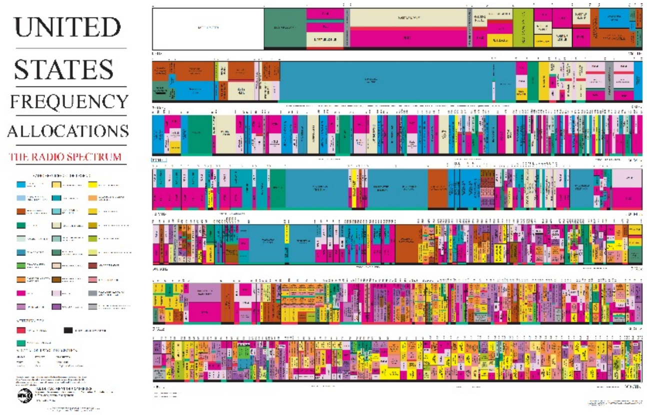
Opportunistic Channel Access Using Reinforcement Learning in Tiered CBRS Networks

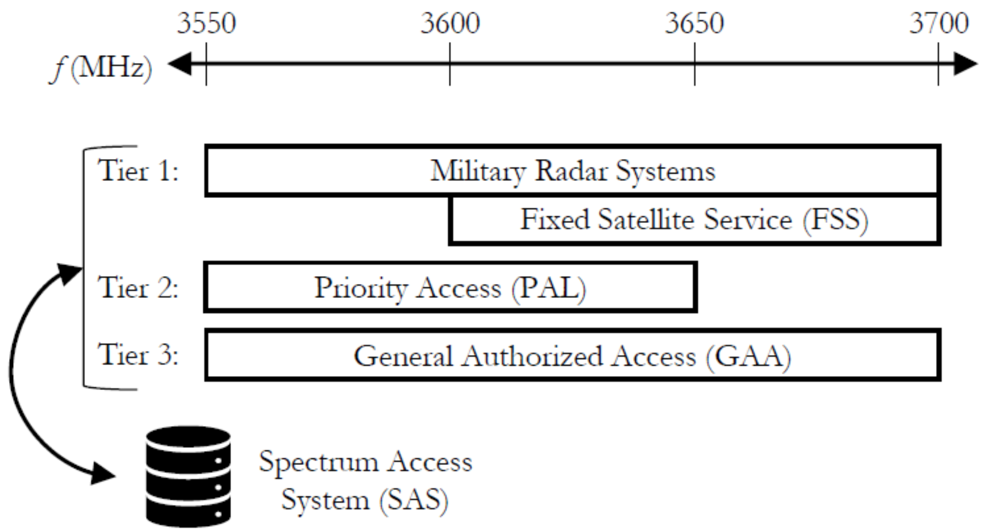
Tonnemacher, Tarver, Chandrasekhar, Chen, Huang, Loong Ng, Zhang, Cavallaro, Camp

Spectrum Availability Crisis



- Spectrum is a highly prized, but finite, resource
- Mobile application bandwidth demands are exponentially increasing over time
- There are two ways of providing for this trend:
  - Increase available spectrum for mobile broadband applications
  - Make better use of the spectrum that is being used
- In 2015, the FCC intends to do both with the Citizens Broadband Radio Service (CBRS)

An FCC Solution: CBRS



- CBRS band (3550-3700 MHz) has an LTE-based centralized spectrum sharing standard lead by CBRS Alliance
- Priority access based in a three tiered system: incumbents, PAL (primary node, PN), GAA (secondary node, SN)
- PAL licensing controlled by SAS, GAA operates in pseudo-unlicensed manner

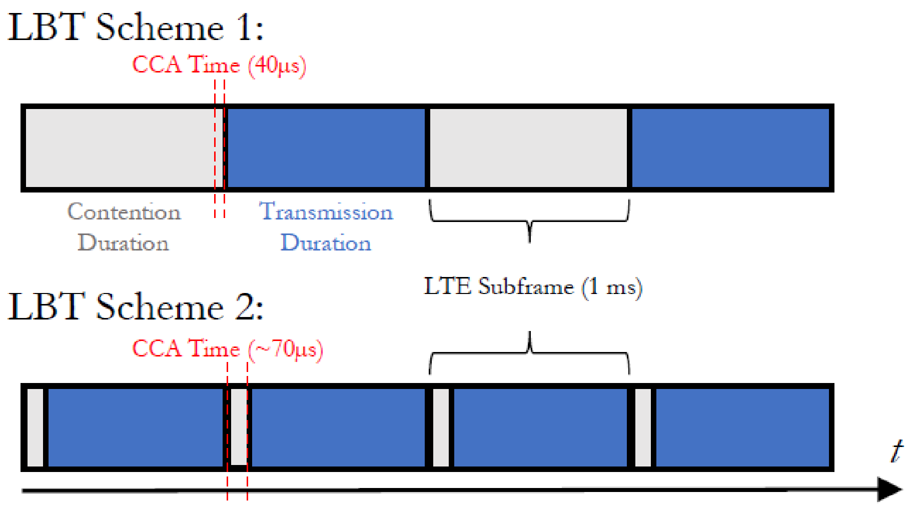
Improving Spectrum Sharing in CBRS Networks

- To improve spectral efficiency in CBRS…
  - We can use Listen-Before-Talk (LBT) schemes to enable opportunistic SN access of PN spectrum
  - LBT schemes rely on a spectrum sensing mechanic to decide if the medium is clear for transmission
  - We consider an energy detection threshold (EDT) to determine channel occupancy
    - If a device wants to transmit, they will sense the spectrum and transmit if it is clear
    - If not, they will defer according to some scheme for another transmit opportunity
- We first show performance when LBT is applied to LTE in a CBRS setup
- We then improve upon this performance using a machine learning framework

Agenda

1. Introduce LBT schemes for LTE
2. LBT performance in different sharing scenarios
3. Problems with LBT in tiered CBRS
4. Introduction to reinforcement learning
5. Our machine learning framework
6. Performance improvements

LTE LBT Schemes Considered

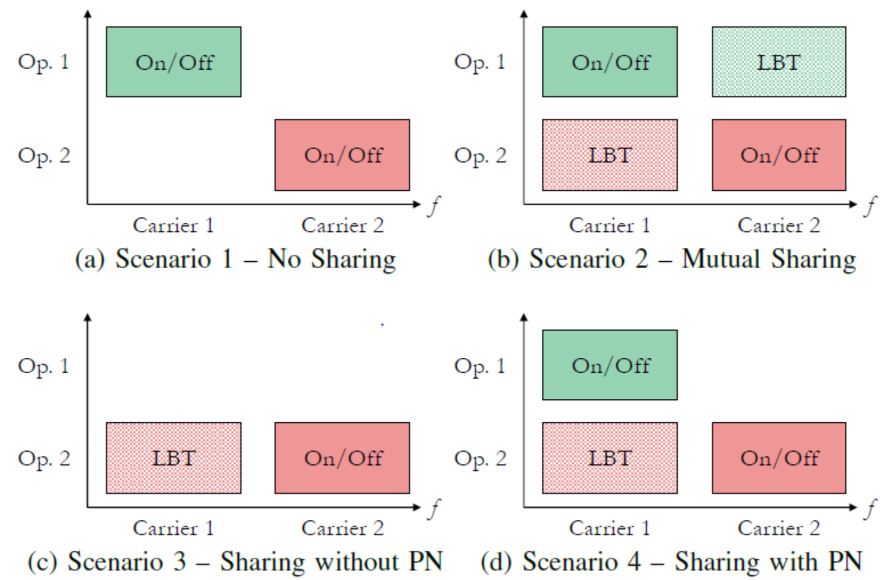


- We compare two spectrum sensing schemes
  - Scheme 1 uses an entire subframe as a contention window, but clear channel assessment relies on last 40 µs
  - Scheme 2 uses first symbol of subframe to sense, with data in the remaining 13 symbols
- Both schemes assume subframe synchronization, which we are able to do in our hardware PoC.
- Similar schemes have been considered in LTE/WiFi coexistence studies.

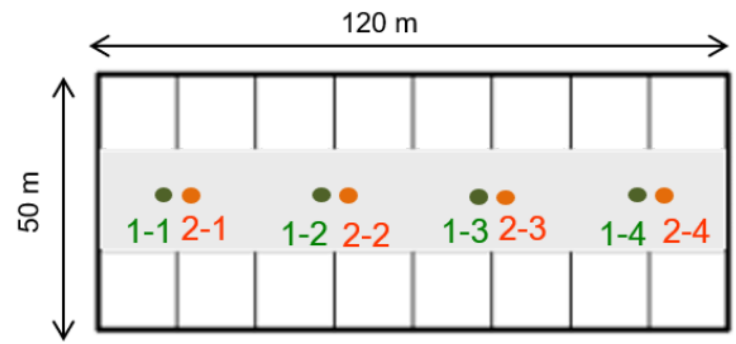
LBT Simulation Scenarios

- We consider four spectrum sharing scenarios to compare the LBT schemes
- Sims use 3GPP indoor scenario for LAA coexistence evaluations
  - Two operators, four small cells in single floor building
  - 18 dBm TX power, -72 dBm EDT
  - 10 randomly distributed UEs per operator
  - 20 MHz system bandwidth
  - 10 drops simulated, 20k subframes per sim
- Performance evaluated in terms of user perceived throughput (UPT) given by:

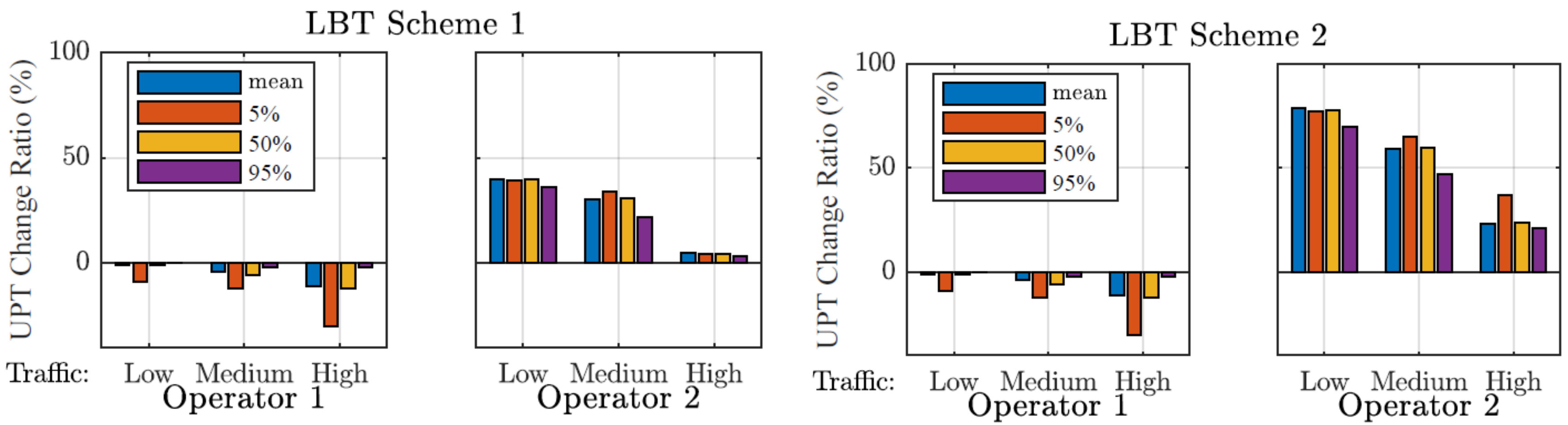
$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{P_{total}}\left[\sum_{j=1}^{P_{served}}\frac{M\cdot r_{ij}}{t_{ij}}+\frac{b_i}{t_{serving,i}}\right]$$



(a) Scenario 1 – No Sharing
(b) Scenario 2 – Mutual Sharing
(c) Scenario 3 – Sharing without PN
(d) Scenario 4 – Sharing with PN

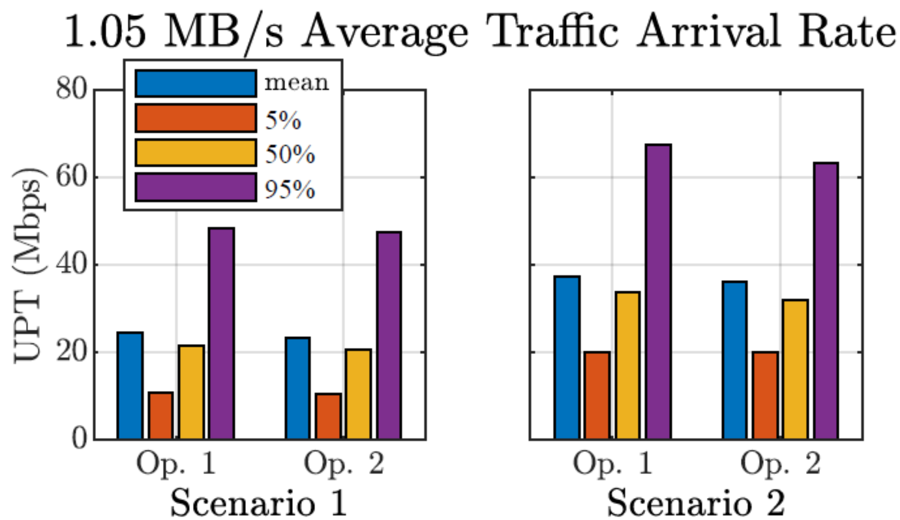


LBT Scheme Comparison



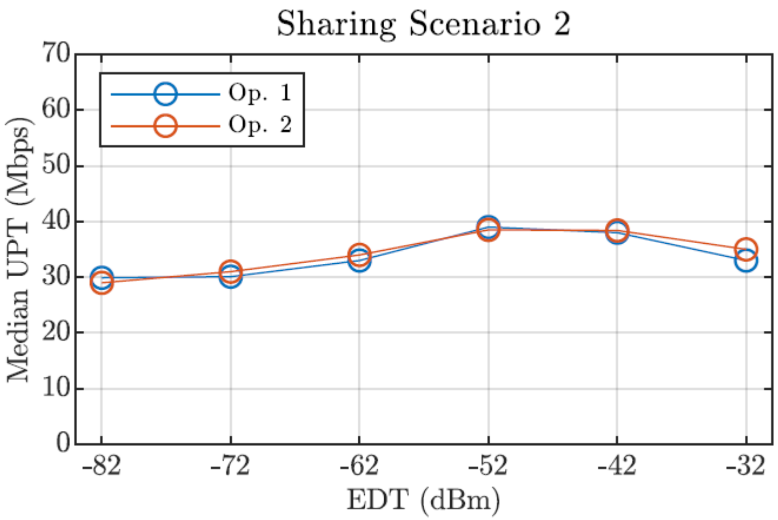
- Consider LBT sharing in scenario 4 (OP 2 sharing on OP 1 carrier)
- As expected, the UPT gains for operator 2 (SN) are significantly better when using scheme 2
  - Duty cycle of opportunistic transmissions is higher
- Both schemes result in similar UPT drop for operator 1 (PN)
- *For remainder of study, we implement scheme 2 because of its performance advantage*

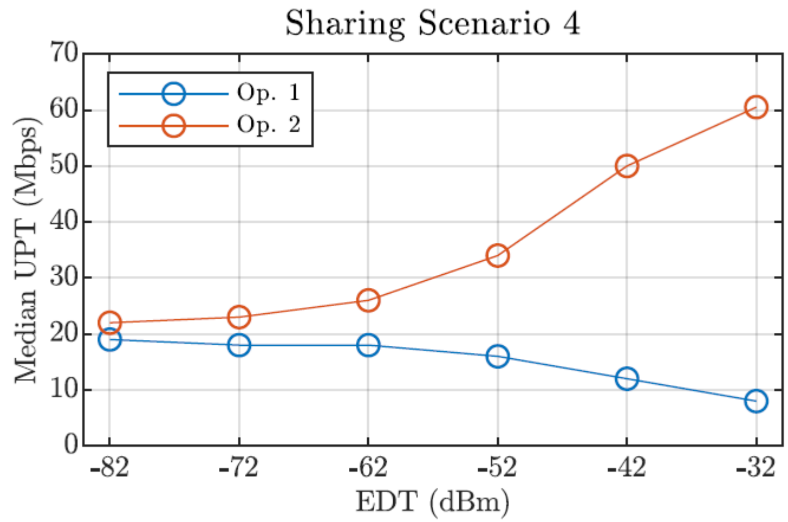
LBT Overall UPT Performance Gain



- We only consider symmetric sharing scenarios, as we are looking for boundary conditions
- Median UPT increase of approximately 55% for both operators
- Joint spectrum sharing sum rate out performs isolated channel operation, showing to better spectrum usage

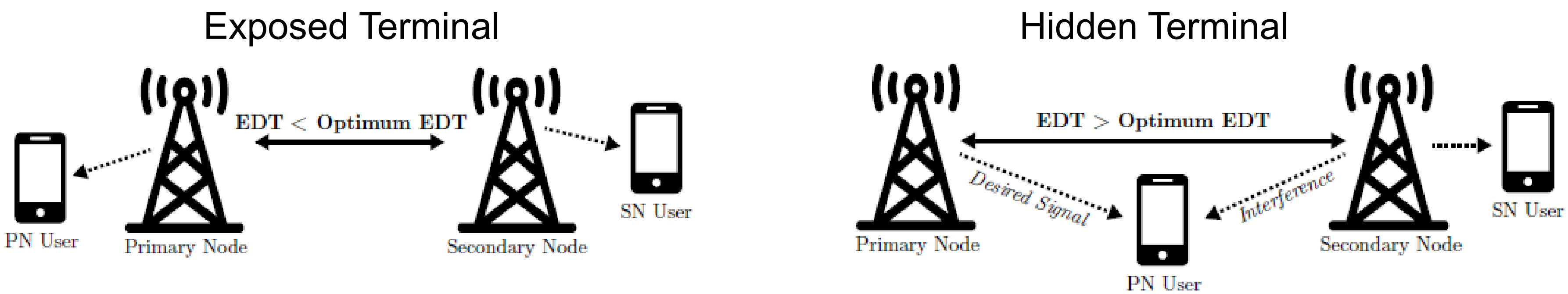
Energy Detection Threshold Effect





- In a mutual sharing scenario (2),
  - "best" EDT can be found for a given topology satisfying both operators
- In a one-sided sharing scenario (4),
  - Higher EDT values result in more sharing opportunities, at the expense of increased interference for Op. 1
- *EDT selection can play a large part in PN and SN performance*

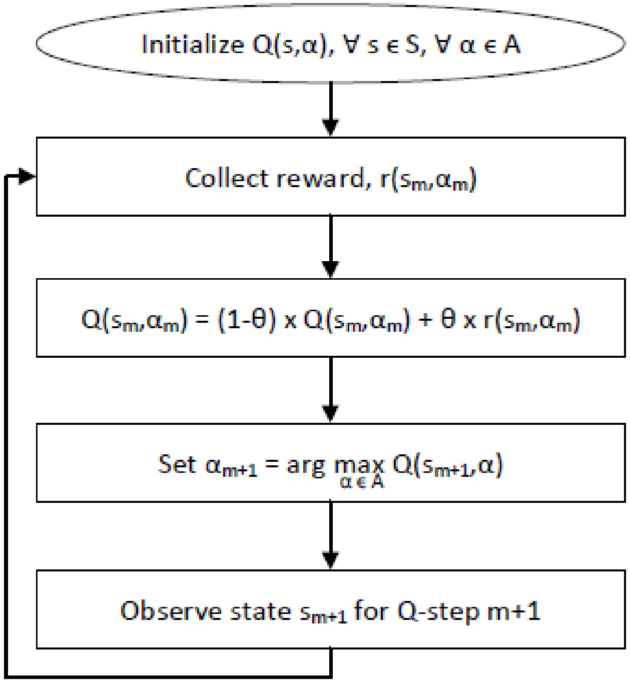
Problems with Licensed LBT



- Statically-defined LBT policies can be vulnerable to problematic topologies
  - Exposed terminal (left), Hidden terminal (right)
- LBT in LTE cannot easily deal with these in the same way as other LBT schemes, such as 802.11 CSMA/CA
- However, proper EDT selection can mitigate consequences of exposed/hidden terminals
- *We propose using a Q-learning based EDT adjustment scheme to reduce the negative consequences of spectrum sharing on the PN*

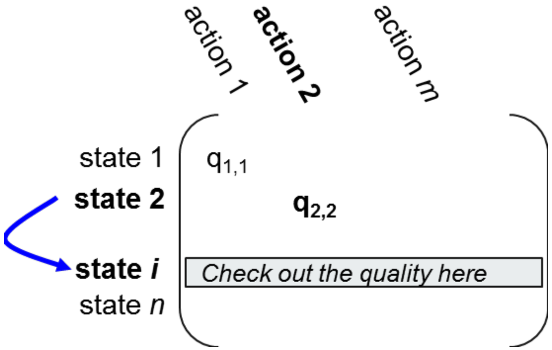
Reinforcement Learning Background

- Generally, reinforcement learning can be described by the interaction between agent and environment at regular epochs, indexed by *m*
  - The agent:
    - Executes action α_m
    - Receives observation α_m of state s_m
    - Receives reward r_m
  - The environment:
    - Receives action α_m
    - Emits observation o_{m+1} of state s_{m+1}
    - Emits scalar reward r_{m+1}



Q-Learning Structure - States and Actions

- Defining a state set (and action set) is always a tradeoff between outcome resolution and convergence time
  - If there are too many states/actions, convergence in the table of state-action pairs (Q-Table) to a clear best action can take too long
  - Too little states, and nuance in the environmental observation can be lost



- In our reinforcement learning design, the observed state is defined by the average queue size in the primary node:

State	Average Primary Node Queue Size	Comment
1	$0 \leq L_m < \gamma_1$	Primary node traffic load is light
2	$L_m \geq \gamma_1$	Primary node traffic load is heavy

- Actions are a discreet set of specific EDT values
- A tunable threshold, γ_1, is set to differentiate the states

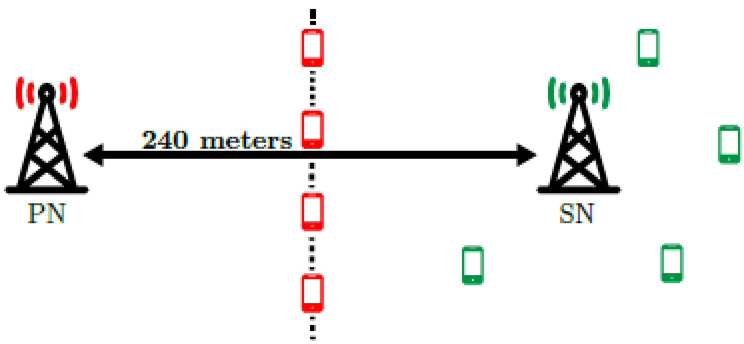
Q-Learning Structure - Reward

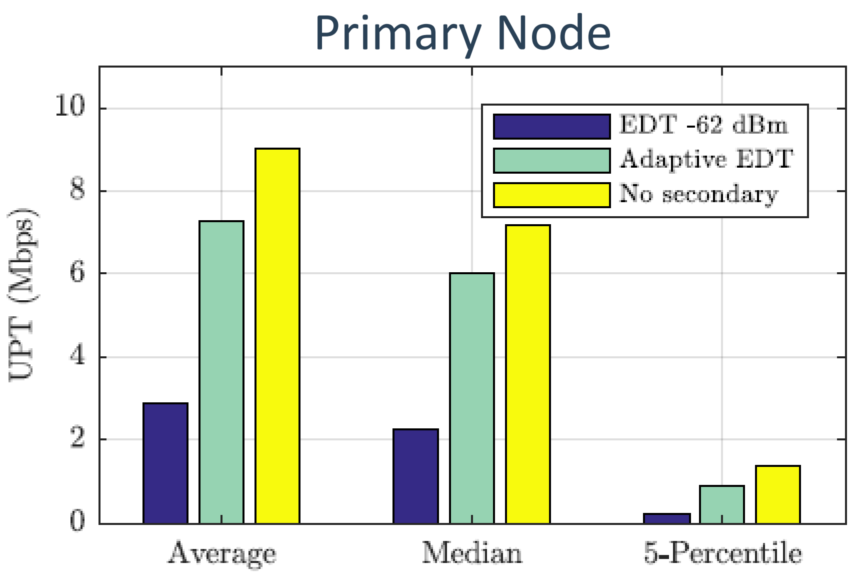
(s_m, s_{m+1})	Reward $r(s_m, \alpha_m)$
$(1,1)$	• γ_4, if $B_m \leq \gamma_2$ and $\alpha_m \geq \gamma_3$ • $-\gamma_4$, if $B_m \leq \gamma_2$ and $\alpha_m < \gamma_3$ • Z_m, otherwise
$(1,2)$	• γ_4, if $B_m \leq \gamma_2$ and $\alpha_m \geq \gamma_3$ • $-\gamma_4$, if $B_m \leq \gamma_2$ and $\alpha_m < \gamma_3$ • $-\gamma_4$, if $B_m > \gamma_2$
$(2,1)$	• 0, if $B_m \leq \gamma_2$ • γ_4, otherwise
$(2,2)$	• 0, if $B_m \leq \gamma_2$ • Z_m, otherwise

- Rewards given to state, action pair depending on state transition, observation, and action
- Here:
  - B_m : average buffer occupancy over previous epoch
  - Z_m : smaller, soft reward when outcomes are between actionable thresholds
  - γ_2 : buffer occupancy threshold; γ_3 : low/high EDT threshold; γ_4 : base reward
- We use discount factor, θ, to control the importance of the reward via:

$$Q(s_m, \alpha_m) = \theta Q(s_m, \alpha_m) + (1-\theta) r(s_m, \alpha_m)$$

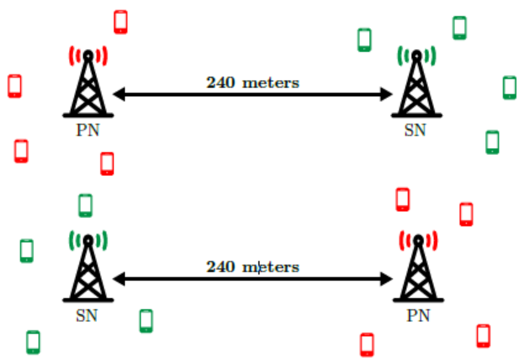
Problematic Topology Interference Mitigation

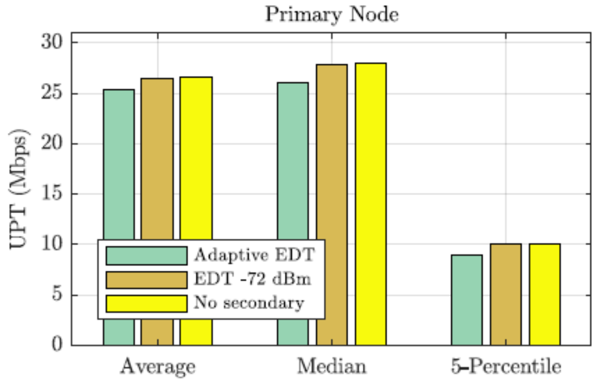




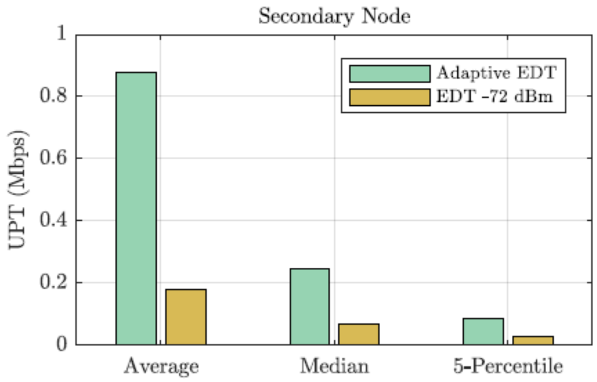
- In hidden node topology, a high EDT can cause unwanted interference on PN, severely hurting PN UPT
- Using Q-learning, EDT drifts downwards, yielding PN UPT gains compared with a fixed high EDT
- In this scenario, no SN (interference free PN) is shown as the PN UPT upper bound

Adapting to Primary Node Load





(a) PN UPT for the case where the SN adapts to PN load.



(b) SN UPT for the case where the SN adapts to PN load.

- In this topology, setting SN EDT low can unnecessarily block non-interfering opportunistic access
- Adapting EDT allow SN to harvest more bandwidth from PN carrier with little loss to PN UPT

Conclusion

- In this work, we:
  - Examined the challenges of using LBT for PAL/GAA spectrum sharing in CBRS networks
  - Evaluated two LBT schemes, showing greatly improved SN UPT with minor PN UPT reduction
  - Reduced negative consequences of PN spectrum sharing via novel Q-learning algorithm that adjusts SN opportunistic access by learning an optimal EDT for carrier sensing
  - Showed that using average and differential PN buffer occupancy as RL environmental observation can improve throughput by up to 350% with only marginal (4%) PN UPT losses

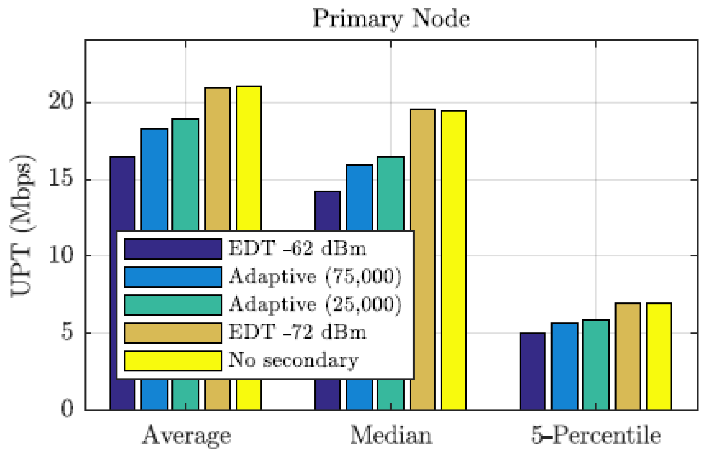
Thank you!

Matthew Tonnemacher

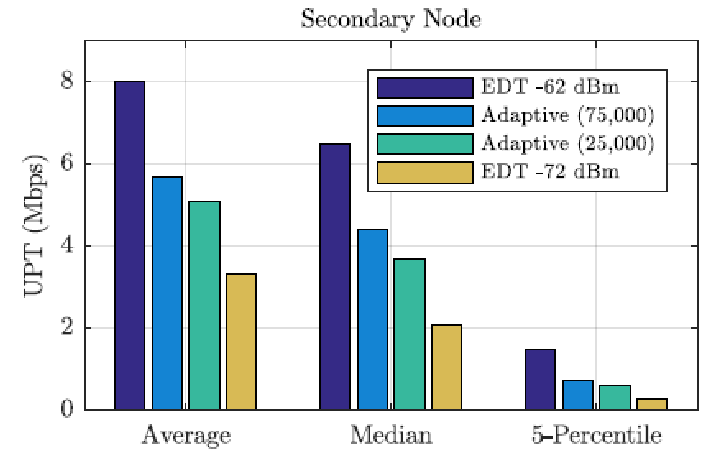
matthew2.t@samsung.com

Q-Learning Parameter Effects

- Setting state threshold plays a big role in learning performance
  - Different traffic load/state mappings result in slightly different UPT
  - Higher thresholds on high load favor SN opportunistic access
  - Thresholds can be adjusted to specific deployment depending on overall network needs:
    - Is it more important to protect the PN traffic?
    - Is it more important to maintain high spectral efficiency?

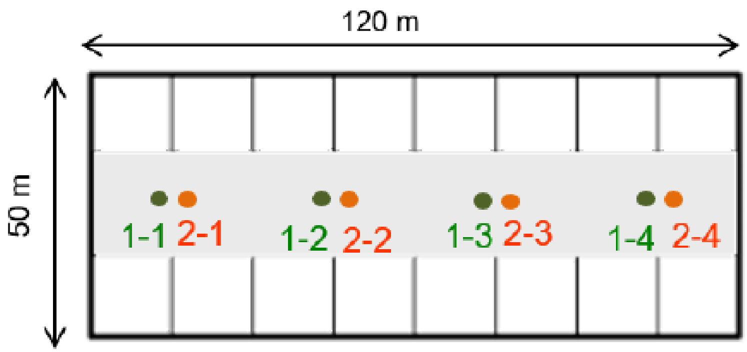


(a) PN UPT with differing γ_1 values.

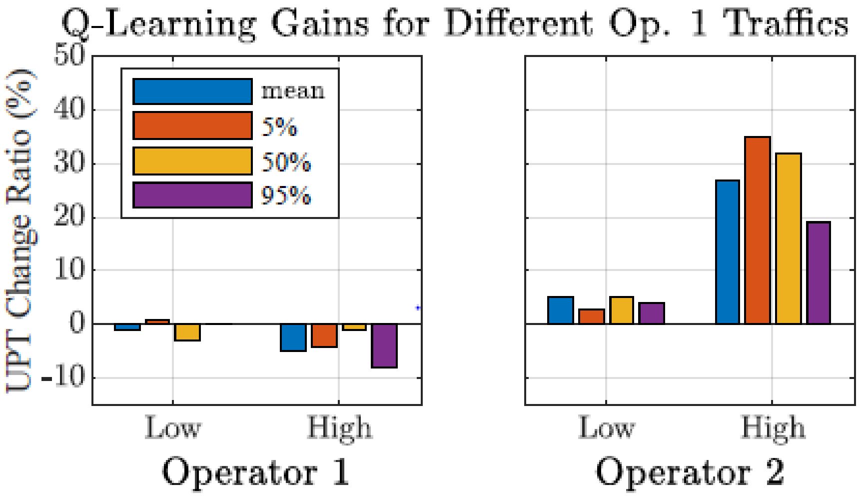


(b) SN UPT with differing γ_1 values.

Multi-Node Scenario Traffic Adaptation



3GPP indoor scenario for LAA coexistence evaluations



- We examine four PN-SN node pairs with low (0.125 MB/s average traffic arrival rate) and high (1.05 MB/s average traffic arrival rate), comparing fixed vs adaptive EDT
- In low-traffic cases, fixed EDT performs similarly to adaptive, as there are rarely collisions in access attempts
- Higher traffic results in adaptive EDT providing substantial gains over the fixed EDT case