Curl-free Scheduling Fields: A Fundamental Characterization of Stability in Wireless Networks

Gerhard Wunder

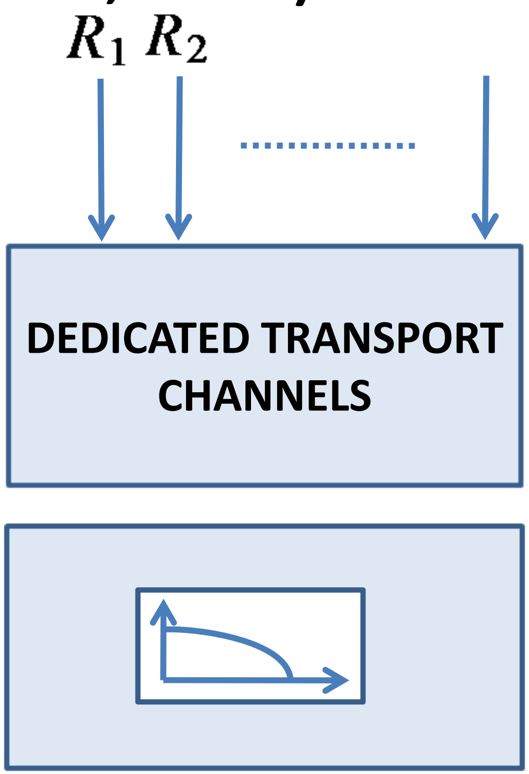
ITG Fachgruppentreffen
Aachen Feb. 12th 2008

Joint Work with Z. Chan (PhD Cand.) and Thomas Michel (PhD)

Fraunhofer German-Sino Lab Mobile Communications

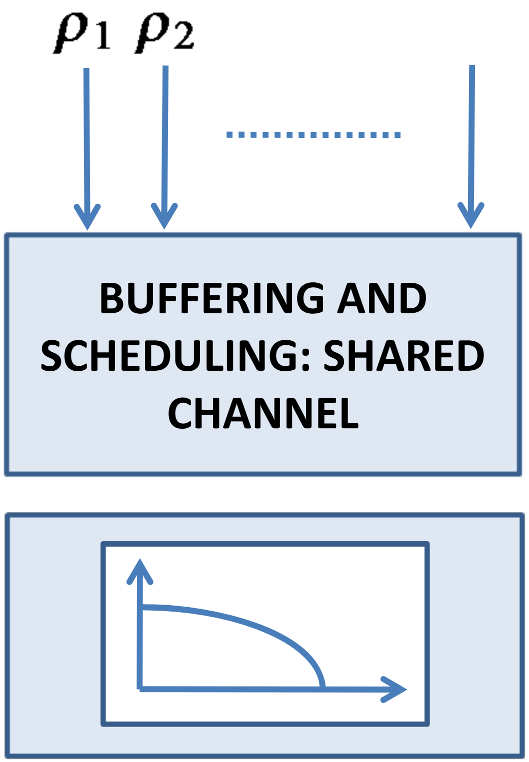
Crosslayer Design Mobile Commun. Networks

„Traditional“ MAC Design (GSM, UMTS)



Rate Region without CSI

„State of the art“ MAC Design (HSDPA, LTE)

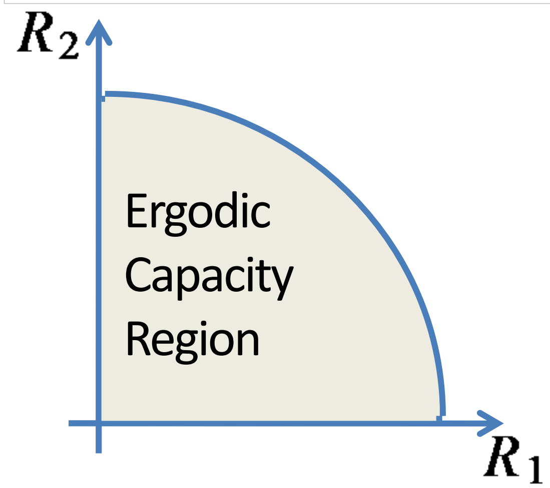


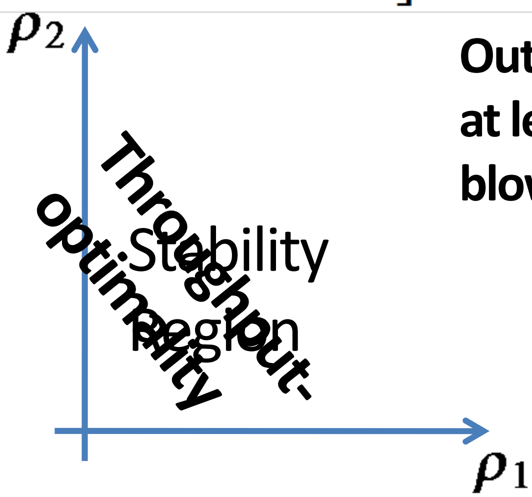
Rate Region with CSI

Crosslayer Design

- **Benefits**: cope with **random traffic**, achieve **multiuser diversity**, and "**learn**" **ergodic** capacity region with CSI, i.e. long term supportable rates by employing **scheduling**.

- Maximum Weight Matching policy [Tassiulas et al ’92]
  Exponential Rule: [Shakkottai & Stolyar ’02]
  Queue Proportional, Idle State Prediction
  [Seong & Cioffi ’06], [Zhou & Wunder ’07]





Outside stability region: at least one buffer blows up over time!!

Results

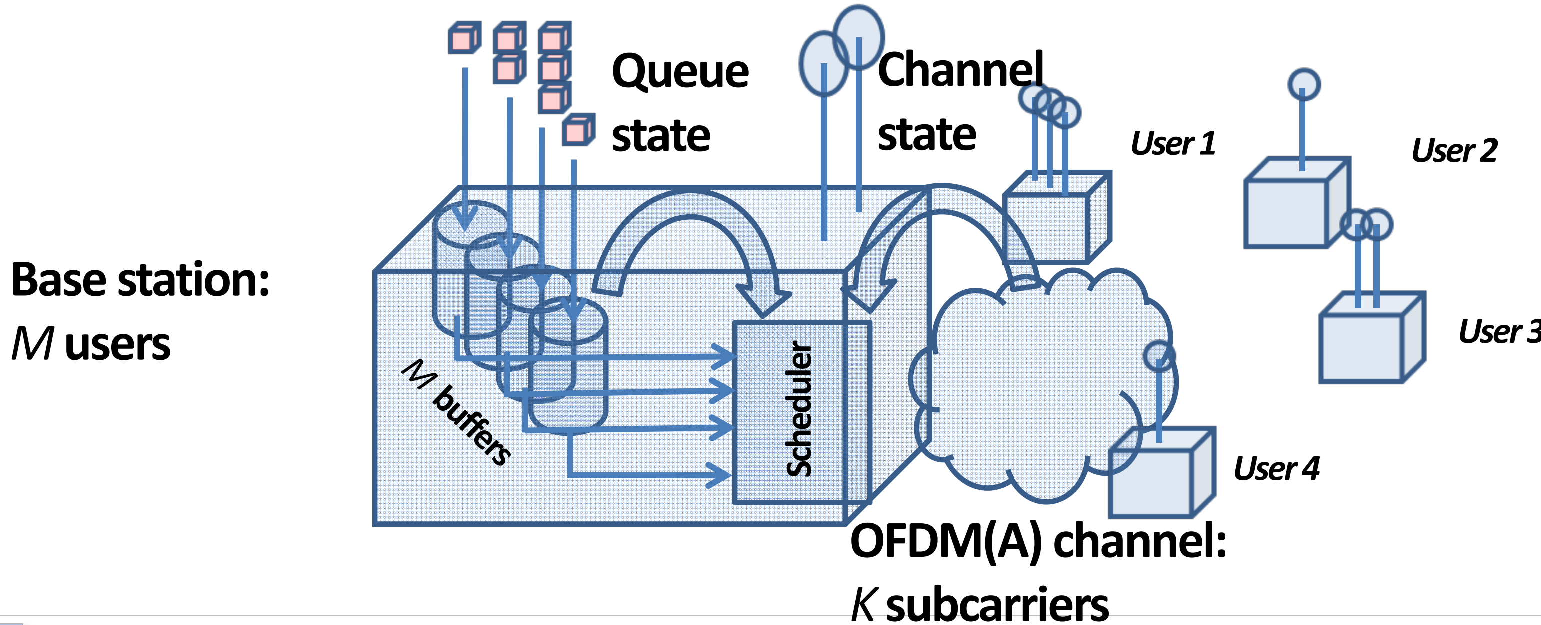
- We show that a general, **comprehensive** representation and **universal decomposition** of scheduling policies exist.
- We provide a **canonical** approach to design **throughput-optimal** scheduling policies (helps to solve the long open-standing problem of delay-optimality).
- We show that the **intrinsic resource allocation problem** has combinatorical nature that can be incorporated "from scratch".

Content

- System model and decomposition
- Curl-free scheduling fields
- Ressource allocation
- Outlook and coclusions

System Model and Decomposition

System Model



- Let $n \in \mathbb{N}$ be the time slot; the packet arrival process $\mathbf{a}(n) \in \mathbb{R}_+^M$ is **iid** with mean rate $\boldsymbol{\rho} := \mathbb{E}(\mathbf{a}(n))$ and $\Pr(\mathbf{a}(n) = 0) > 0$.

- The rate process $\mathbf{r}(n) \in \mathbb{R}_+^M$ is **iid** and $\mathbf{r}(n) \in \mathcal{C}(\mathbf{h}(n), P(n))$ where $\mathcal{C}(\mathbf{h}(n), P(n)) \subset \mathbb{R}_+^M$ is instantanous (**discrete**) rate region; $\mathbf{h}(n) \in \mathbb{R}_+^{MK}$ is **vector of channel gains**, $P(n) \in \mathbb{R}_+$ is **power budget**.

LTE OFDMA Downlink Channel

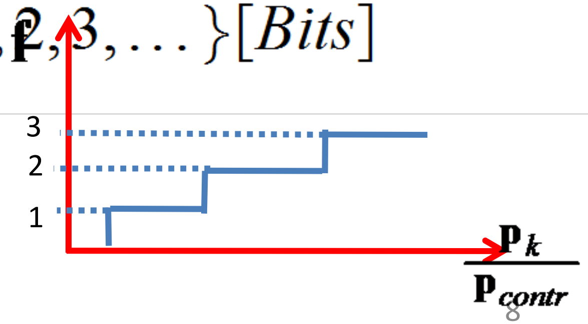
- Denote backlog as $\mathbf{q}(n) \in \mathbb{R}_+^M$; by our assumptions the **queueing system** evolves as δ_0-irreducible **Markov chain**:

$$\mathbf{q}(n+1) = [\mathbf{q}(n) - \mathbf{r}(n) + \mathbf{a}(n)]^+$$

- Due to OFDM(A) $\mathcal{C}(\mathbf{h}, P)$ is generated by 1.) **exclusive** assignment of **subcarrier sets** $\mathcal{S}_1, \ldots, \mathcal{S}_M \subset \mathcal{K} := \{1, \ldots, K\}$ to users and 2.) **powers** $p_k, k \in \mathcal{K}$, to subcarriers subject to budget $\sum_k p_k \leq P$.

- Subcarrier rate $r_{m,k}(h_{m,k}, p_k)$ is a function of the channel gain and the power. The achievable rate of user m on subcarrier k is then

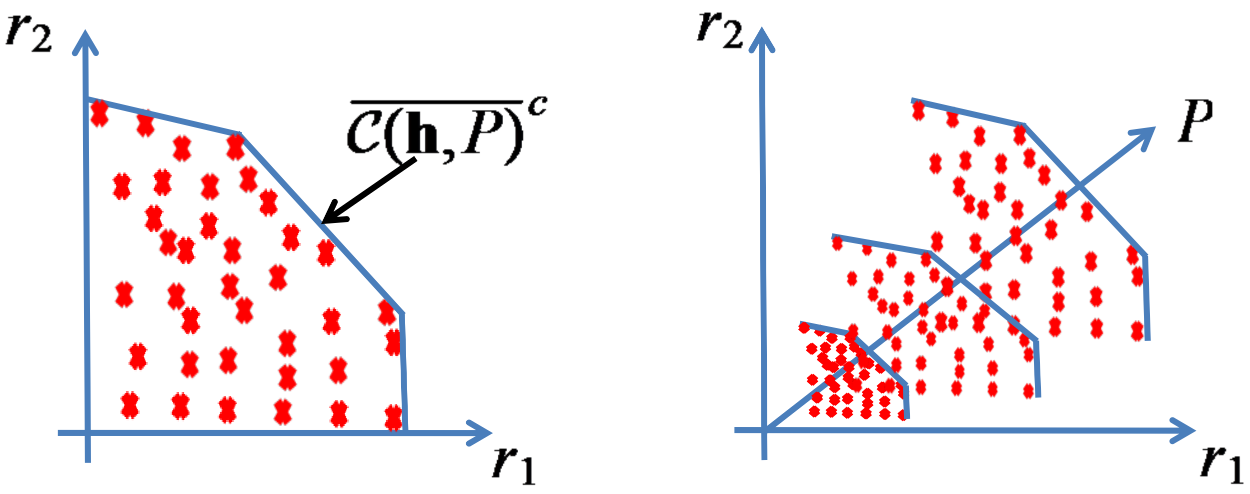
$$r_{m,k}(p_k(n)) = f(h_{m,k}, p_k) \in \{1, 2, 3, \ldots\}[Bits]$$



LTE OFDMA Downlink Channel

- Hence, the instantanous rate region $\mathcal{C}(\mathbf{h}, P)$ is a set of **discrete rate points**!

- **More general**: $\mathcal{CP}(\mathbf{h}) \subset \mathbb{R}_+^{M+1}$ is the set of rate-power tuples.

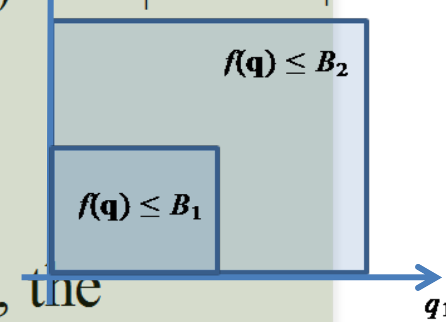


Notion of Stability

Definition 1

The queueing system is **f-stable** if there is a function $f: \mathbb{R}_+^M \to \mathbb{R}_+$ which is **unbounded in any direction** and it holds:

$$\limsup_{n \to +\infty} \mathbb{E}(\, f(\mathbf{q}(n)) \,) < +\infty$$



Choosing $f(\mathbf{q}) = \|\mathbf{q}\|$, where $\|\cdot\|$ is any vector norm, the queueing system is **strongly stable**.

Definition 2

A policy is **throughput-optimal**, if it keeps the system f-stable for any arrival rate vector $\rho \in \mathrm{int}(C_{erg}(P))$, i.e. in the **interior of the ergodic capacity region** (it is not possible to stabilize the system outside this region!).

Scheduling Policies

Definition 3

A scheduling policy $\mathcal{P}$ is a mapping from the current queue state $\mathbf{q}(n)$ **and** channel state $\mathbf{h}(n)$ to the set of rates $\mathbf{r} \in \mathcal{C}(\mathbf{h}, P)$. Denote this mapping by $\mathbf{r}^{\mathcal{P}}(\cdot,\cdot)$ we define the rate allocation **here** as:

$$\mathbf{r}^{\mathcal{P}}(\mathbf{h},\mathbf{q}) = \arg\max_{\tilde{\mathbf{r}} \in \mathcal{C}(\mathbf{h},P), \tilde{\mathbf{r}} \geq \bar{\mathbf{r}}} \left(\boldsymbol{\mu}^{\mathcal{P}}(\mathbf{h},\mathbf{q})\right)^T \bullet \tilde{\mathbf{r}}$$

i.) $\boldsymbol{\mu}^{\mathcal{P}}(\mathbf{h},\mathbf{q}) \in \mathbb{R}_+^M$ is a policy-specific weight vector which („generalized weight matching”) might depend **both on queue and channel state**.

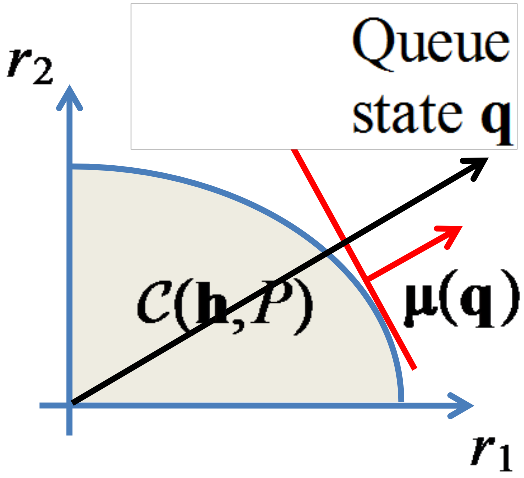
ii.) Obviously: $\mathbf{r}^{\mathcal{P}}(\mathbf{h},\mathbf{q}) \in \mathrm{bd}(\overline{\mathcal{C}(\mathbf{h},P)}^c)$.

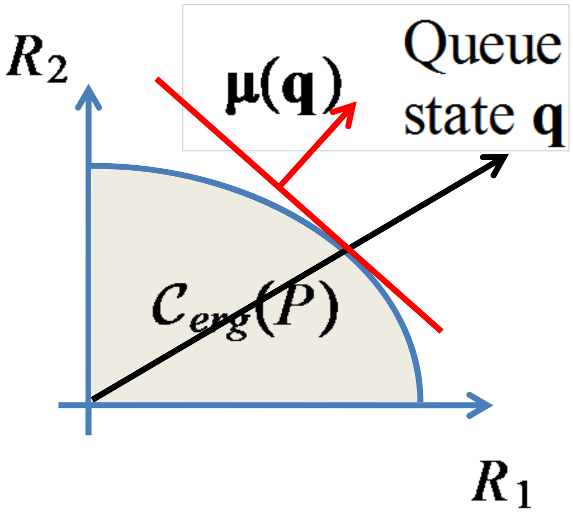
iii.) $\bar{\mathbf{r}}$ are minimum rate constraints for e.g. **H-ARQ users**.

Scheduling examples

- Maximum weight matching (MWM) scheduling: $\boldsymbol{\mu}^{\mathcal{P}}(\mathbf{q}) = \mathbf{q}$.

- Queue Proportional (QP) scheduling





Decomposition

Theorem 1

If $\|\mathbf{q}\|$ is sufficiently large, then the following is true:

i.) Any **throughput-optimal** policy **almost surely** allocates a rate point on $\mathrm{bd}(\overline{\mathcal{C}(\mathbf{h},P)}^{c})$, i.e. „generalized weight matching" is optimal.

ii.) The mapping $\boldsymbol{\mu}^{\mathcal{P}}(\mathbf{h},\mathbf{q})$ which characterizes a throughput-optimal scheduling policy is **independent** of the current channel state $\mathbf{h}$.

Universal Decomposition

MAC LAYER

Weight matching: find appropriate vector-valued mapping:

$$\boldsymbol{\mu} : \mathbb{R}_+^{M+1} \to \mathbb{R}_+^{M} : \mathbf{q} \hookrightarrow \boldsymbol{\mu}(\mathbf{q})$$

Ressource Allocation: solve

$$\mathbf{r} = \arg\max_{\tilde{\mathbf{r}} \in C(\mathbf{h},P),\, \tilde{\mathbf{r}} \geq \bar{\mathbf{r}}} \boldsymbol{\mu}^T \cdot \tilde{\mathbf{r}}$$

for given $\boldsymbol{\mu}$ and rate/power constraints $\bar{\mathbf{r}}/P$.

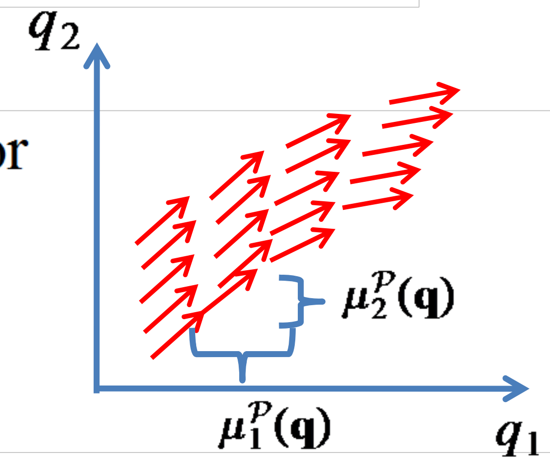
When is a weight matching policy throughput-optimnal?

Can we solve the optimization problem efficiently?

Curl-free Scheduling fields

Main Theorem

- First of all, observe that $\bar{\mu}_1(\mathbf{q}), \bar{\mu}_2(\mathbf{q}), \ldots, \bar{\mu}_M(\mathbf{q})$ defines an M-dimensional vector (**scheduling**) **field**.



- Without loss of generality, the weight vector can be normalized:

$$\bar{\boldsymbol{\mu}}^{\mathcal{P}}(\mathbf{q}) := \frac{\boldsymbol{\mu}^{\mathcal{P}}(\mathbf{q})}{\|\boldsymbol{\mu}^{\mathcal{P}}(\mathbf{q})\|_1}$$

- **Note that not** all policies are feasible! **Counterexample**: E.g. the function $\mu_i(q_i) = e^{q_i}$ is not feasible (only known by simulations so far but we have shown in our recent paper).

  So, what is the common of all policies such as MWM, QP etc.?

The Main Theorem

Main Theorem

The scheduling policy $\mathcal{P}$ is throughput-optimal, if the mapping $\bar{\boldsymbol{\mu}}^{\mathcal{P}}$ fulfills the following two conditions:

i.) Let $\|\boldsymbol{\Delta}\mathbf{q}\| \leq C_1$, then:

$$\lim_{\|\mathbf{q}\|\to+\infty,\text{ any path in }\mathbb{R}_+^M} \bar{\mu}_m(\mathbf{q}+\boldsymbol{\Delta}\mathbf{q}) = \lim_{\|\mathbf{q}\|\to+\infty} \bar{\mu}_m(\mathbf{q})$$

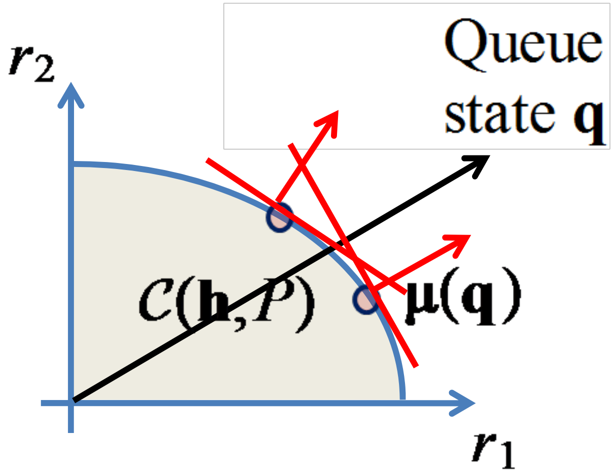
ii.) Let $q_m \leq C_2$, then:

$$\lim_{\|\mathbf{q}\|\to+\infty,\text{ any path in }\mathbb{R}_+^M, q_{m\leq}C_2} \bar{\mu}_m(\mathbf{q}) = 0$$

Main Theorem: Interpretation

- If $\|\mathbf{q}\|$ becomes large, the weight vector varies smoothly between two time slots.

- If $\|\mathbf{q}\|$ becomes large, no rate is wasted on "nonurgent" users.



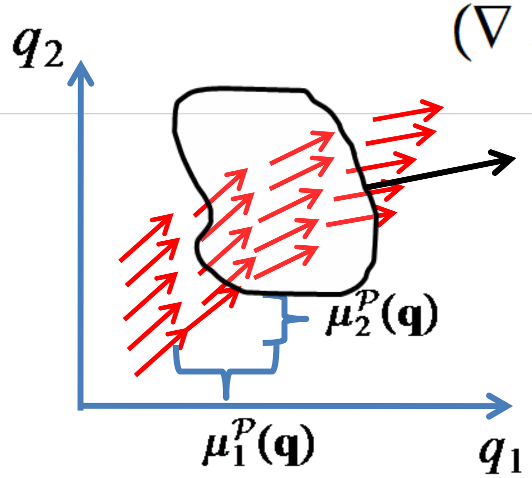
Main Theorem: Proof Sketch

- **Suppose** that there are unbounded functions $V(\mathbf{q}), f(\mathbf{q}) : \mathbb{R}_+^M \rightarrow \mathbb{R}_+$ so that:

$$\frac{\partial V(\mathbf{q})}{\partial q_i} = f(\mathbf{q})\bar{\mu}_i(\mathbf{q})$$

- If so $\bar{\mu}(\mathbf{q})$ **must satisfy** the conditions of the **Poincaré Lemma**, i.e. $\bar{\mu}(\mathbf{q})$ is a continuous, **totally integrable** function, e.g. in 3 dimensions:

$$(\nabla \times \bar{\boldsymbol{\mu}}(\mathbf{q})) = \mathrm{curl}(\bar{\boldsymbol{\mu}}(\mathbf{q})) = \mathbf{0}$$



All line integrals along lines are zero: a curl-free scheduling field!

Main Theorem: Proof Idea

- The first part of the proof shows: if $\bar{\boldsymbol{\mu}}$ is integrable then for some constants $\theta, B > 0$ the so-called **Lyapunov drift** becomes:

$$\mathbf{E}(V(\mathbf{q}(n+1)) - V(\mathbf{q}(n))|\mathbf{q}(n)) \leq -\theta f(\mathbf{q}),$$

$$\forall \|\mathbf{q}\| > B$$

This implies: The Markov chain is f-stable [see e.g. Meyn 1992].

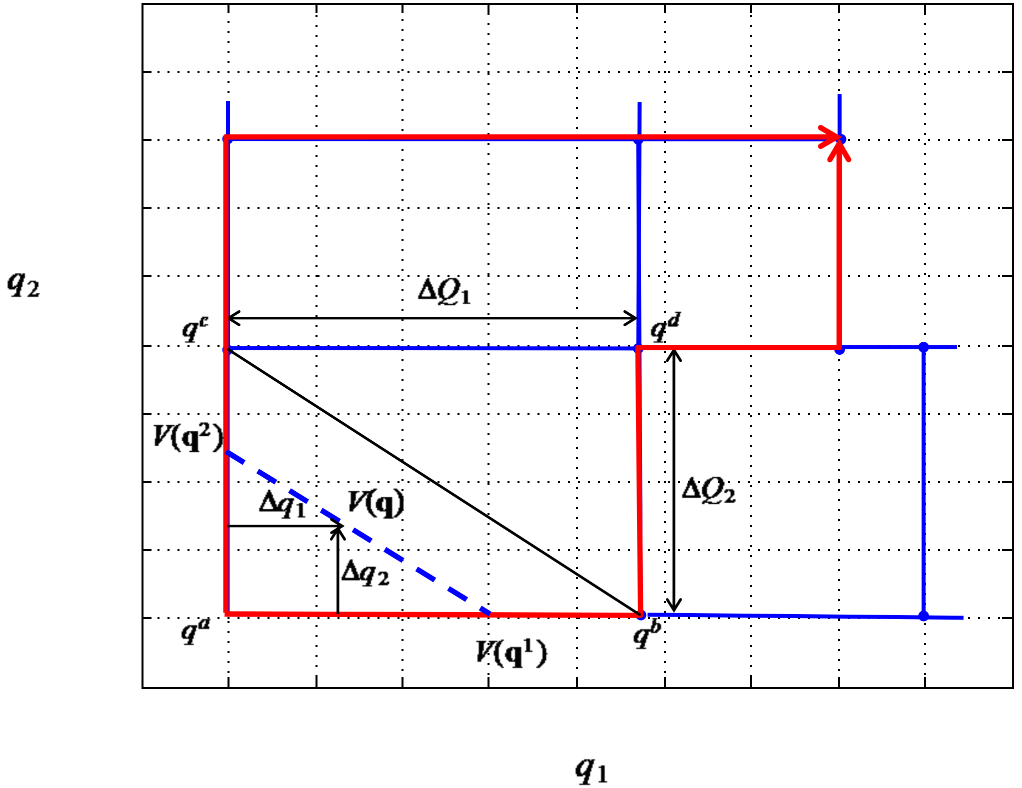
- **BUT**, even MWM scheduling does not fulfill Poincaré's Lemma!!

- Hence, in the second part we show: if $\bar{\boldsymbol{\mu}}(\mathbf{q})$ fulfills the condition of the theorem it can be arbitralily closely approximated by some integrable function constructed as follows:

Main Theorem: Proof Sketch

First, integrable continuation is constructed along grid lines ....

Then, we ensure independence of path as well!



Finally, it is shown that the difference between continuation and original scheduling becomes arbitrarily small.

Resource Allocation

Recall:

Ressource Allocation: solve

$$\mathbf{r} = \arg\max_{\tilde{\mathbf{r}} \in C(\mathbf{h},P), \tilde{\mathbf{r}} \geq \bar{\mathbf{r}}} \boldsymbol{\mu}^T \cdot \tilde{\mathbf{r}}$$

for given $\boldsymbol{\mu}$ and rate/power constraints $\bar{\mathbf{r}}/P$.

Resource Allocation

- **Obviously**: Resource allocation problem is combinatorial problem in $\mathcal{S}_1, \ldots, \mathcal{S}_M$: **brute force prohibitive** when K is large!

- **Trick**: Solution is **forced** to lie on $\mathrm{bd}(\overline{\mathcal{CP}(\mathbf{h})}^c)$; introducing **power prize** $\lambda \in \mathbb{R}_+$ and **user revenues** $\mu'_m \in \mathbb{R}_+$ the maximization problem can be written as:

$$\max_{\mathbf{p} \in \mathbb{R}_+^K, \mathcal{S}_1, \ldots, \mathcal{S}_M} \sum_{m=1}^{M} (\mu'_m + \mu_m) \sum_{k \in \mathcal{S}_m} r_{m,k}(p_k) - \lambda \sum_{k=1}^{K} p_k$$

- Here, λ and μ'_m **ensure** that:

$$\sum_{k=1}^{K} p_k \le P, \quad \sum_{k \in \mathcal{S}_m} r_{m,k}(p_k) \ge \bar{r}_m \ \forall m$$

for some given power budget P and rate constraints $\bar{r}_m, \forall m$.

Resource Allocation

- **Observation**: The problem **decouples** into K independent problems even for the our combinatorial problem.
- **Idea**: Find smallest possible λ, μ'_m such that constraints are fulfilled.
- This opens up an efficient way to solve the combinatorical problem by viewing it as a (non-standard) "ressource allocation game".

Resource Allocation

Rate of user 1 when all other weights are fixed!

$M+1$ players resource allocation game:

Player 1: $\min \mu'_1$ s.t. $\sum_k r_{1,k}^{(\mu'_1,\mu'_{-1})} \geq \bar{r}_1$

Player 2: $\min \mu'_2$ s.t. $\sum_k r_{2,k}^{(\mu'_2,\mu'_{-2})} \geq \bar{r}_2$

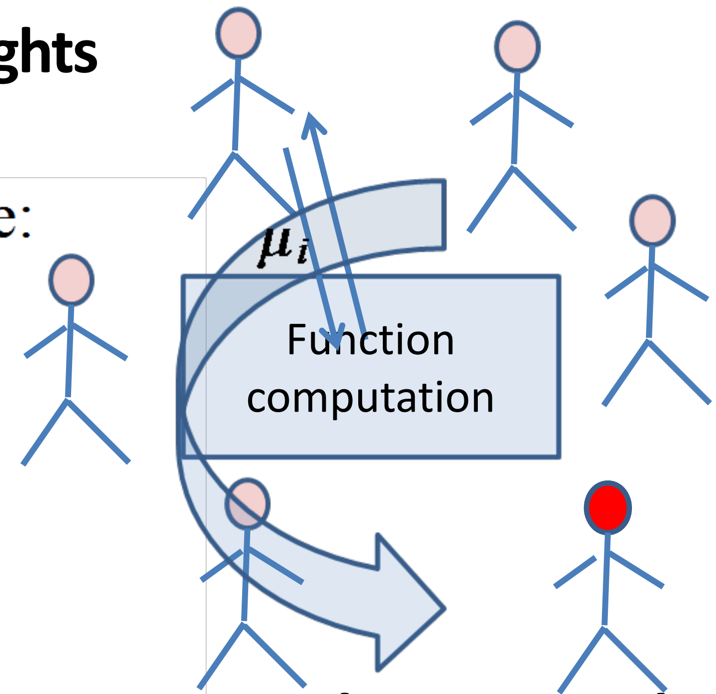
.

.

.

Player M: $\min \mu'_M$ s.t. $\sum_k r_{1,k}^{(\mu'_M,\mu'_{-M})} \geq \bar{r}_1$

Power Player $M+1$: $\min(-\lambda)$ s.t. $\sum_k p_k \geq \bar{P}$



Asynchronous updates!

Resource Allocation

Theorem 2

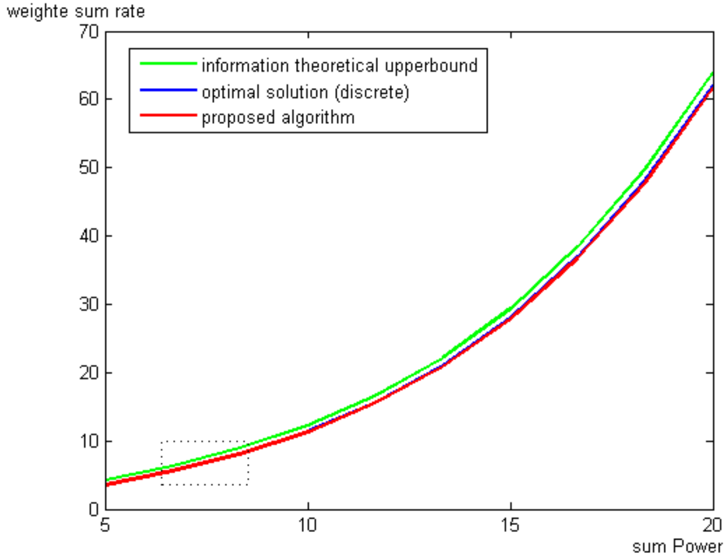
The outcome can be characterized as follows:

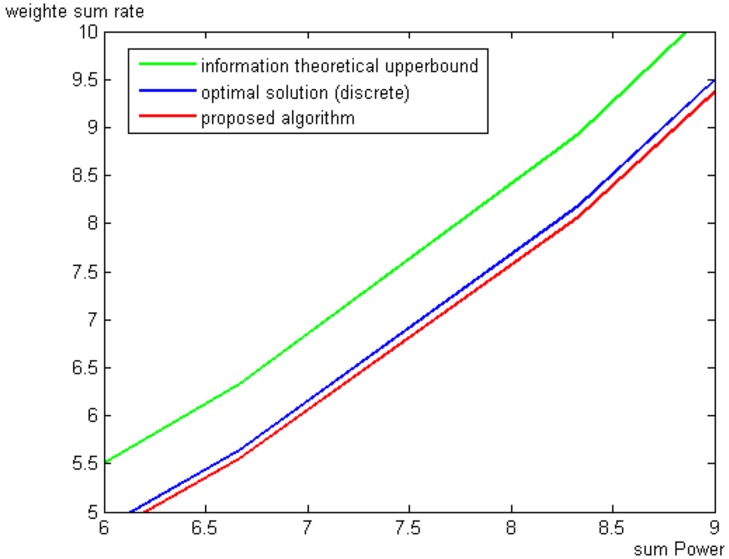
The sequence $\boldsymbol{\mu}^{(n)}, \lambda^{(n)}$ generated with asynchronous updates of this ressource allocation game converges to a **smallest** (Pareto-optimal) solution $\boldsymbol{\mu}^*$ such that the rate constraints are satisfied.

Note: Proof is based on formulating the update rule as an operator which carries, interestingly, properties of an interference function [Yates 95].

Resource Allocation

- Number of users : 5, 1000 channel runs
- Number of subcarriers : 256
- μ = [0.1, 0.1, 0.2, 0.2, 0.4]
- Min. Rate constraint: [3, 3, 2, 1, 0]





Note: 3dB loss in average compared to standard utility optimization!

Conclusions with Outlook:

We have presented a invaluable example of applying successfully queuing-, information- and optimization theory to solve a fundamental problem.

Research is only the beginning: What about:
i.) non-ergodic processes ii.) past-dependent policies
iii.) non-cooperative scheduling in multicell scenarios

We want emphasize two cases:
- MIMO: Even per subcarrier computation is infeasible
  (new patent filed, graph theoretic approaches!)
- Networks: MWM appears naturally in networks with flow control; framework can be applied?