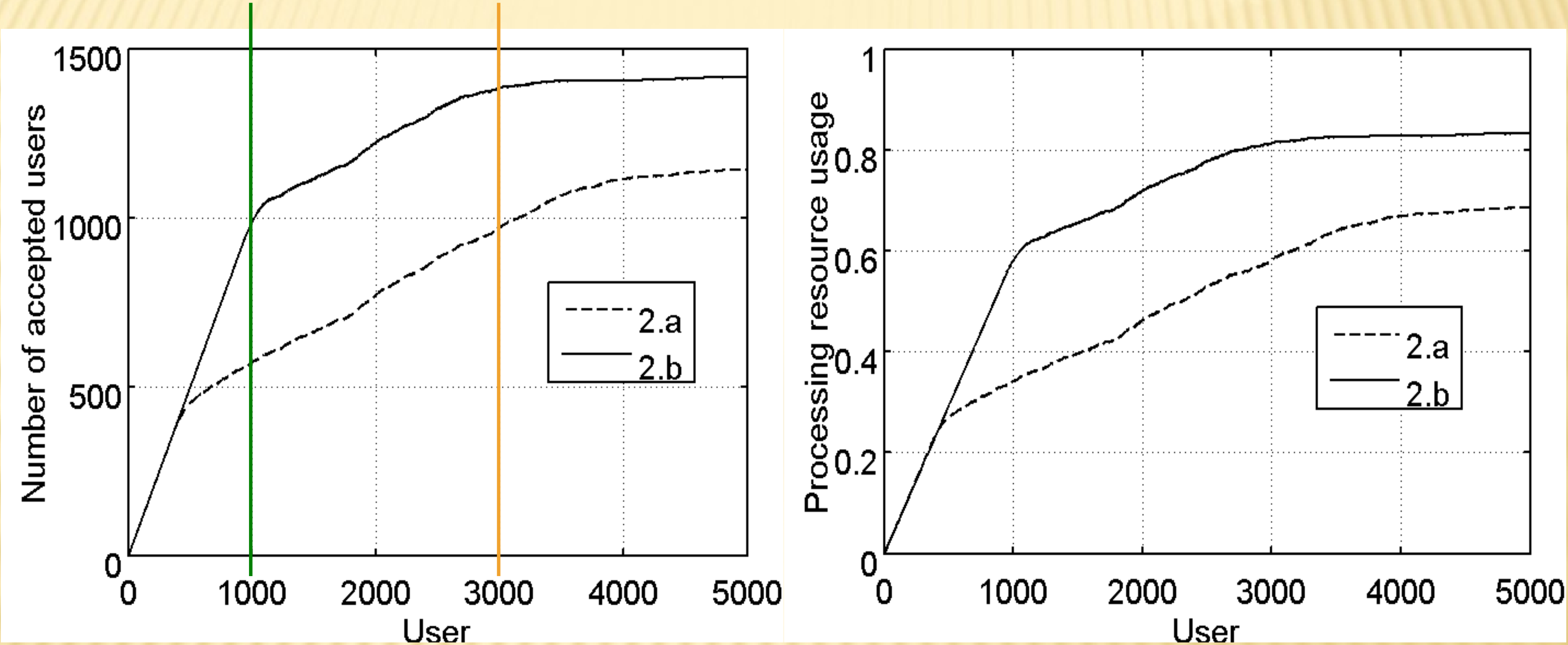
Strategy 2.a

8	8	8	8
8	8	8	8
8	8	8	8
8	8	8	8

Strategy 2.b

2	6	6	2
6	18	18	6
6	18	18	6
2	6	6	2

Scenario IV: Results



CONCLUSIONS

- SDR clouds: merge SDR with cloud computing
- Scalable solution for wireless communications
- Computing resource management strategies
  - Tradeoff between resource allocation efficiency and flexibility
- Results for different resource management strategies
- Dynamic adaptations needed
- Dynamically definable strategies