ML-Based C-DRX Configuration Optimization

Philipp Bruhn & Germán Bassi

Content

- Motivation, background, and problem
- Our machine-learning-based solution
- Results, insights, and lessons learned
- Conclusions and future work

Motivation, Background, and Methodology

Motivation

- Discontinuous Reception (DRX) allows UE to check for incoming downlink traffic intermittently to reduce UE energy consumption
- DRX creates a trade-off between UE energy saving and delay or throughput
- Number of possible DRX configurations very large [1],[2] → difficult to choose

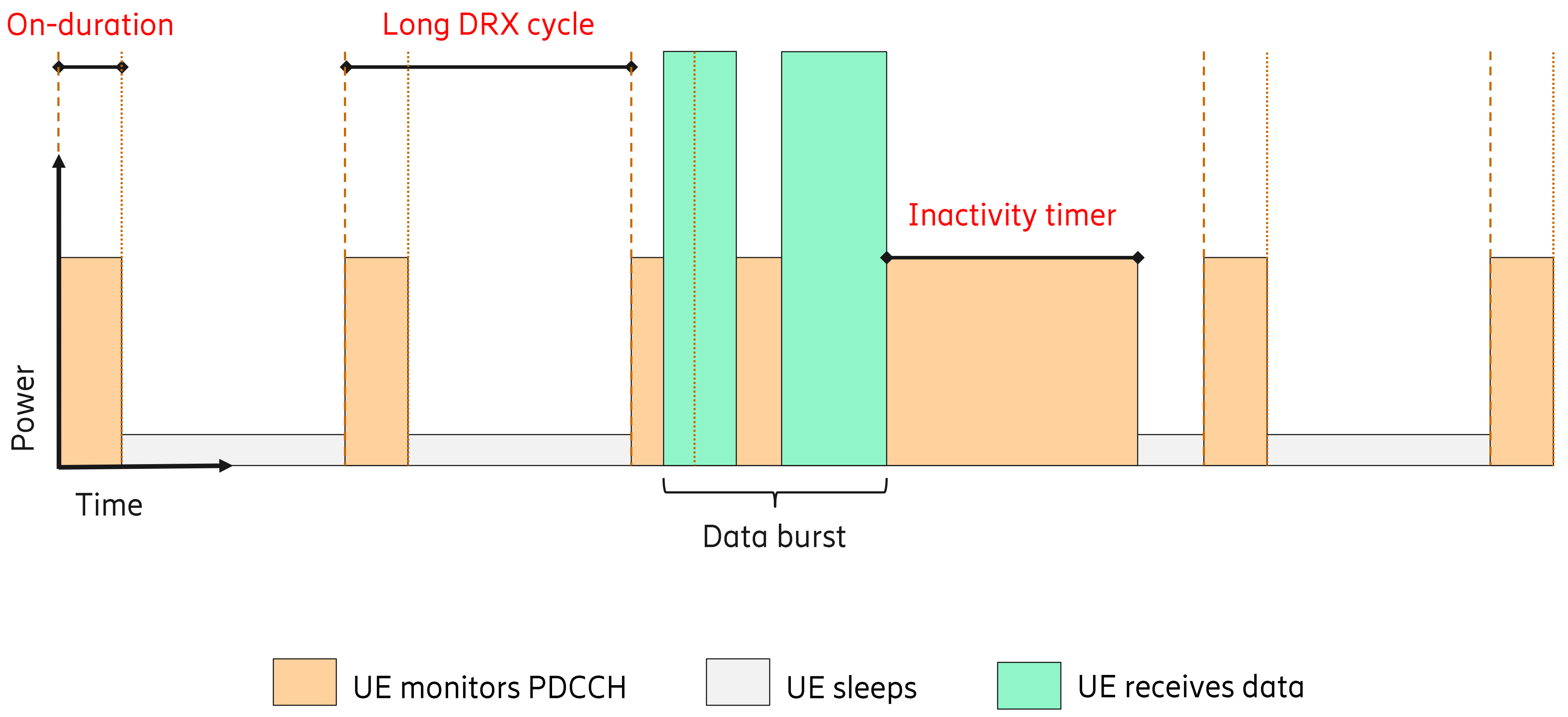
Goal of this work:

- Optimize DRX configuration per UE according to performance goals or intents using Machine Learning (ML) techniques
- Demonstrate value of energy-consumption-related feedback from UE

[1] 3GPP TS 38.321 "NR; Medium Access Control (MAC) Protocol Specification" [2] 3GPP TS 38.331 "NR; Radio Resource Control (RRC) Protocol Specification"

Connected Mode DRX

Long DRX cycle only



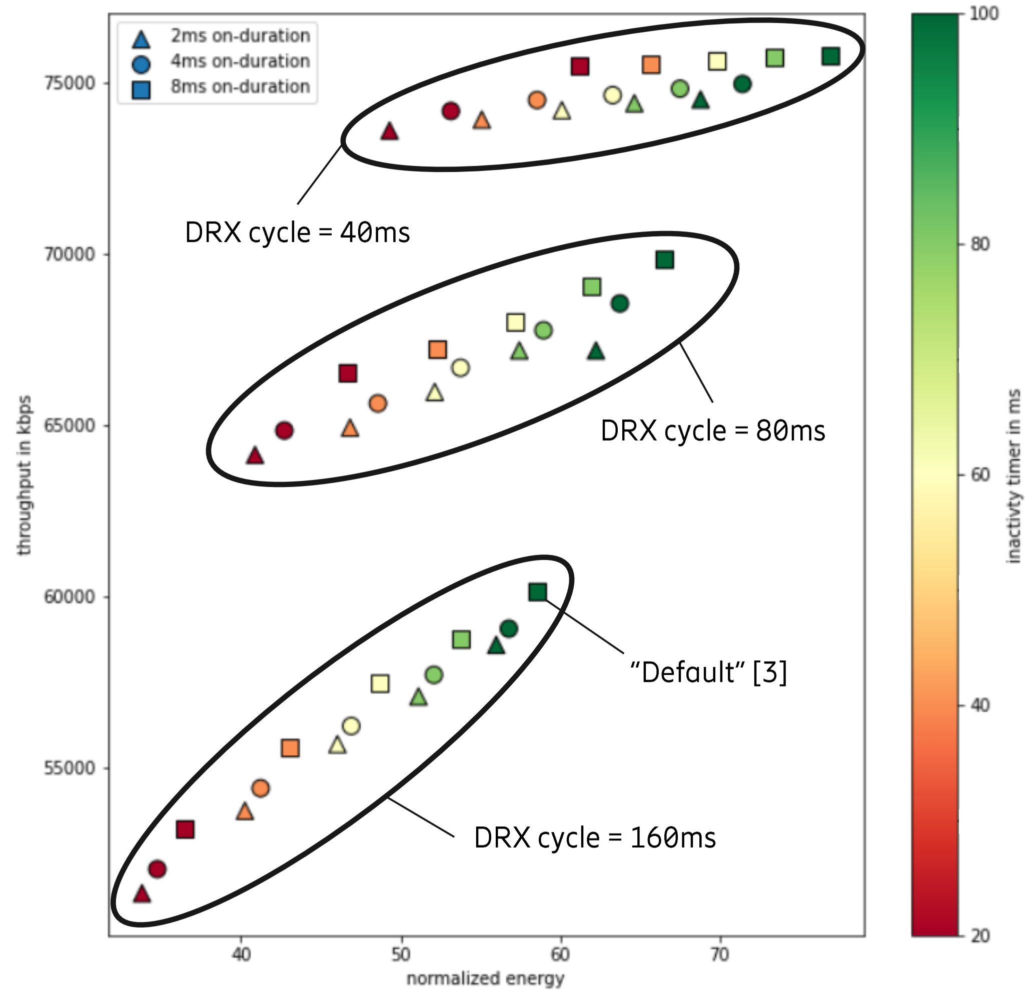
Related C-DRX Study

Default DRX setting according to 3GPP:

Long DRX cycle	160 ms
On-duration	8 ms
Inactivity timer	100 ms

FTP traffic model 3

- Packet size = 500 kB
- Mean inter-arrival time = 200 ms



[3] 3GPP TR 38.840 "Study on User Equipment (UE) Power Saving in NR"

Contextual Bandit

<u>Approach:</u>

Use contextual bandit to set optimal DRX configuration depending on UE energy consumption (and other) feedback

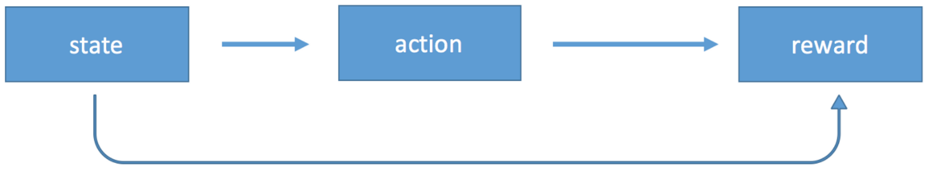
→ Realized using Vowpal Wabbit

Bandit behavior mainly affected by:

- Policy evaluation approach
  → Direct method
- Exploration strategy
  → Epsilon-greedy

VOWPAL WABBIT

"Contextual bandit is a machine learning framework designed to tackle [...] complex situations. [...] A learning algorithm can test out different actions and automatically learn which one has the most rewarding outcome for a given situation." [4]



Contextual bandit problem, where state and action effect reward [5].

[4] How to Build Better Contextual Bandits Machine Learning Models | Google Cloud Blog [5] Contextual Bandits and Reinforcement Learning | Towards Data Science

Exploration vs. Exploitation

*"A learning algorithm can **test out different actions** and automatically learn which one has the most rewarding outcome for a given situation."*

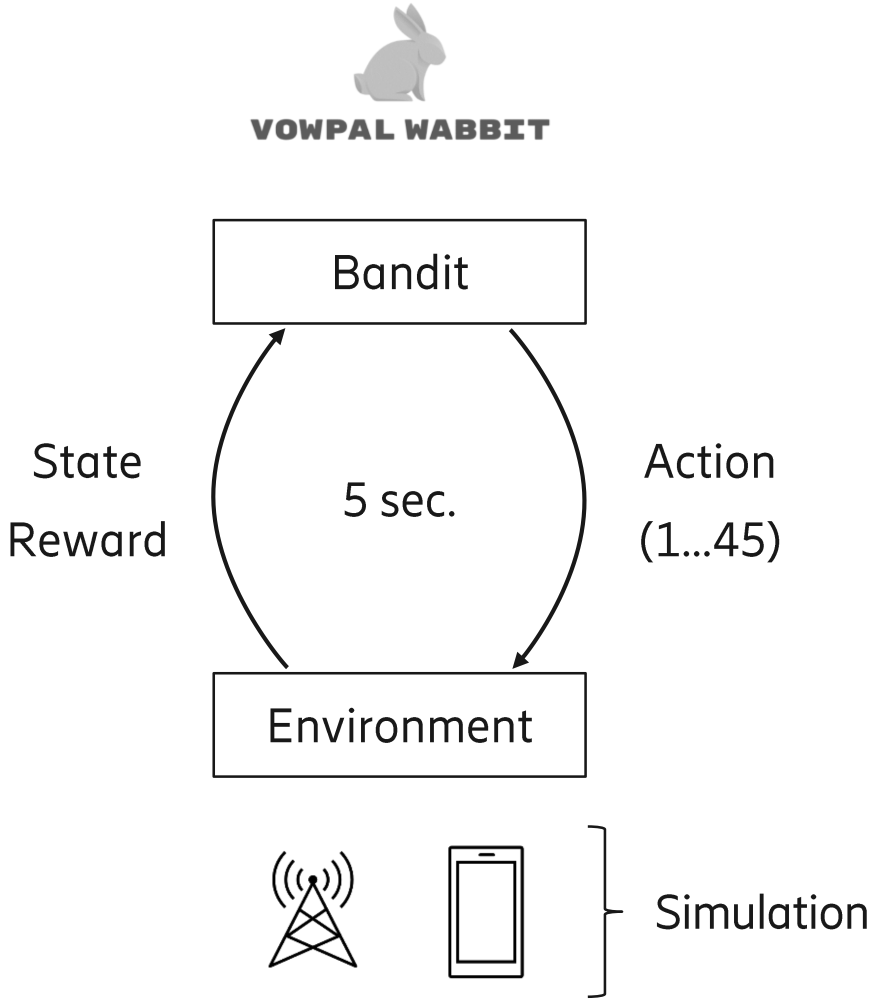
ϵ-greedy

- Parameter ϵ controls trade-off between exploration vs. exploitation (for $0 < \epsilon < 1$)
- Exploitation with probability $1 - \epsilon$: Bandit chooses action based on (assumed) best reward
- Exploration with probability ϵ: Bandit chooses action uniformly at random
- ϵ can be fixed, adjusted over time ("ϵ-decay"), or adapted in other ways, e.g., based on heuristics

Our choice: Linear "ϵ-decay" from 100% to 5% during first 1000 learning steps

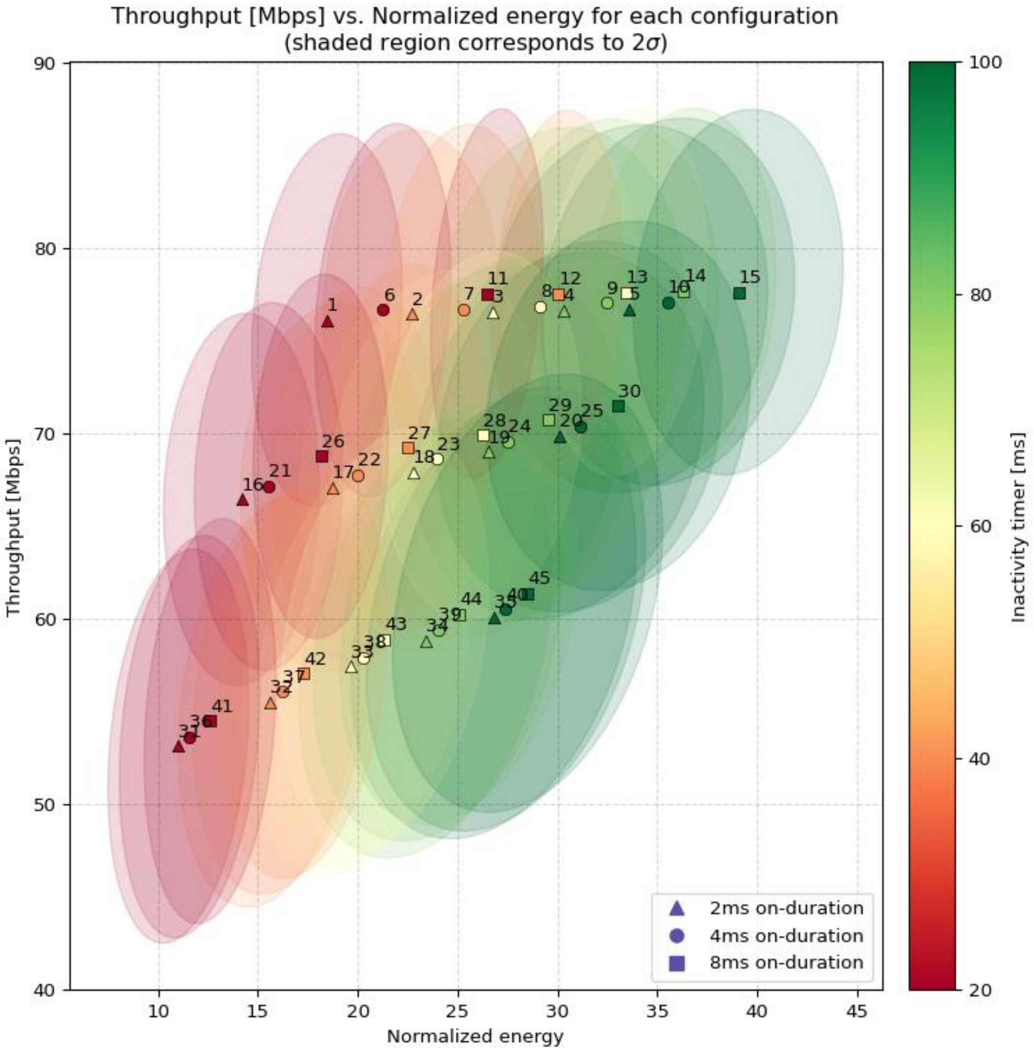
Observation-Action Cycle

- State and reward created based on 5 sec. observation (5 sec. averaging period)
- Bandit chooses among 45 actions corresponding to 45 DRX configurations (labeled as 1…45)
- Chosen action is translated into DRX configuration upon actuation in simulation



Results and Insights

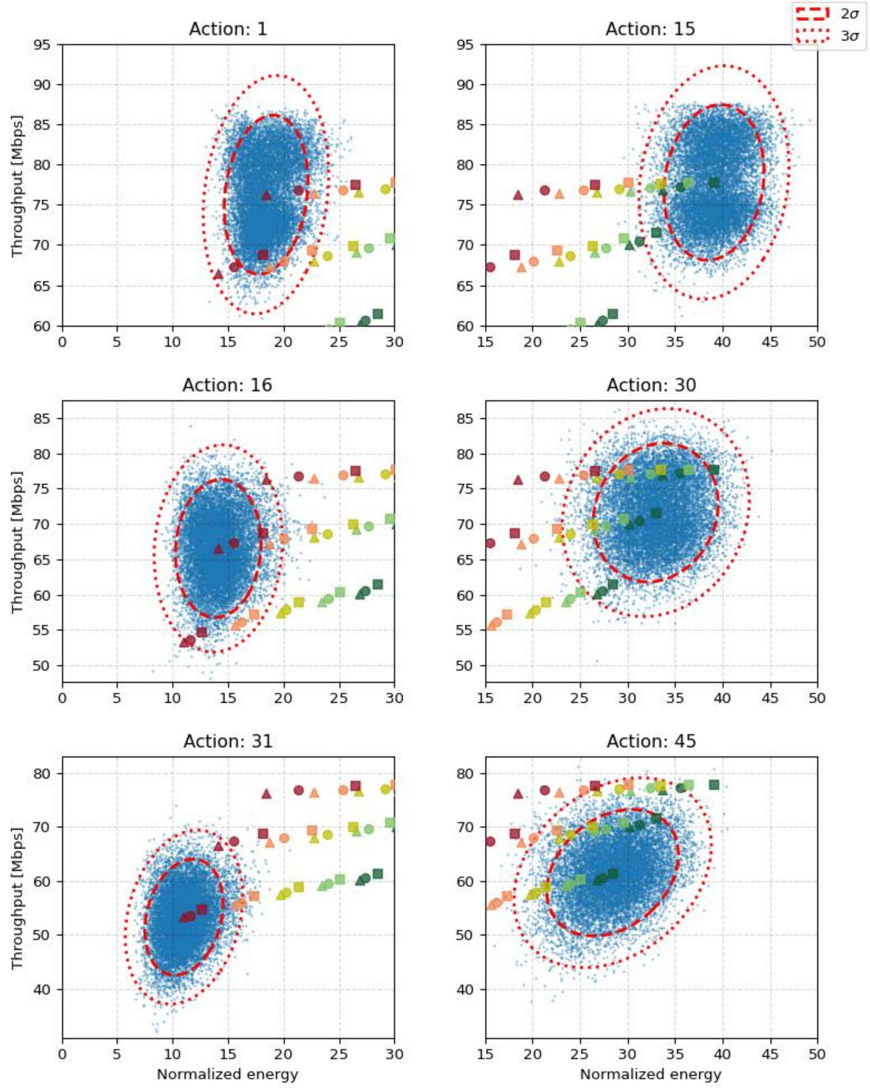
Statistics of the DRX configurations



By choosing a certain DRX configuration (a.k.a. action 1-45) the UE experiences a variable throughput/energy consumption

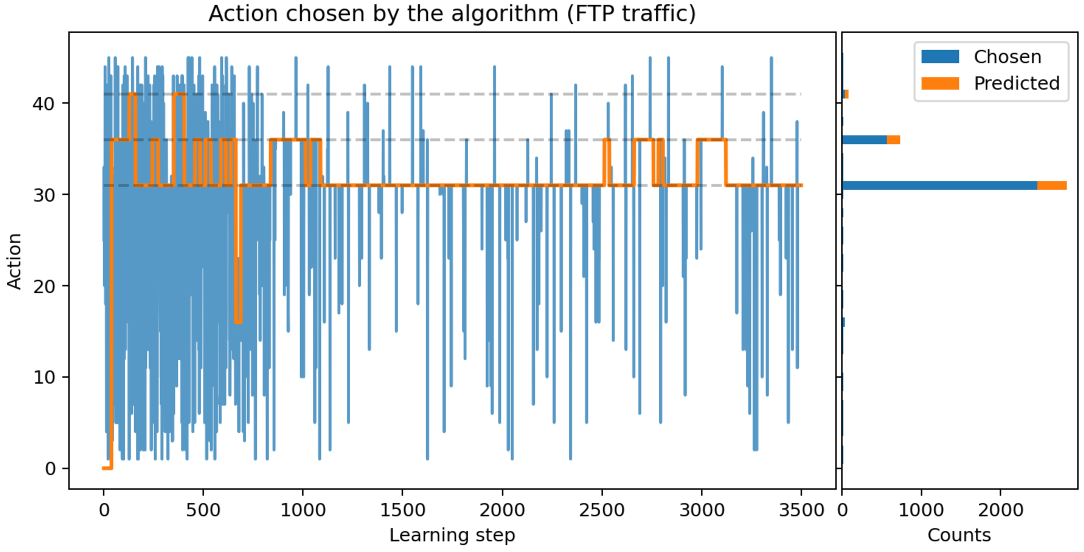
The fine granularity of the different DRX parameters results in a large overlap between configurations

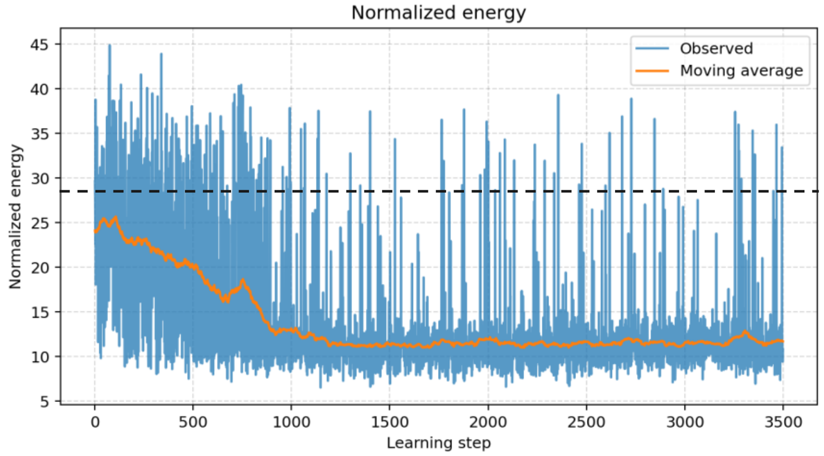
More in detail



Experiment #1

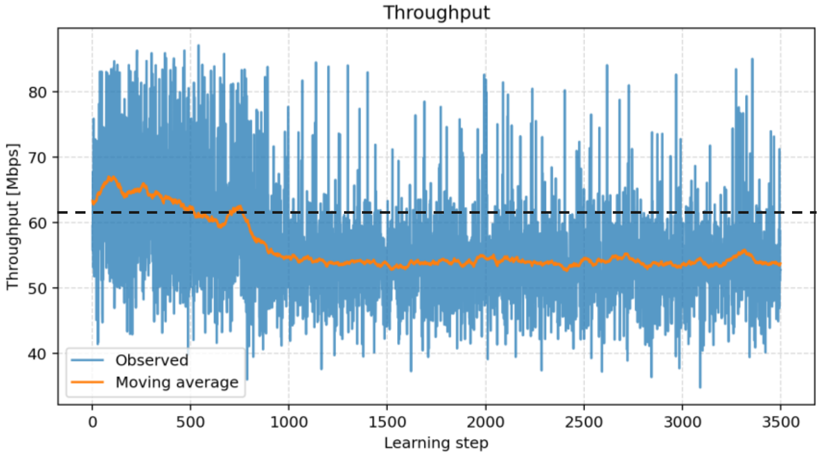
- Traffic: FTP-3 model
- Intent: Energy minimization
- Reward ~ $Energy_{Monitor}^{-1} \in (0,1)$





28.4 Units

Mean power of default DRX configuration (acc. 3GPP)



61.5 Mbps

Mean throughput of default DRX configuration (acc. 3GPP)

→ Compared to <u>default</u> DRX configuration, approximately 52% "useless" energy saving, but only 9% throughput loss.

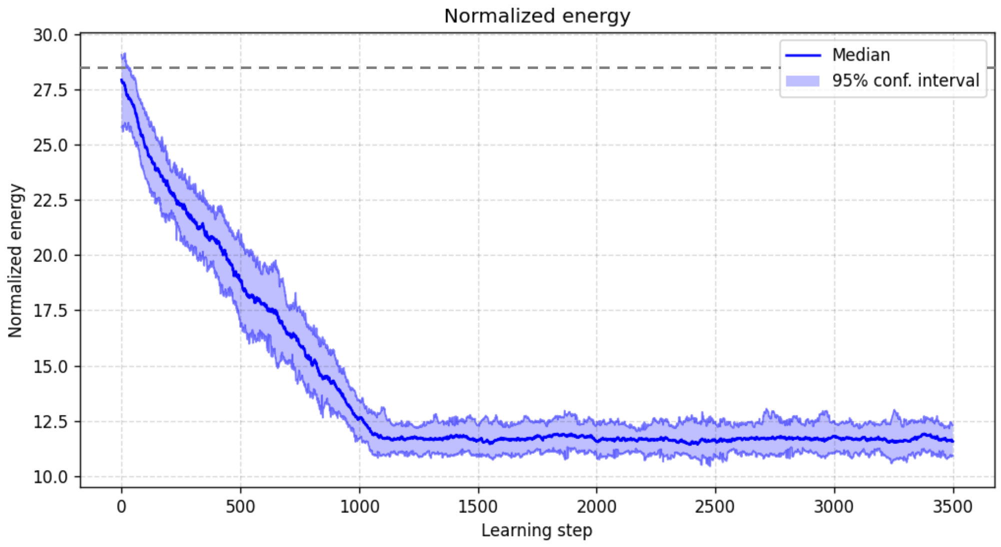
$Energy_{Monitor}$: Energy consumed for monitoring PDCCH (excluding rx on PDSCH and tx on PUSCH)

Experiment #1

- Traffic: FTP-3 model
- Intent: Energy minimization
- Reward ~ $Energy_{Monitor}^{-1} \in (0, 1)$

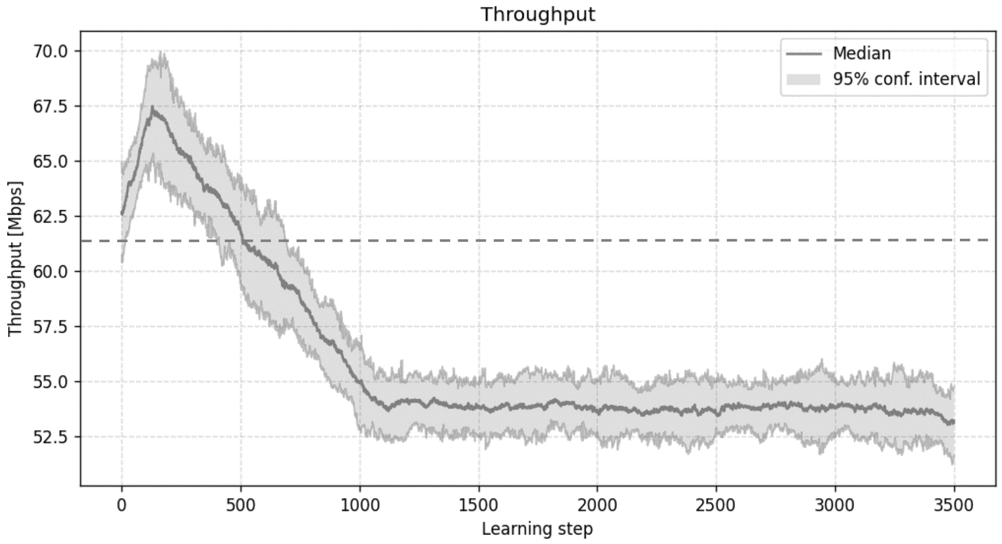
Simplified representation: Median and 95% confidence interval for 40 learning passes (with averaging window of 101 steps per simulation)

→Reflects level of stability of learning



28.4 Units

default DRX configuration



61.5 Mbps

default DRX configuration

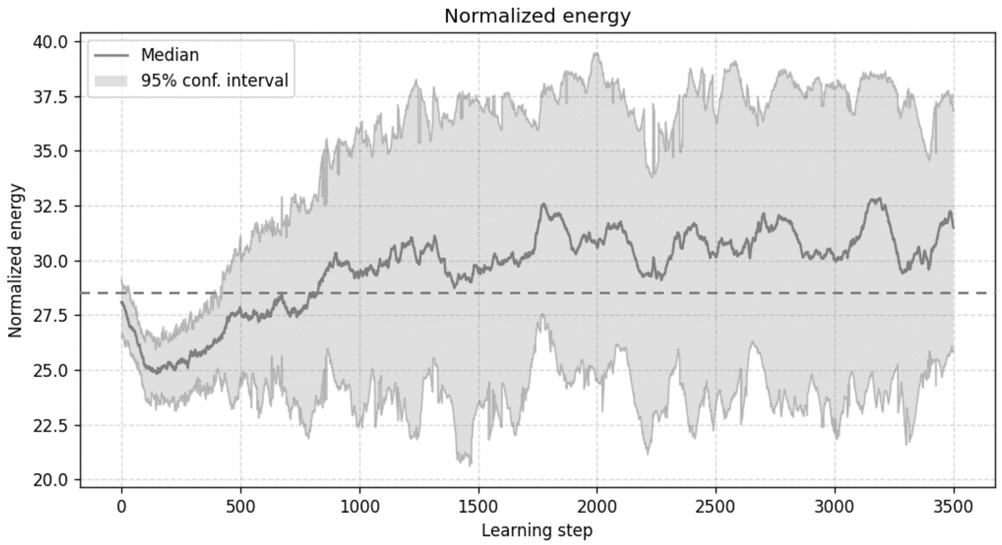
$Energy_{Monitor}$: Energy consumed for monitoring PDCCH (excluding rx on PDSCH and tx on PUSCH)

Experiment #2a

- Intent: Throughput maximization
- Reward ~ $\left(\frac{Throughput}{Max.Throughput}\right)^6 \in (0,1)$

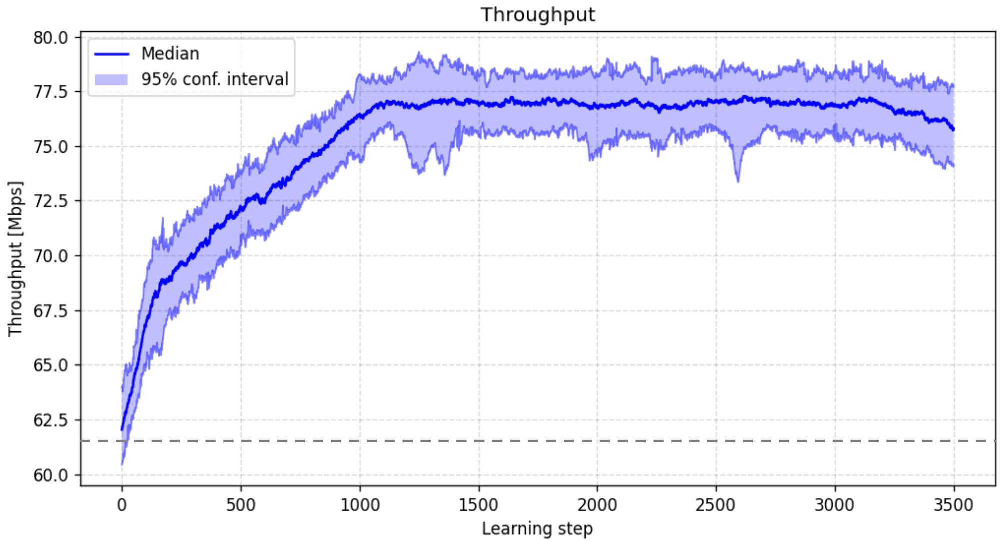
  → Non-linearity improves learning
- Observation: SINR

- Note: 37-71% correlation between throughput and SINR (mainly depending on the DRX cycle)



28.4 Units

default DRX configuration



61.5 Mbps

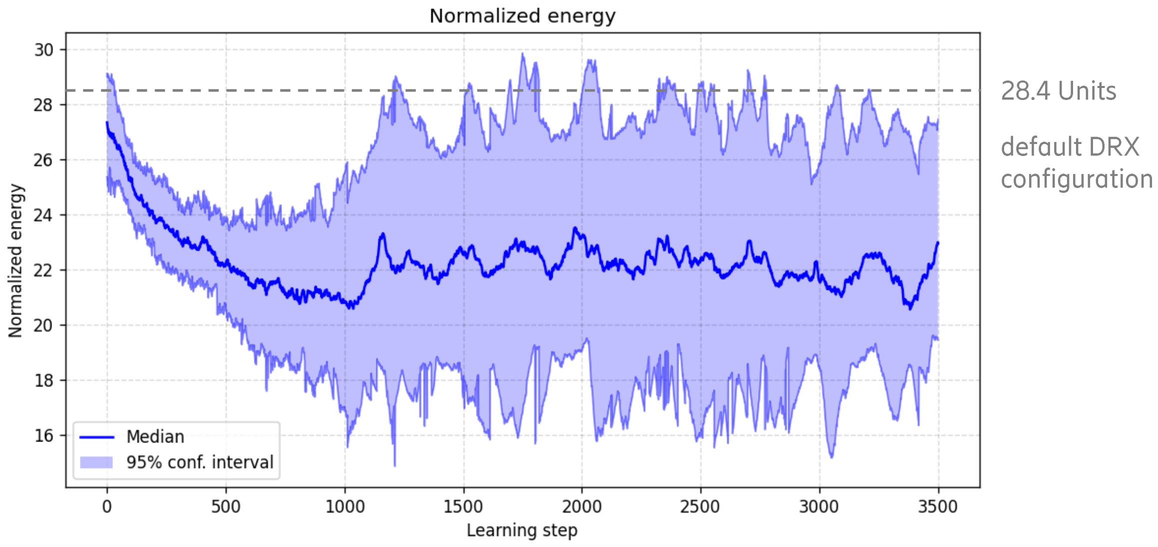
default DRX configuration

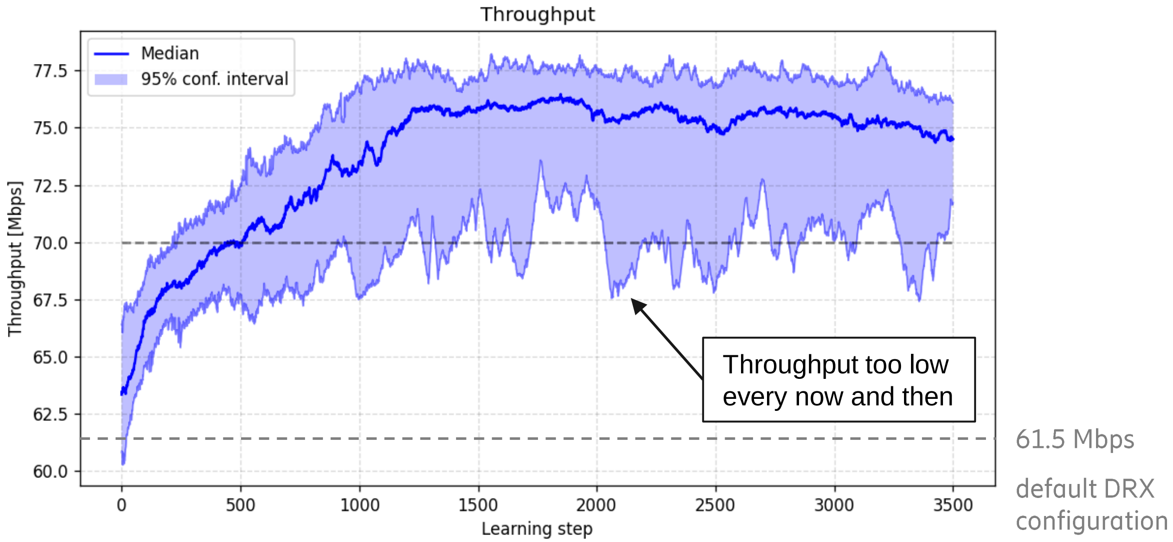
Experiment #3

- Intent: Performance-bounded energy minimization
- Constraint: Throughput min. 70 Mbps
- Reward $\begin{cases} \sim Energy_{Monitor}^{-1} & if\ TP \geq 70\ Mbps \\ -1 & if\ TP < 70\ Mbps \end{cases}$

Compared to Exp. 2a:

→ Throughput loss <u>2.4%</u>; Energy saving <u>26.2%</u>

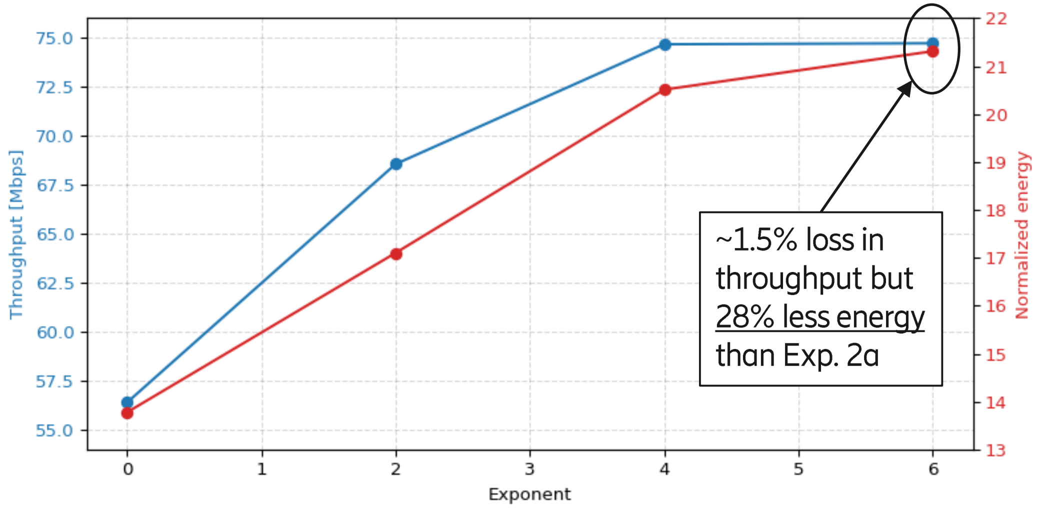


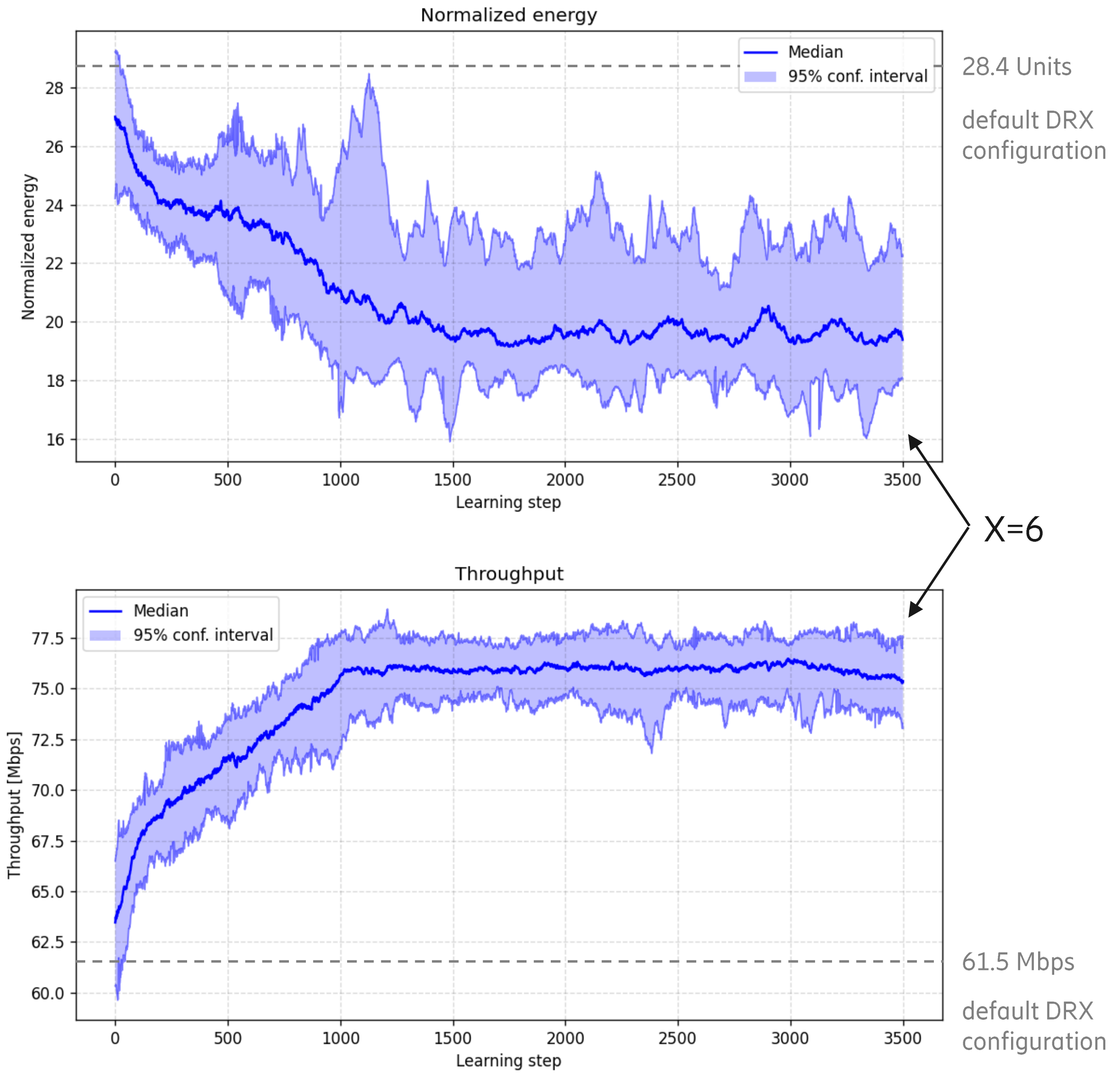


$Energy_{Monitor}$: Energy consumed for monitoring PDCCH (excluding rx on PDSCH and tx on PUSCH)

Experiment #4

- Intent: Adjustable joint optimization
- Reward ~ $Energy_{Monitor}^{-1} * \left(\frac{Throughput}{Max.Throughput}\right)^{X}$
- Varying the exponent, we give more or less relevance to maximizing throughput





$Energy_{Monitor}$: Energy consumed for monitoring PDCCH (excluding rx on PDSCH and tx on PUSCH)

Closing Remarks

Takeaways

Contextual bandit fast and simple approach to select good DRX configuration according to UE feedback

Randomness in rewards increases uncertainty and impairs convergence/stability

Hard performance bounds cause rewards that make learning more challenging

Multiple types of users (e.g., multiple types of traffic and intents) can be handled simultaneously

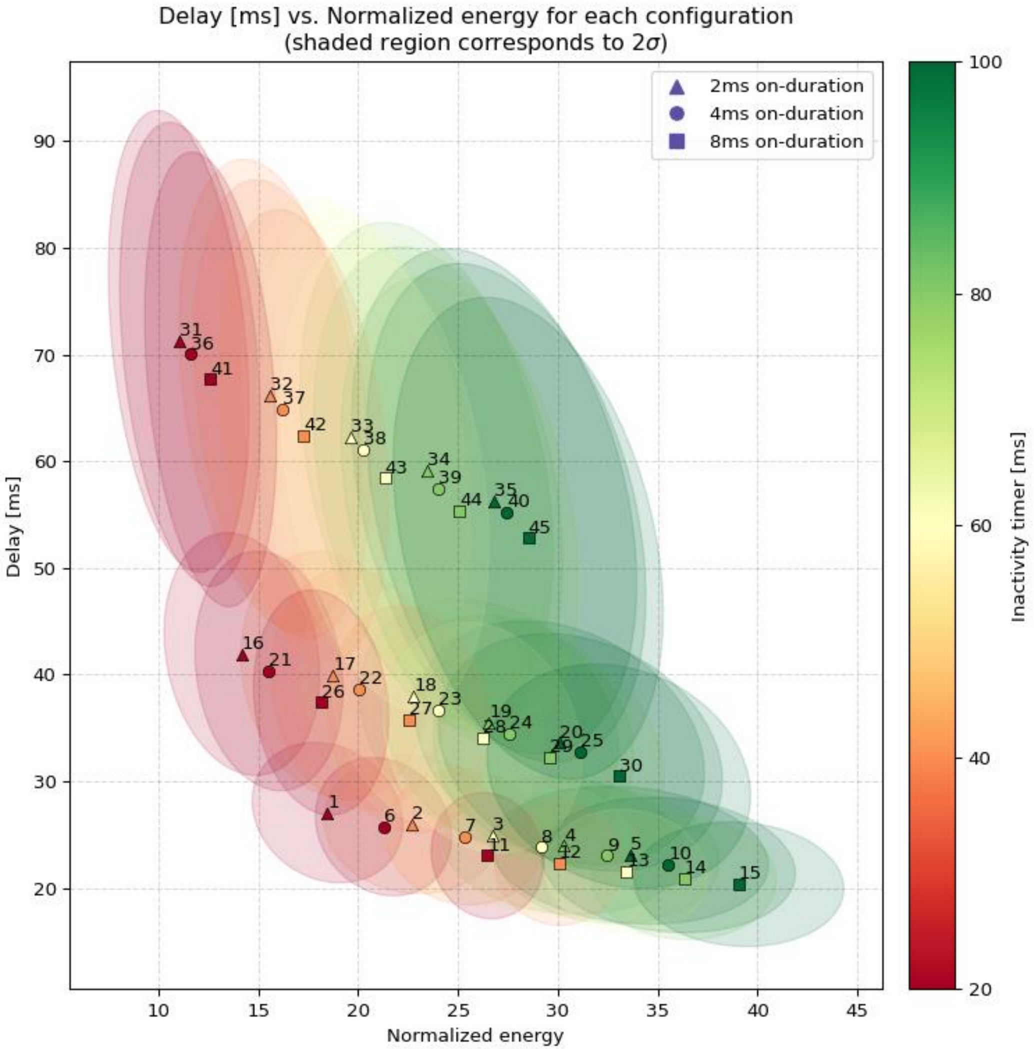
<u>Future work:</u>

- Consider other traffic types (e.g., real-time video)
- Explore softer thresholds
- Perform safe exploration

german.bassi@ericsson.com
philipp.bruhn@ericsson.com

Additional Slides

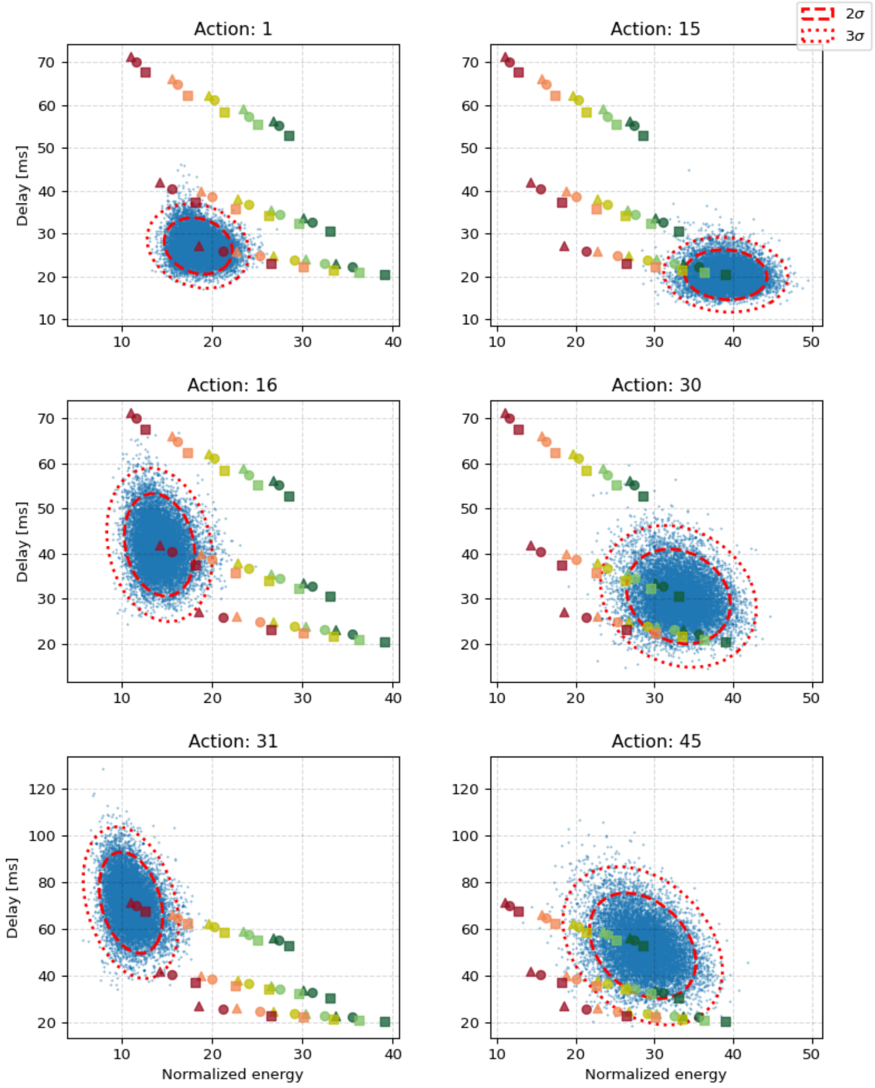
Statistics of the DRX configurations



By choosing a certain DRX configuration (a.k.a. action 1-45) the UE experiences a variable delay/energy consumption

The fine granularity of the different DRX parameters results in a large overlap between configurations

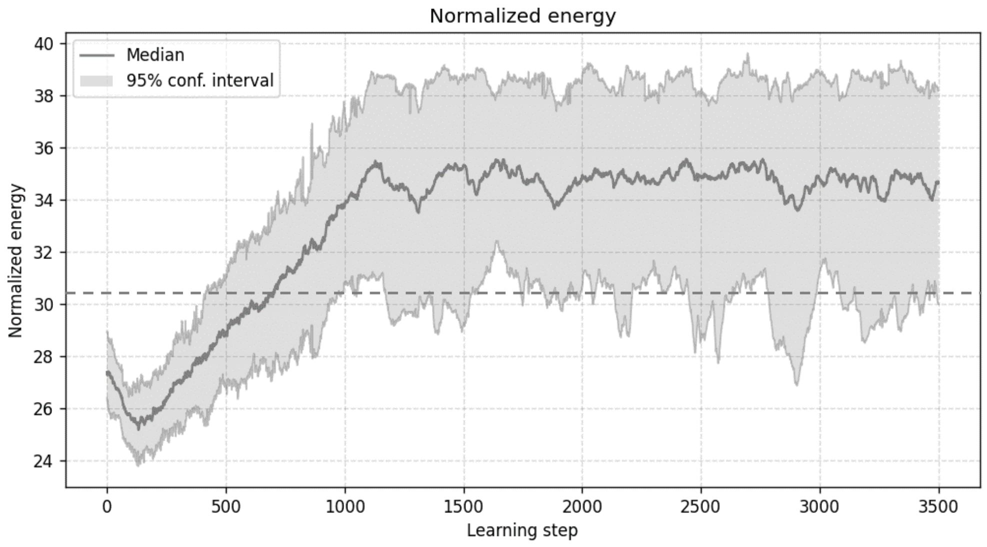
More in detail



Experiment #2b

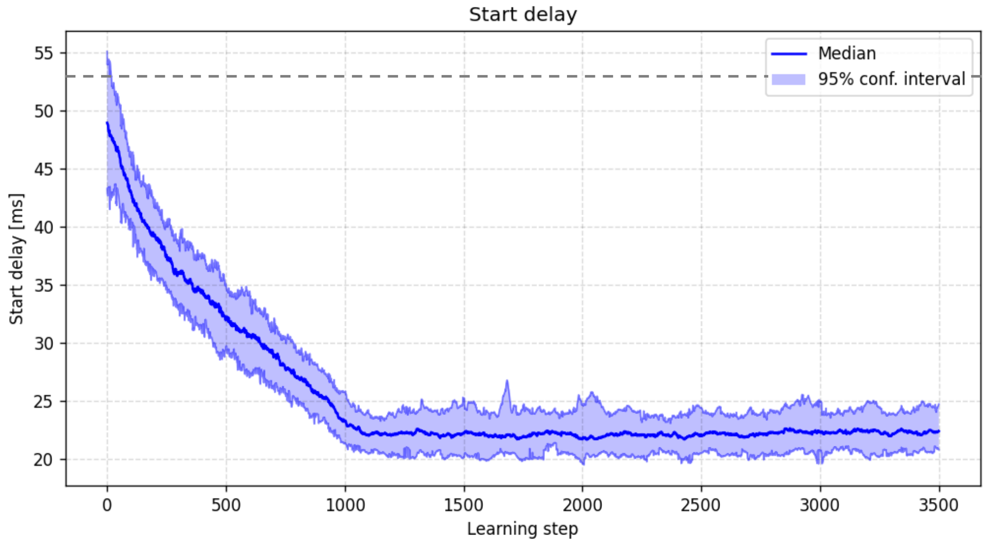
- Intent: Delay minimization
- Reward ~ $\frac{Min.\,Delay}{Delay} \in (0, 1)$

"Start delay" refers to delay of first segment of object transmission → Uncorrelated with SINR



28.4 Units

default DRX configuration



53 ms

default DRX configuration