Hybrid Scheduling in Heterogeneous Half- and Full-Duplex Wireless Networks

Tingjun Chen[*], Jelena Diakonikolas[†], Javad Ghaderi[*], and Gil Zussman[*]

[*]Electrical Engineering, Columbia University

[†]Computer Science, Boston University

IEEE INFOCOM

Apr. 17, 2018

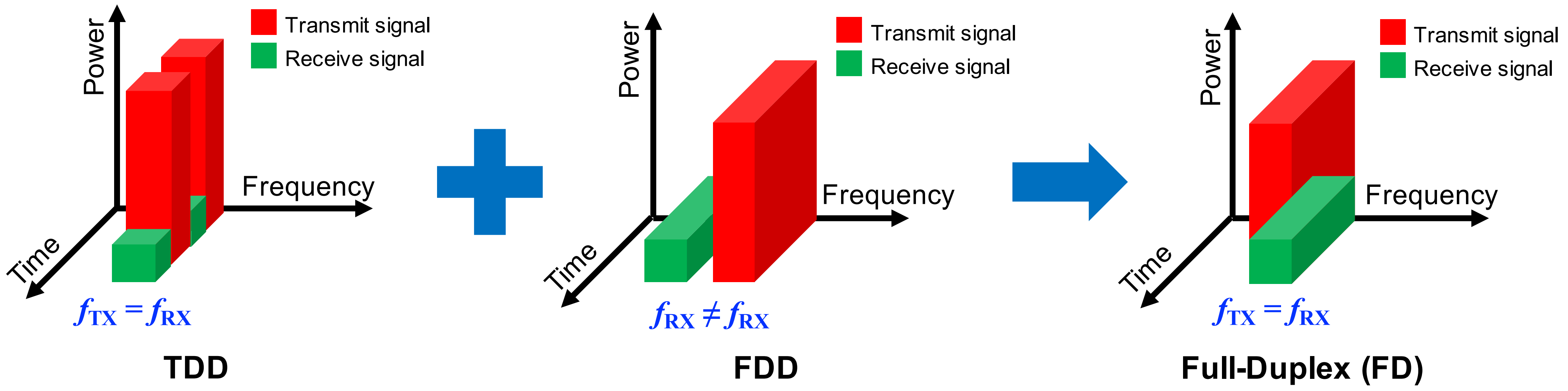
COLUMBIA UNIVERSITY
IN THE CITY OF NEW YORK

wim.net
Wireless & Mobile Networking Lab

Full-Duplex Wireless

- Legacy half-duplex wireless systems separate transmission and reception in either:
  - Time: Time Division Duplex (TDD)
  - Frequency: Frequency Division Duplex (FDD)

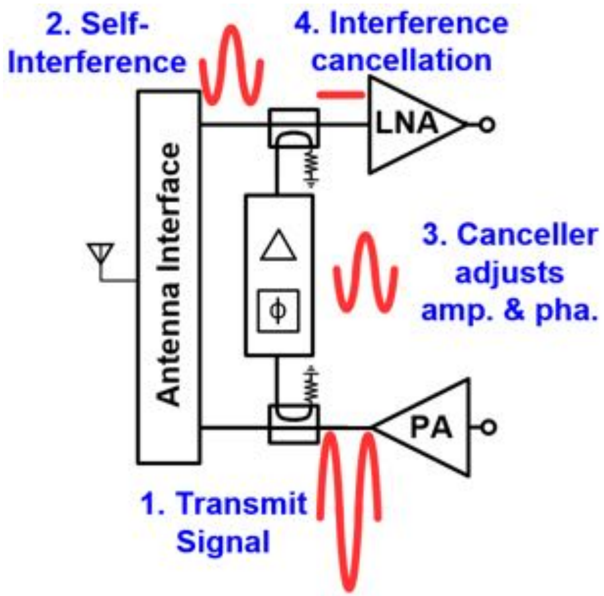
- (Same channel) Full-duplex communication: simultaneous transmission and reception on the same frequency channel

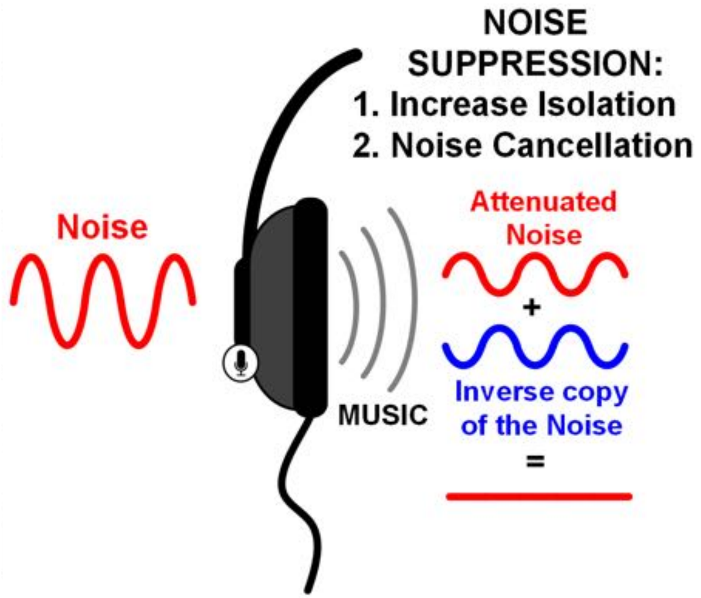


Full-Duplex Wireless

- Benefits of full-duplex wireless:
  - Increased system throughput and reduced latency
  - More flexible use of the wireless spectrum and energy efficiency

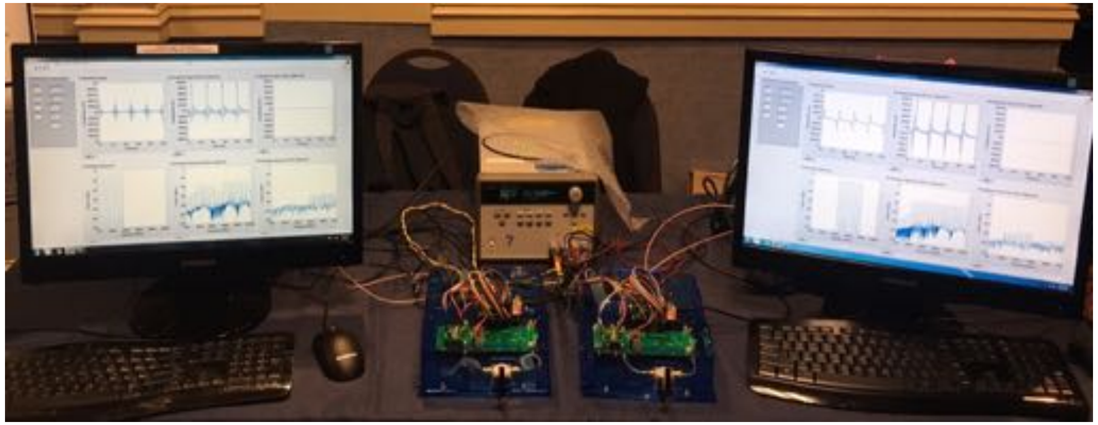
- Viability is limited by self-interference
  - Transmitted signal is **billions** of times (**10^9 or 90dB**) stronger than the received signal
  - Requiring extremely powerful self-interference cancellation



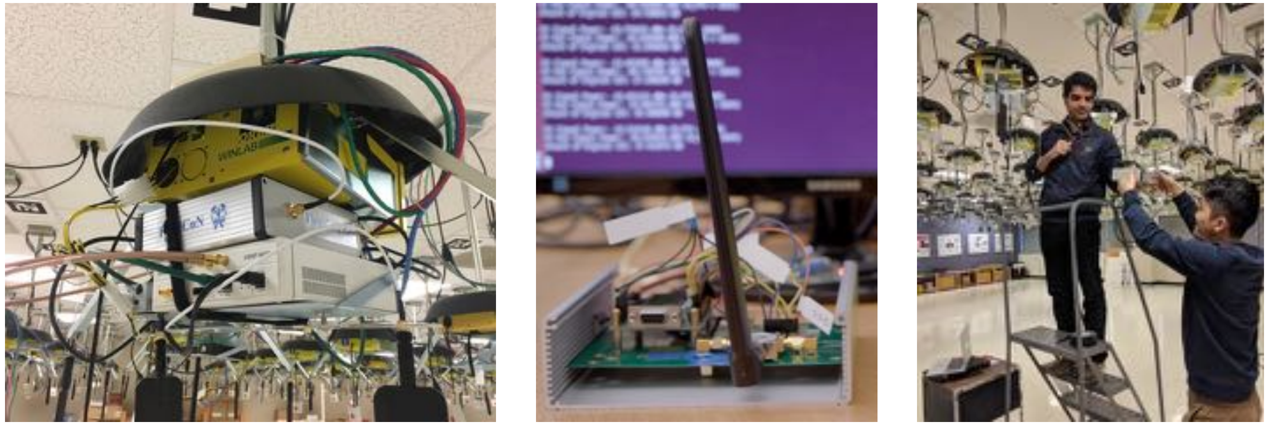


The Columbia FlexICoN Project

- **F**ull-Dup**lex** Wireless: From **I**ntegrated **C**ircuits t**o** **N**etworks (FlexICoN)
  - Development of full-duplex transceiver/system, algorithm design, experimental evaluation, etc.
  - Integration of full-duplex capability with the open-access ORBIT testbed



Gen-2 wideband full-duplex link
(Demo at INFOCOM'17)

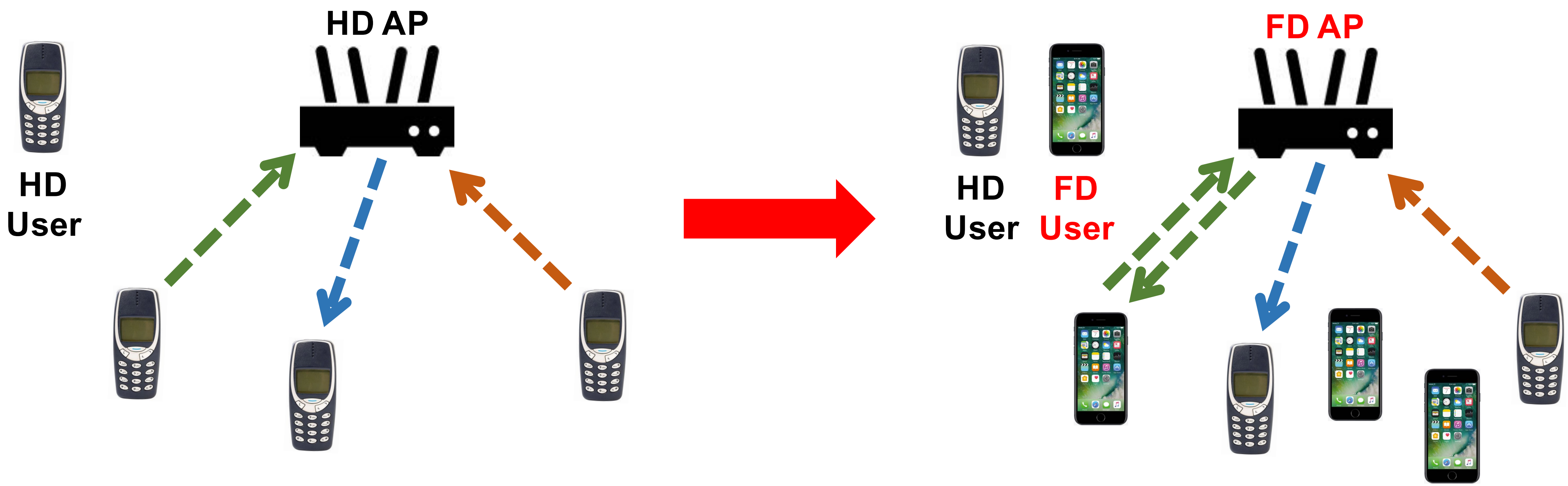


A programmable Gen-1 full-duplex node installed in ORBIT
(Demo Session 2 on Wed. at 9:30am in Palace Lounge)

  - Future integration with the ***PAWR COSMOS*** city-scale testbed (NSF PAWR Session on Wed. at 15:30pm in Tapa 1)

Motivation

- Gradual replacement and introduction of full-duplex (FD) devices into legacy half-duplex (HD) networks



- ***Goal***: Develop ***efficient*** and ***fair*** scheduling algorithms in such ***heterogeneous*** half-duplex and full-duplex networks with performance guarantees

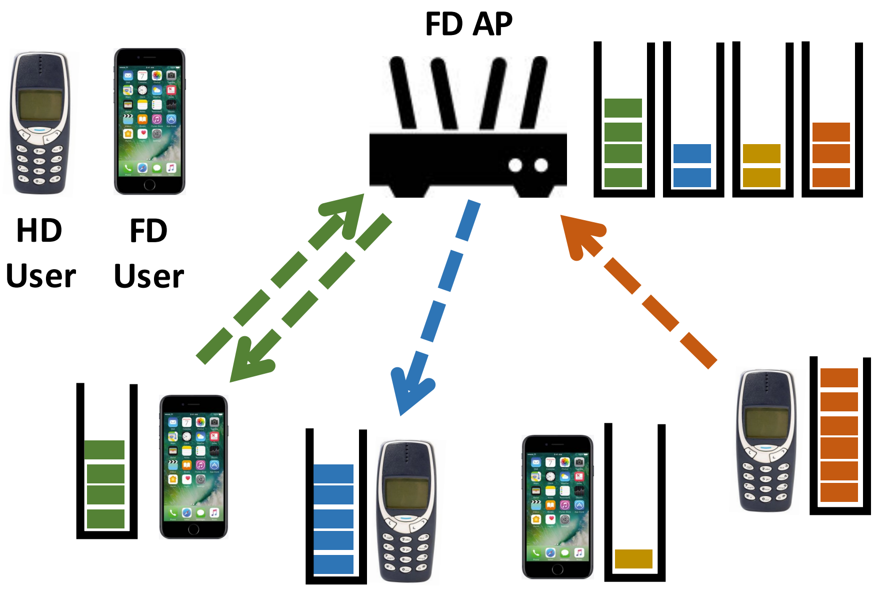
Related Work

- Full-duplex radio/system design
  - Laboratory bench-top design: [Choi et al. 2010], [Duarte & Sabharwal, 2010], [Aryafar et al. 2012], [Bharadia et al. 2013/2014], [Kim et al. 2013/2015], [Korpi et al. 2016], [Sayed et al. 2017]
  - Integrated circuits (small form-factor) design: [Zhou et al. 2014/2015], [Debaillie et al. 2015], [Yang et al. 2015], [Reiskarimian et al. 2016/2017], [Zhang et. al 2017/2018]
- Throughput gains from full-duplex:
  - [Xie & Zhang, 2014], [Nguyen et al. 2014], [Korpi et al. 2015], [Marasevic et al. 2017/2018]
- Cellular/WiFi scheduling:
  - [Duarte et al. 2014], [Yang & Shroff, 2015], [Alim et al. 2016], [Chen et al. 2015/2016], [Goyal et al. 2016/2017]
- CSMA/Scheduling in legacy half-duplex networks:
  - CSMA, Max-Weight, Greedy-Maximal, Longest-Queue-First, Q-CSMA, etc. [Kleinrock & Tobagi, 1975], [Tassiulas & Ephremides 1992], [Dimakis & Walrand, 2006], [Brzezinski et al. 2006], [Ni et al. 2012], [Birand et al. 2012], etc.

- ***Heterogeneous*** networks with both half- and full-duplex users were not considered
- **Fairness** between half- and full-duplex users was not considered
- Very little work provided ***performance guarantees*** (e.g., throughput optimality)

Model

- Time is slotted ($t = 1, 2, \ldots$)
- A single-channel, collocated, ***heterogeneous*** network with one access point (AP) and N users:
  - The AP and N_F users are full-duplex (FD)
  - $N_H = N - N_F$ users are half-duplex (HD)



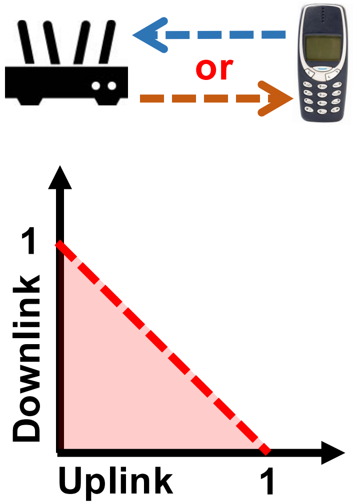
A heterogeneous network with $N_F = N_H = 2$

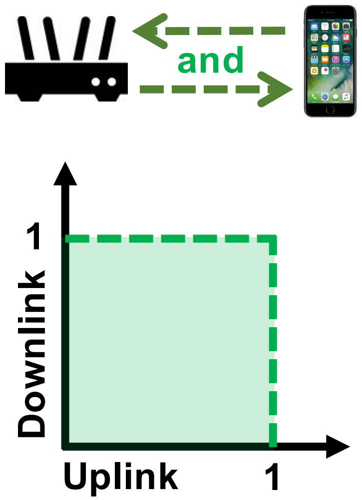
- N downlink queues at the AP and one uplink queue at each user
  - The AP has information about ***all downlink queues***
  - A user has information about ***only its uplink queue***
- Unit link capacity and perfect self-interference cancellation
- ***Feasible schedules***: a single half-duplex uplink or downlink, or a pair of full-duplex uplink and downlink
- A pair of full-duplex uplink and downlink are always scheduled at the same time

Problem Formulation

- Capacity Region: Convex hull of all feasible schedules
- For a legacy half-duplex user: $\lambda_{\text{uplink}} + \lambda_{\text{downlink}} \leq 1$
- For a full-duplex user:

$$\left.\begin{array}{r} \lambda_{\text{uplink}} \leq 1 \\ \lambda_{\text{downlink}} \leq 1 \end{array}\right\} \max\{\lambda_{\text{uplink}}, \lambda_{\text{downlink}}\} \leq 1$$



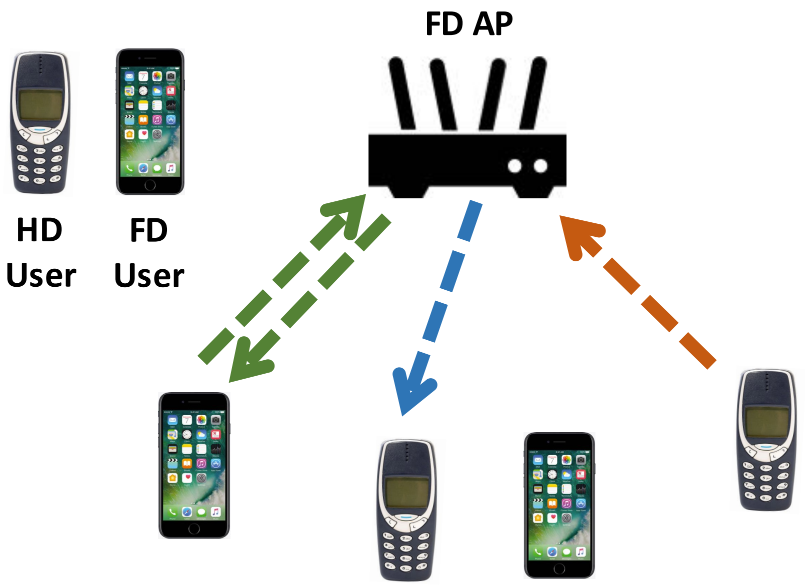


- A scheduling algorithm is ***throughput-optimal*** if it can keep the network queues stable for all arrival rate vectors in the interior of the capacity region

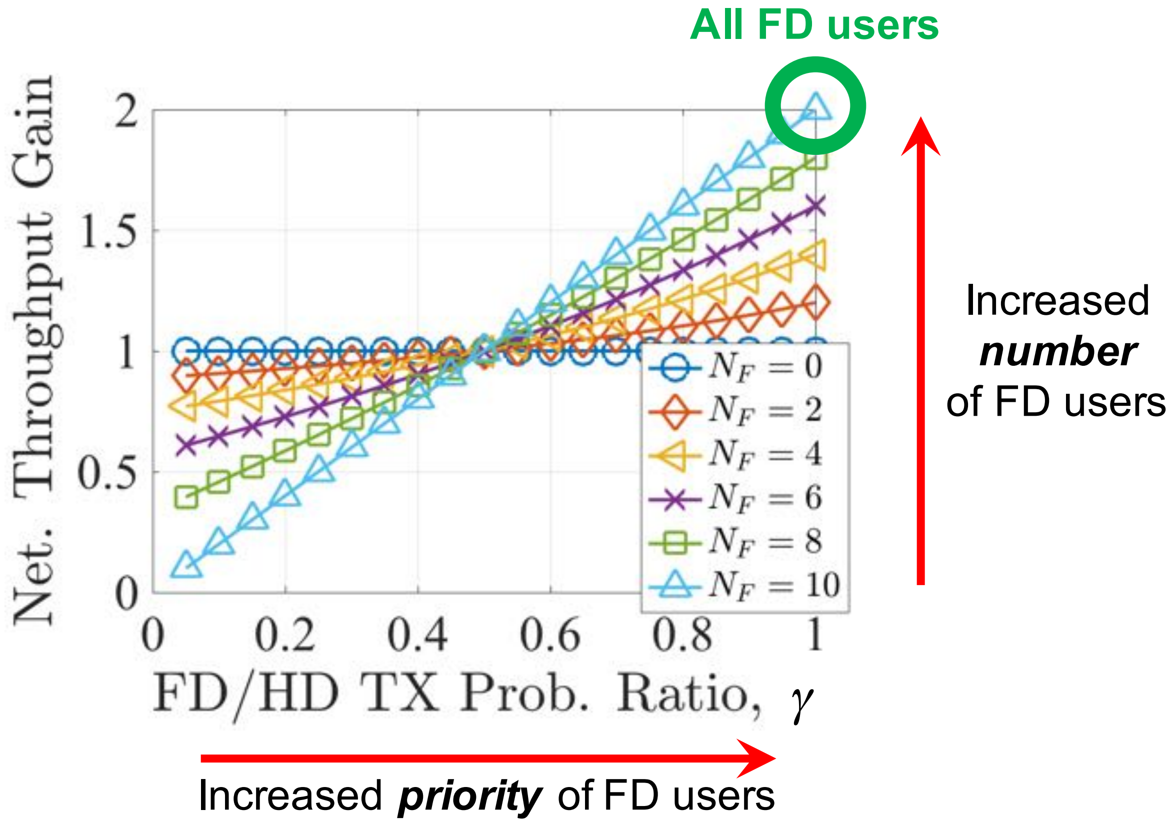
- ***Goal***: Achieve maximum throughput in networks with ***heterogeneous*** half-duplex and full-duplex users in a distributed manner, while being ***fair*** to all the users and having favorable delay performance
- ***Solution***: ***H-GMS*** – A Hybrid scheduling algorithm that combines centralized Greedy Maximal Scheduling (GMS) and distributed Q-CSMA

Introducing Full-Duplex Users – Everyone Gains!

- A ***homogeneous*** network with $N = 10$ half-duplex users **vs.** A ***heterogeneous*** network with N_H half-duplex users and N_F full-duplex users ($N_H + N_F = N = 10$)
- Consider the a static CSMA algorithm with fixed transmission probabilities p_H and p_F for half-duplex and full-duplex users. Let $p_F = \gamma\, p_H$ with $\gamma \in (0, 1]$
- With $p_H = 0.5$, throughput gain of the ***network***:



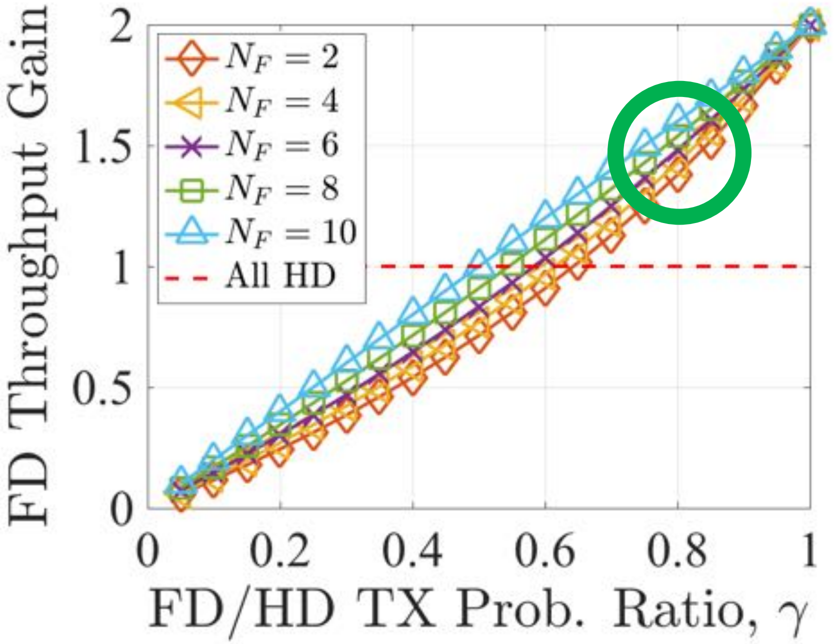
A heterogeneous network
with fixed N and varying N_F



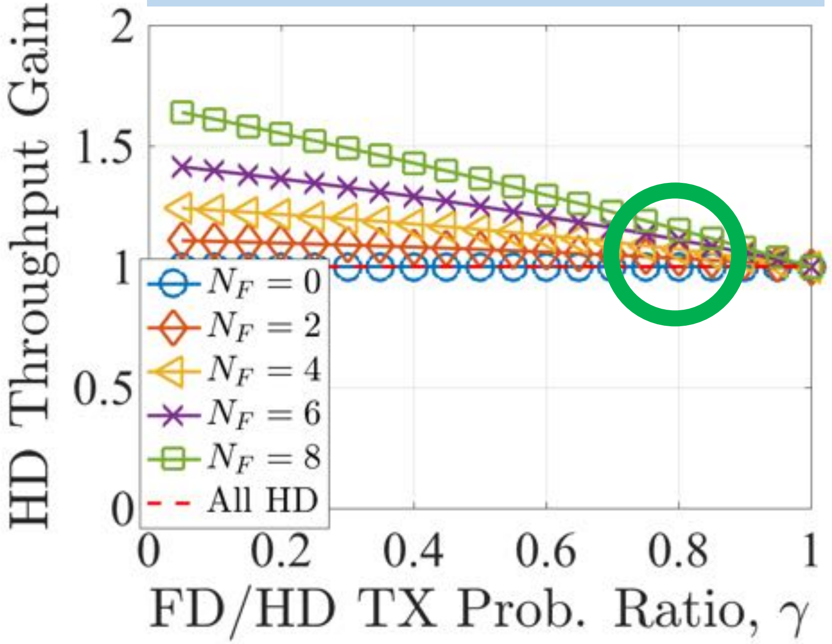
Introducing Full-Duplex Users – Everyone Gains!

- A ***homogeneous*** network with $N = 10$ half-duplex users **vs.** A ***heterogeneous*** network with N_H half-duplex users and N_F full-duplex users ($N_H + N_F = N = 10$)
- Consider the a static CSMA algorithm with fixed transmission probabilities p_H and p_F for half-duplex and full-duplex users. Let $p_F = \gamma\, p_H$ with $\gamma \in (0, 1]$
- With $p_H = 0.5$, throughput gain of ***individual users***:

Even half-duplex users can gain!



Increased ***priority*** of FD users



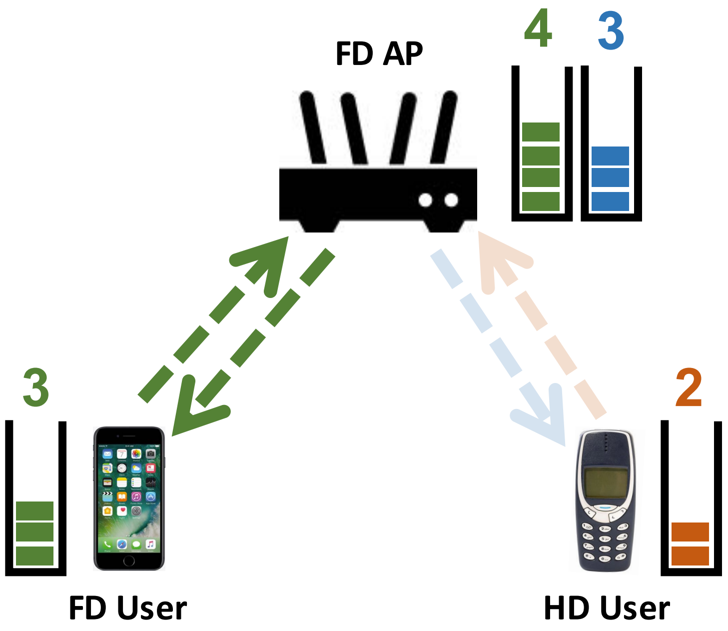
Increased ***priority*** of FD users

Increased ***number*** of FD users

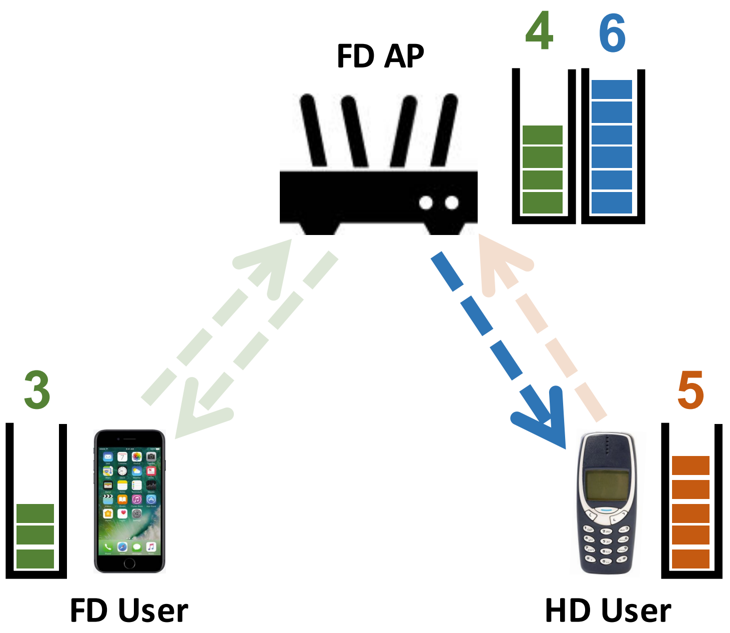
Scheduling Algorithms

- Max-Weight Scheduling (MWS) is throughput-optimal
  - Q-CSMA can be applied

- What about the Greedy Maximal Scheduling (GMS)?
  - The returned schedule may not be Max-Weight



MWS = GMS



MWS ≠ GMS

Scheduling Algorithms

- Max-Weight Scheduling (MWS) is throughput-optimal
  - Q-CSMA can be applied

- What about the Greedy Maximal Scheduling (GMS)?
  - The returned schedule may not be Max-Weight

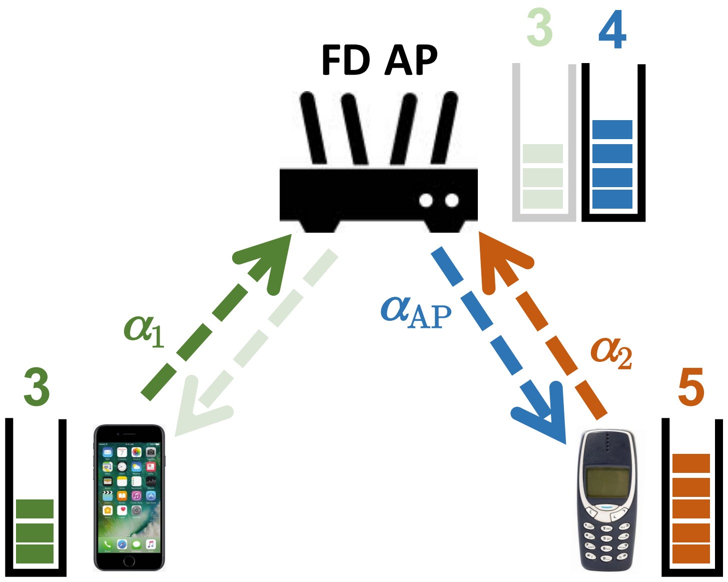
- ***Proposition***: The centralized Greedy Maximal Scheduling (GMS) algorithm is throughput-optimal in ***any*** collocated ***heterogeneous*** half-duplex and full-duplex networks
  - Proof is based on local-pooling

- ***Question***: How to achieve GMS is a distributed manner?
- ***Solution***: ***H-GMS*** – a **H**ybrid scheduling algorithm that combines centralized **GMS** and distributed Q-CSMA

Proposed Algorithm: H-GMS in slot t

If the previous slot is an ***idle*** slot:

- **Step 1: Initiation** (centralized GMS at the AP)
  - The AP selects the downlink with the longest queue
  - The AP draws an ***initiator link*** from all the uplinks and the selected downlink according to an access probability distribution α



Step 1

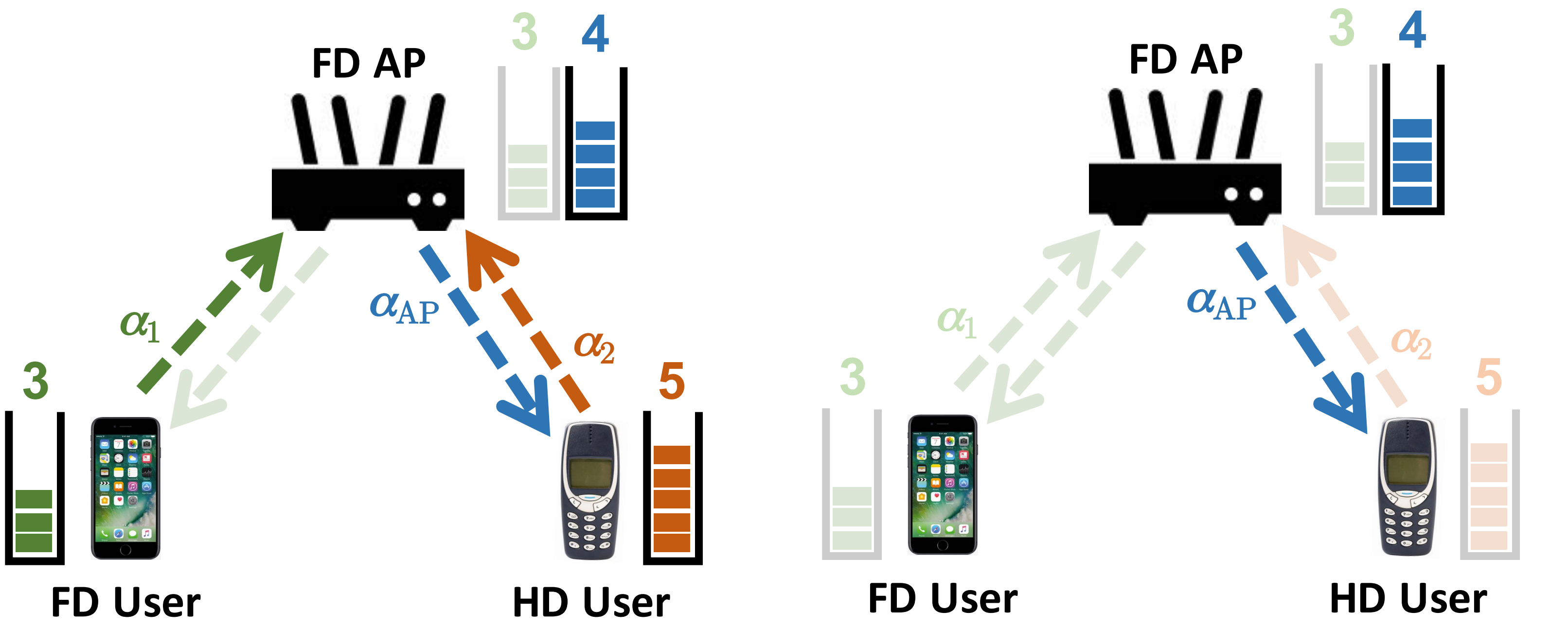
Proposed Algorithm: H-GMS in slot t

If the previous slot is an ***idle*** slot:

- **Step 2: Coordination** (distributed Q-CSMA)
  - If link l is selected as the initiator link, it is activated w.p. $p(Q_l(t))$

Transmission probability and weight functions $f(Q(t))$

$$p(Q(t)) = \frac{\exp(f(Q(t)))}{1+\exp(f(Q(t)))}$$



Step 1

Step 2: if the HD downlink is selected

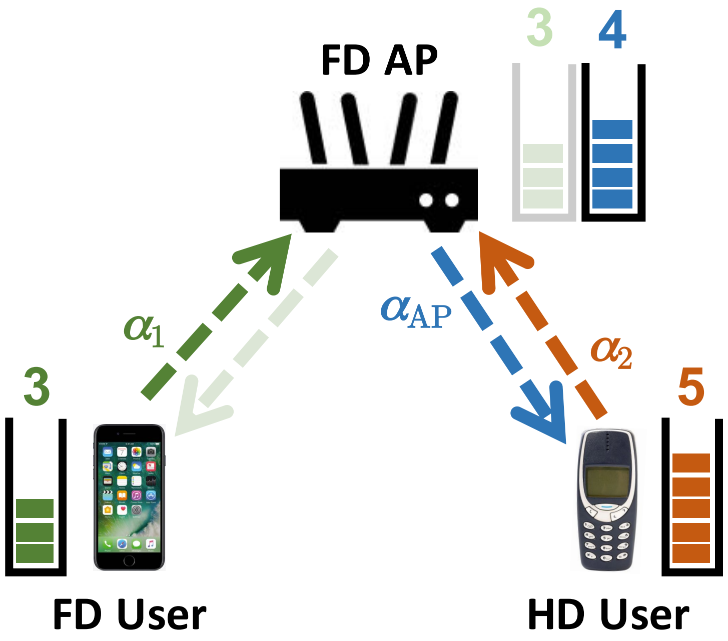
Proposed Algorithm: H-GMS in slot t

Transmission probability and weight functions $f(Q(t))$

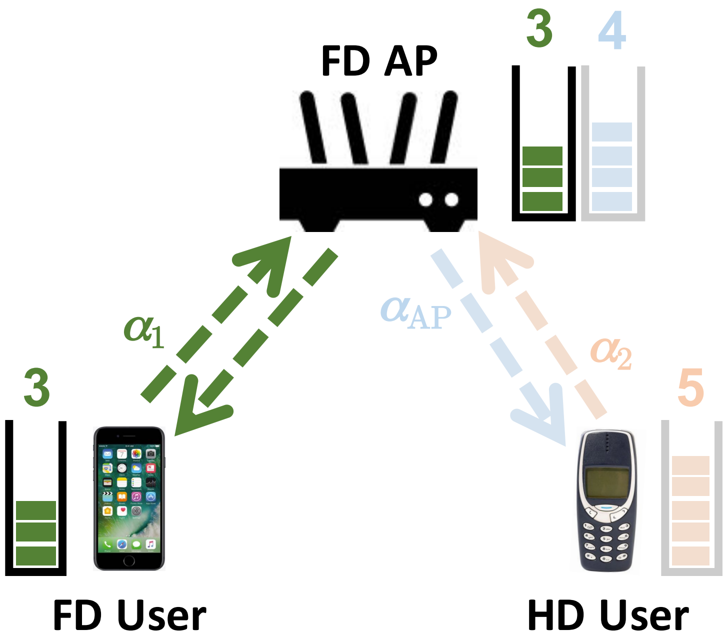
$$p(Q(t)) = \frac{\exp(f(Q(t)))}{1+\exp(f(Q(t)))}$$

If the previous slot is an ***idle*** slot:

- **Step 2: Coordination** (distributed Q-CSMA)
  - If link l is selected as the initiator link, it is activated w.p. $p(Q_l(t))$
  - If the initiator link is a full-duplex uplink (downlink), the corresponding downlink (uplink) will also be activated



Step 1

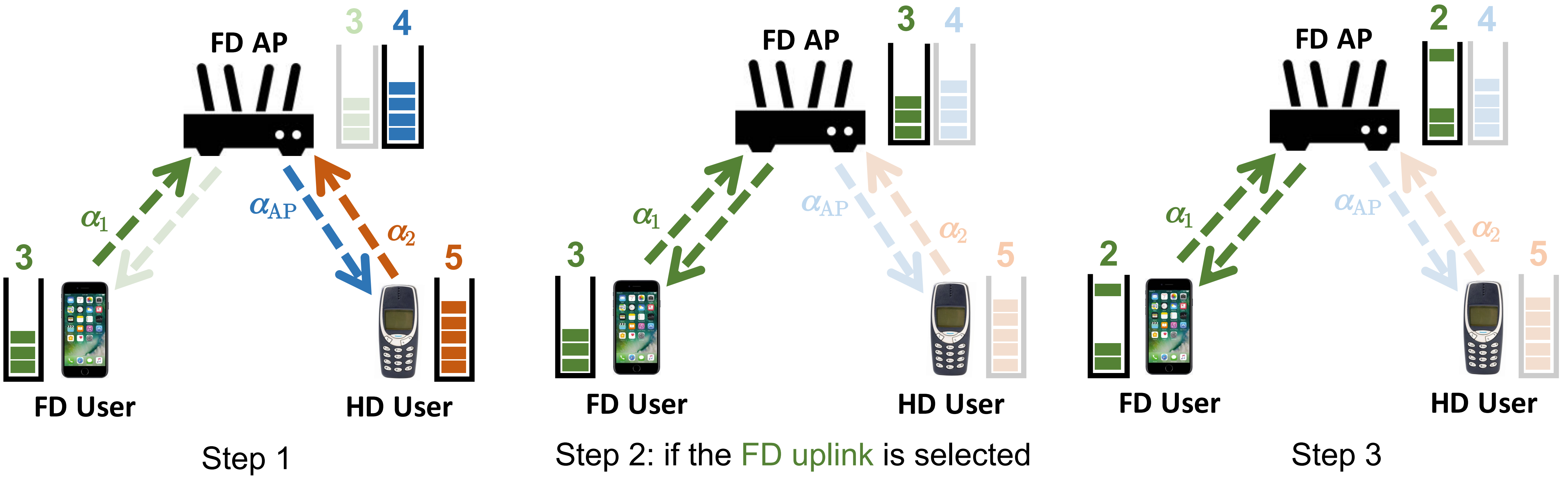


Step 2: if the FD uplink is selected

Proposed Algorithm: H-GMS in slot t

If the previous slot is an ***idle*** slot:

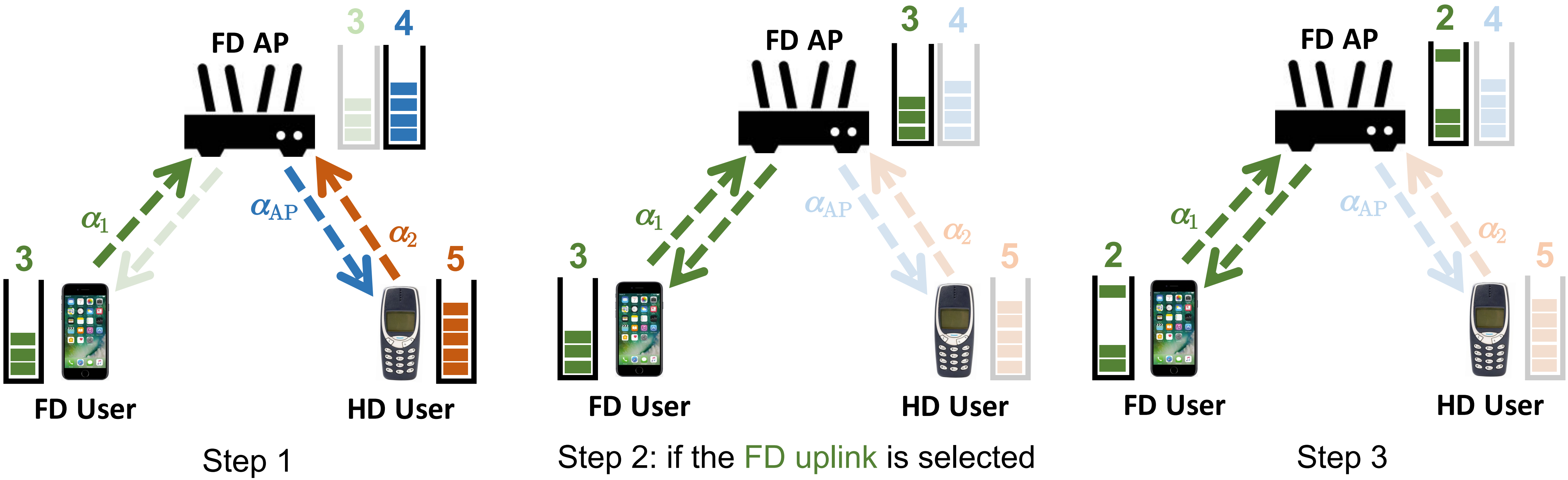
- **Step 3: Transmission**
  - One packet is transmitted on each activated link



Proposed Algorithm: H-GMS in slot t

If the previous slot is a ***busy*** slot:

- The AP keeps the same initiator link and repeats steps 2 & 3

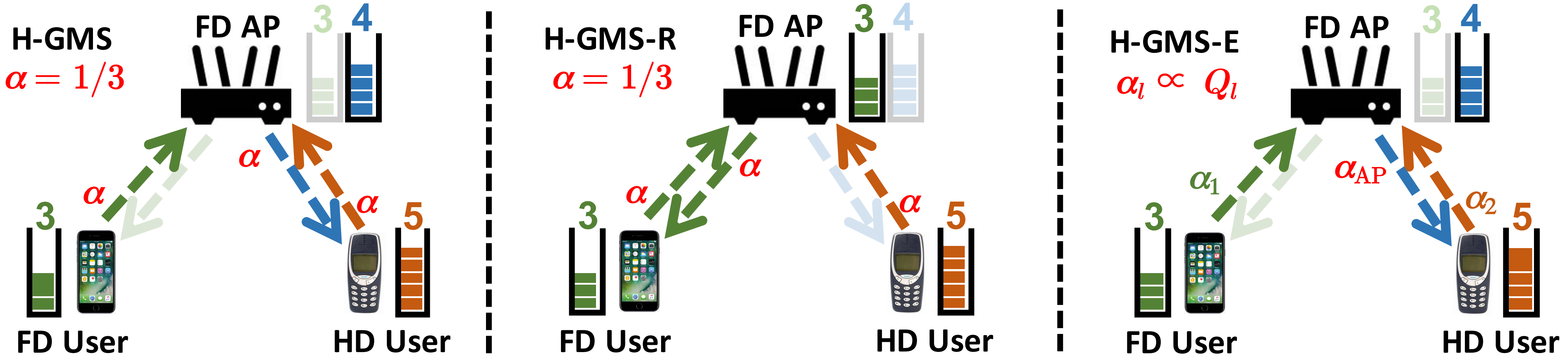


Main Results

- ***Theorem***: For any arrival rate vector inside the capacity region, the system Markov chain $(\boldsymbol{X}(t), \boldsymbol{Q}(t))$ is positive recurrent under the H-GMS algorithm. The weight function f can be any nonnegative increasing function such that $\lim_{x\to\infty} f(x)/\log(x) < 1$ or $\lim_{x\to\infty} f(x)/\log(x) > 1$.
  - Proof is based on fluid limit analysis

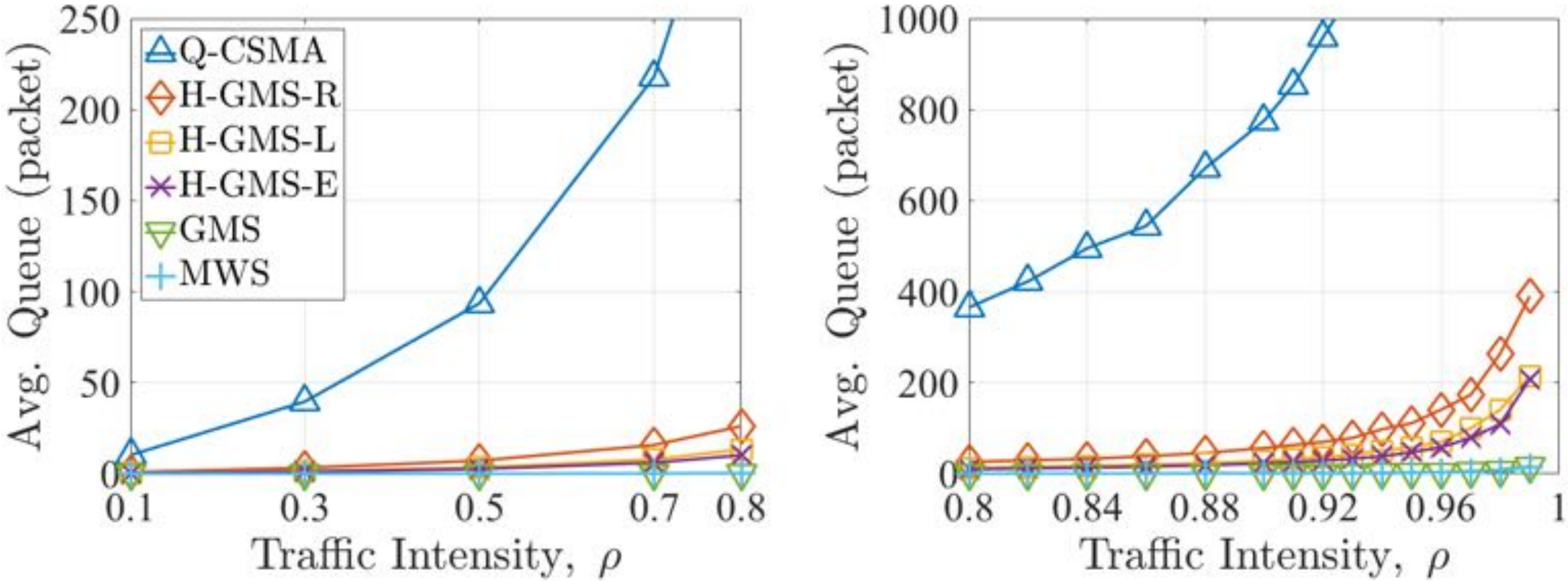
$$p(Q(t)) = \frac{\exp(f(Q(t)))}{1+\exp(f(Q(t)))}$$

- Variants of ***H-GMS***:
  - ***H-GMS*** (or ***H-GMS-L***)
  - ***H-GMS-R***: the AP selects a downlink queue uniformly at Random, α is uniformly distributed
  - ***H-GMS-E***: the AP selects the downlink with the longest queue, α is proportional to the Estimated uplink queues



Performance Evaluation – Queue Length

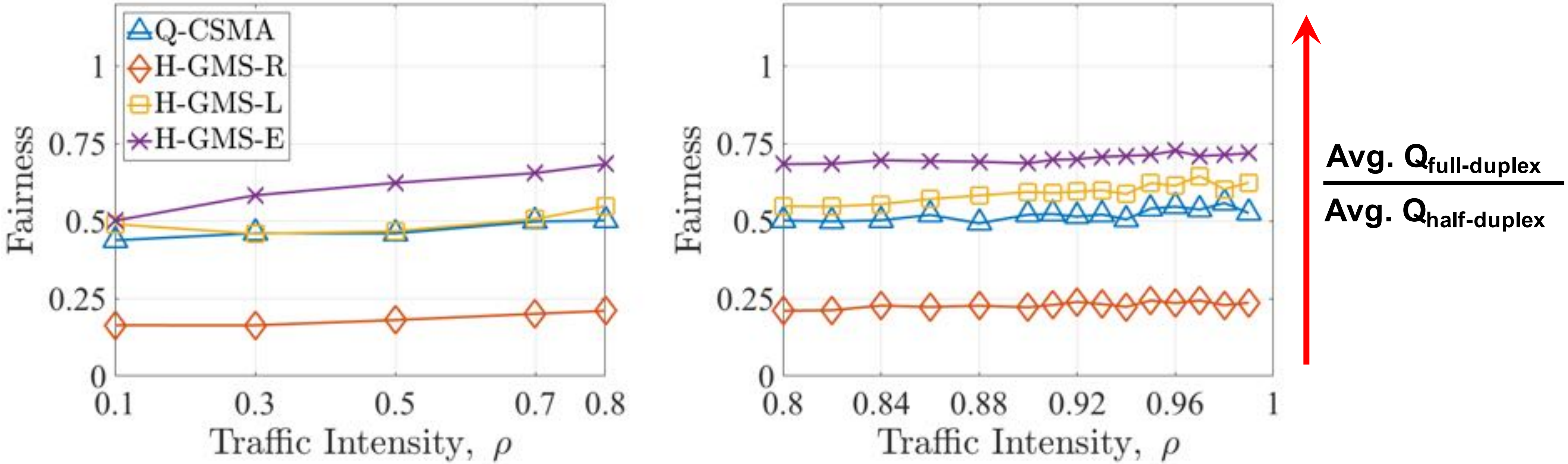
- Simulations with $N = 10$ users with $N_F = N_H = 5$ in a heterogeneous network
- Equal arrival rate on all the uplinks and downlinks with total arrival rate $\rho \in (0, 1]$
- Average queue length (packet) for every link



The largely reduced queue length resulted from (i) utilizing the centralized downlink queue information at the AP, and (ii) the introduction of full-duplex users

Performance Evaluation – Fairness

- Simulations with $N = 10$ users with $N_F = N_H = 5$ in a heterogeneous network
- Equal arrival rate on all the uplinks and downlinks with total arrival rate $\rho \in (0, 1]$
- ***Fairness*** between full-duplex and half-duplex users (i.e., ratio between their queue lengths)

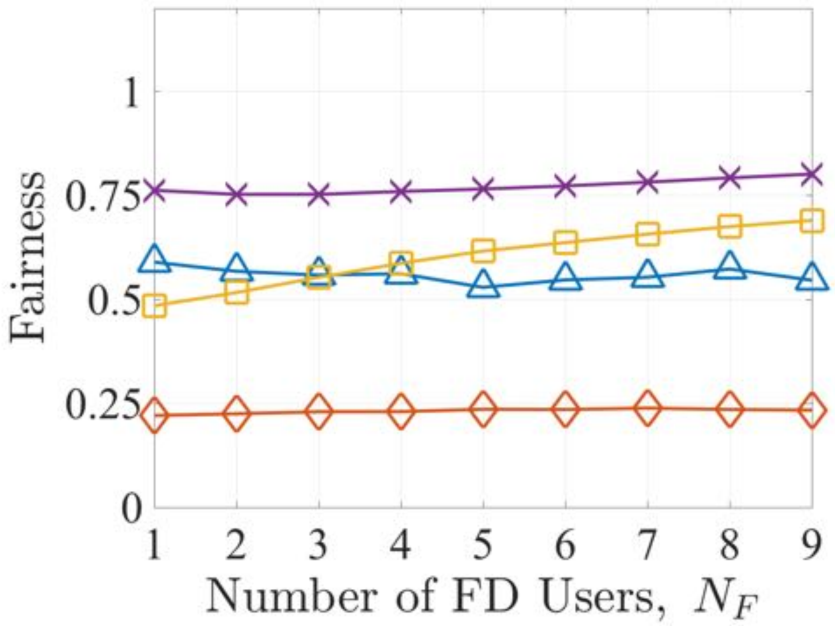


H-GMS-L and ***H-GMS-E*** improve fairness by selecting the initiator link differently

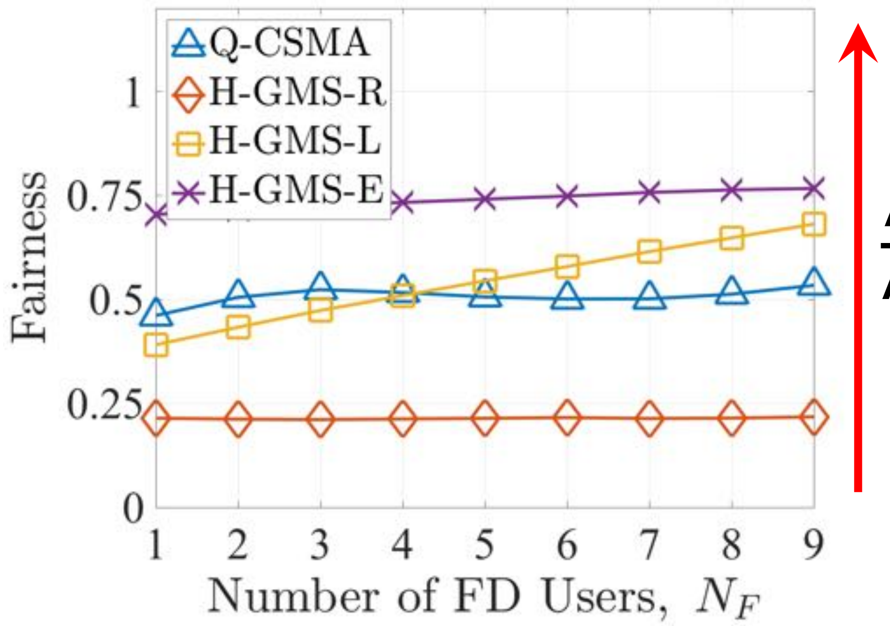
Performance Evaluation – Effect of N_F

- Simulations with $N = 10$ users with $N_F = N_H = 5$ in a heterogeneous network
- Equal arrival rate on all the uplinks and downlinks with total arrival rate $\rho \in (0, 1]$
- ***Fairness*** under different values of N_F

Medium traffic intensity, $\rho = 0.8$



High traffic intensity, $\rho = 0.95$



$$\frac{\text{Avg. } Q_{\text{full-duplex}}}{\text{Avg. } Q_{\text{half-duplex}}}$$

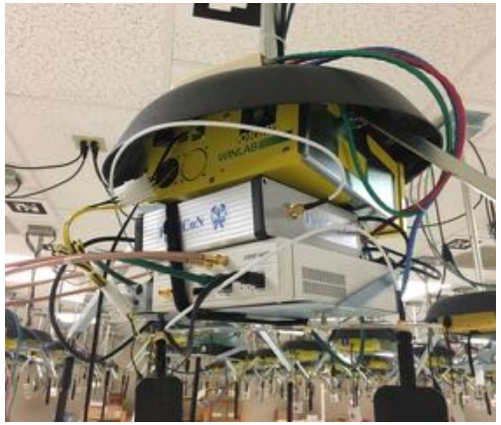
Summary

- Scheduling in heterogeneous half-duplex and full-duplex wireless networks
- All the users can gain (even for half-duplex users!) in terms of throughput when introducing full-duplex users into legacy half-duplex networks

- ***H-GMS*** – a hybrid scheduling algorithm combining centralized GMS and distributed Q-CSMA, and is proven to be throughput-optimal
- Performance evaluation of H-GMS

- Future directions:
  - Delay analysis of H-GMS
  - Experimental evaluation using existing/customized full-duplex testbeds

- Please come to our full-duplex demo ***tomorrow at 9:30am*** if you are interested!



Thank you!

tingjun@ee.columbia.edu

http://www.ee.columbia.edu/~tc2668

Tingjun Chen, Jelena Diakonikolas, Javad Ghaderi, and Gil Zussman,
“Hybrid Scheduling in Heterogeneous Half- and Full-Duplex Wireless Networks”.

Columbia University in the City of New York

wim.net Wireless & Mobile Networking Lab