# Deep Learning for Proactive Resource Allocation in LTE-U Networks

Ursula Challita*, Li Dong*, and Walid Saad†
*School of Informatics, The University of Edinburgh, Edinburgh, UK. Emails: {ursula.challita, li.dong}@ed.ac.uk.
†Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA. Email: walids@vt.edu.

***Abstract*—LTE in unlicensed spectrum (LTE-U) is a promising approach to overcome the wireless spectrum scarcity. However, to reap the benefits of LTE-U, a fair coexistence mechanism with other incumbent WiFi deployments is required. In this paper, a novel deep learning approach is proposed for modeling the resource allocation problem of LTE-U small base stations (SBSs). The proposed approach enables multiple SBSs to perform dynamic channel selection, carrier aggregation, and fractional spectrum access proactively while guaranteeing fairness with existing WiFi networks and other LTE-U operators. SBSs are modeled as Homo Egualis agents that aim at predicting a sequence of future actions and thus achieving long-term equal weighted fairness with WLAN and other LTE-U operators over a given time horizon. Simulation results using real data traces show that the proposed scheme can yield up to 28% gains over a conventional reactive approach. The results also show that the proposed framework prevents WiFi performance degradation for a densely deployed LTE-U network.**

## I. Introduction

LTE in unlicensed bands (LTE-U) has emerged as an effective solution to overcome the scarcity of the radio spectrum [1]. Using LTE-U, a cellular small base station (SBS) can access the unlicensed spectrum thus improving the overall network capacity and spectral efficiency. However, to achieve the promised quality-of-service (QoS) improvements from LTE-U, many challenges must be addressed ranging from effective co-existence with existing WiFi networks to resource allocation and multiple access over licensed and unlicensed bands [1].

If not properly deployed, LTE-U can significantly degrade the performance of the wireless local area network (WLAN) in the absence of an efficient spectrum sharing mechanism [1]. There has been a number of recent works [2]–[7] that investigated this challenge. This prior art can be categorized into two groups: channel access [2]–[4] and channel selection [6], [7]. The authors in [2] and [3] propose different channel access mechanisms based on listen-before-talk (LBT) that rely on either a fixed/random contention window (CW) size [2] or an adaptive CW size [3]. Nevertheless, a fixed CW size cannot handle time-varying traffic loads thus yielding unfair outcomes. The authors in [4] develop a holistic approach for both traffic offloading and resource sharing for one LTE-U SBS. In [5], the authors study the problem of resource allocation with uplink-downlink decoupling for LTE-U. However, none of these works jointly account for both channel selection and channel access. In other words, they do not analyze the potential gains that can be obtained upon aggregating or switching between different unlicensed channels.

In terms of LTE-U channel selection, the authors in [6] propose a matching-based solution, which is both distributed and cooperative. Moreover, the work in [7] combines channel selection along with channel access. Despite the promising results, the work in [6] and [7] consider a reactive sense-and-avoid approach that does not account for the future dynamics of the network and thus potentially incurring loss in terms of performance. On the other hand, in a *proactive* approach, rather than reactively responding to incoming demands and serving them when requested, SBSs can predict traffic patterns and determine future off-peak times so that incoming traffic demand can be properly allocated over a given time window and thus minimizing disruptions to WLAN.

The main contribution of this paper is to introduce a novel deep reinforcement learning algorithm based on long short-term memory (RL-LSTM) cells for proactively allocating LTE-U resources over the unlicensed spectrum. The LTE-U resource allocation problem is formulated as a noncooperative game in which the players are the SBSs. To solve this game, we propose an RL-LSTM framework which enables the SBSs to autonomously learn which unlicensed channels to use along with the corresponding channel access probability on each channel taking into account future environmental changes, in terms of WLAN activity on the unlicensed channels and LTE-U traffic loads. Unlike previous studies which are either centralized [7] or rely on the coordination among SBSs [3], our approach is based on a self-organizing proactive resource allocation scheme in which the SBSs utilize past observations to build predictive models on spectrum availability and intelligently plan channel usage over a finite time window. The use of LSTM cells enables the SBSs to predict a sequence of interdependent actions over a long-term time horizon thus achieving long-term fairness among different underlying technologies. Moreover, we show that the proposed framework converges to a mixed-strategy distribution which constitutes a mixed-strategy Nash equilibrium (NE) for the studied game. To the best of our knowledge, *this is the first work that exploits the framework of LSTMs for proactive resource allocation in LTE-U networks*. Simulation results show that the proposed approach yields significant rate improvements compared to conventional reactive solutions.

The rest of this paper is organized as follows. In Section II, we present the system model. Section III describes the proposed coexistence game model. The LSTM-based algorithm is proposed in Section IV. In Section V, simulation results are analyzed. Finally, conclusions are drawn in Section VI.

## II. System Model

Consider the downlink of an LTE-U network composed of a set $\mathcal{J}$ of $J$ LTE-U SBSs belonging to different LTE operators, a set $\mathcal{W}$ of $W$ WiFi access points (WAPs), and a set $\mathcal{C}$ of $C$ unlicensed channels. Each SBS $j \in \mathcal{J}$ has a set of $\mathcal{K}_j$ of $K_j$ LTE-U UEs associated with it. We focus on the operation of the SBSs over the unlicensed band, while the licensed spectrum resources are allocated in a conventional way. Both SBSs and WAPs adopt the LBT access scheme and, thus, at a particular time, a given unlicensed channel is occupied by either an SBS or a WAP. We consider the LTE carrier aggregation feature using which the SBSs can aggregate up to five component carriers.

Our goal is to jointly determine the dynamic channel selection, carrier aggregation, and fractional spectrum access for each

This research was supported by The University of Edinburgh Principal's Career Development PhD Scholarship and the U.S. National Science Foundation under Grants CNS-1460316 and IIS-1633363.

SBS, while guaranteeing long-term airtime fairness with WLAN and other LTE-U operators. We therefore need to dynamically analyze the usage of various unlicensed channels. To this end, we divide our time domain into multiple time windows, of duration $T$, each of which consists of multiple time epochs $t$. Our objective is to proactively determine the spectrum allocation vector for each SBS over $T$ while guaranteeing long-term equal weighted airtime share. To guarantee a fair spectrum allocation among SBSs belonging to different operators, we consider inter-operator interference along with inter-technology interference. Next, we define the variables $x_{j,c,t}$ =1 if channel $c$ is selected by SBS $j$ during time epoch $t$, and 0, otherwise, and $\alpha_{j,c,t} \in [0,1]$. $x_{j,c,t}$ determines which channel $c$ SBS $j$ is using during time $t$ and $\alpha_{j,c,t}$ is the channel access probability of SBS $j$ on the unlicensed channel $c$ during time $t$.

A contention-based protocol is used for the channel access over the unlicensed band. In this protocol, prior to transmission, the SBS applies clear channel assessment to detect the state of the channel (idle or busy) based on the detected energy level. If the channel is idle, the SBS gets a transmit opportunity for up to 10 LTE sub-frames; otherwise, it keeps monitoring the channel until it becomes idle. We consider an exponential backoff scheme for WiFi while the SBSs adjust their contention window size (and thus the channel access probability) on each of the selected channels in a way that would guarantee a long-term equal weighted fairness with WLAN and other SBSs.

To derive the throughput achieved by an LTE-U user equipment (UE) and a WAP, we first define the stationary probability of each WAP $w$ and each SBS $j$, $\tau_w$ and $\tau_{j,c,t}$ respectively. The stationary probability denotes the probability with which a given base station attempts to transmit in a randomly chosen slot. Considering an exponential backoff scheme for WiFi, the stationary probability with which WAPs transmit a packet, $\tau_w$, [8]:

$$\tau_w = \frac{2(1-2q_w)}{(1-2q_w)(\mathrm{CW}_{\min}+1) + q_w \mathrm{CW}_{\min}(1-(2q_w)^m)}, \quad (1)$$

where $q_w$ is the collision probability of a WAP, $m$ is the maximum backoff stage with $\mathrm{CW}_{\max} = 2^m \mathrm{CW}_{\min}$, and $\mathrm{CW}_{\min}$ and $\mathrm{CW}_{\max}$ are the minimum and maximum contention window size, respectively. For LTE-U, $m$=0 since no exponential backoff is considered, and, thus the stationary probability of an SBS on a given unlicensed channel $c$ during time epoch $t$ will be $\tau_{j,c,t} = \frac{2}{\mathrm{CW}_{j,c,t}+1}$, where $\mathrm{CW}_{j,c,t}$ is the contention window size of SBS $j$ on channel $c$ during time epoch $t$. Therefore, we do not consider a contention stage for LTE-U. Instead, the SBSs adjust their CW size adaptively to control their channel access probability over the unlicensed band. The collision probability of a WAP is defined as $q_w = 1 - \prod_{v=1,v\neq w}^{W}(1-\tau_v)\prod_{j=1}^{J}(1-\tau_{j,c,t})$, where $c$ is the channel used by WAP $w$. The throughput $R_w$ of a WAP $w$ will be:

$$R_w = \frac{P_{w,\text{succ}} \cdot E[D_w]}{P_{w,\text{idle}} \cdot \theta + P_{w,\text{busy}} \cdot T_b}, \quad (2)$$

where $E[D_w]$ is the expected payload size for WAP $w$, $P_{w,\text{succ}} = \tau_w \prod_{v=1,v\neq w}^{W}(1-\tau_v)\prod_{j=1}^{J}(1-\tau_{j,c,t})$ is the probability of a successful transmission, $P_{w,\text{idle}} = \prod_{j=1}^{J}(1-\tau_{j,c,t})\prod_{w=1}^{W}(1-\tau_w)$ is the probability of an idle slot, and $P_{w,\text{busy}} = 1 - \prod_{j=1}^{J}(1-\tau_{j,c,t})\prod_{w=1}^{W}(1-\tau_w)$ is the probability of a busy slot, regardless of whether it corresponds to a collision or a successful transmission. $\theta$ and $T_b$ are, respectively, the average durations of an idle and a busy slot and, thus, the denominator in (2) corresponds to the mean duration of a WiFi medium access control (MAC) slot.

On the LTE-U side, the achievable airtime fraction for an SBS $j$ on channel $c$ during time epoch $t$, can be expressed as:

$$\alpha_{j,c,t} = \tau_{j,c,t} \prod_{i=1,i\neq j}^{J}(1-\tau_{i,c,t}) \prod_{w=1}^{W}(1-\tau_w). \quad (3)$$

The airtime fraction essentially represents the time allocated for an SBS on channel $c$ during time $t$. Thus, the total throughput of all $K_{j,t}$ UEs that are served by SBS $j$ during time epoch $t$ is:

$$R_{j,t} = \sum_{c=1}^{C} \alpha_{j,c,t} r_{j,c,t}, \quad (4)$$

where

$$r_{j,c,t} = \sum_{k=1}^{K_{j,t}} B_c \log\Big(1 + \frac{P_{j,c,t} h_{j,k,c,t}}{I_{j,c,t} + B_c N_0}\Big). \quad (5)$$

Here, $I_{j,c,t} = \sum_{w=1}^{W} P_{w,c,t} h_{w,k,c,t} + \sum_{i=1,i\neq j}^{J} P_{i,c,t} h_{i,k,c,t}$ is the interference level on SBS $j$ when operating on channel $c$ during time $t$ and $B_c$ is the bandwidth of channel $c$. $P_{j,c,t}$ is the transmit power of SBS $j$ on channel $c$ during time $t$. $h_{j,k,c,t}$ is the channel gain between SBS $j$ and UE $k$ on channel $c$ during time $t$. $N_0$ is the power spectral density of additive white Gaussian noise. Since SBSs and WAPs both adopt LBT, one cell may then occupy the entire channel at a given time thus transmitting *exclusively* on a given channel $c$. However, hidden and exposed terminals could be present on a given channel which can result in interference and thus a degradation in the throughput.

Given this system model, next, we develop an effective spectrum allocation scheme that can allocate the appropriate unlicensed channels along with the corresponding channel access probabilities to each SBS simultaneously over $T$, at $t = 0$.

## III. Proactive Resource Allocation Scheme for Unlicensed LTE

In this section, we propose a proactive approach for allocating spectrum resources to SBSs in the unlicensed bands. In this regard, we formulate the resource allocation problem as a noncooperative game $\mathcal{G}$=($\mathcal{J}, \mathcal{A}_j, u_j$) with the SBSs in $\mathcal{J}$ being the players, each of which must choose a channel selection and channel access pair $a_{j,c,t}$=($x_{j,c,t}$,$\alpha_{j,c,t}$) $\in \mathcal{A}_j$ for each time $t$.

The objective of each SBS $j$ is to maximize its total throughput over the selected channels over $T$ given by:

$$u_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j}) = \sum_{t=1}^{T}\sum_{c=1}^{C} \alpha_{j,c,t} r_{j,c,t}, \quad (6)$$

where $\boldsymbol{a}_j = [(a_{j,1,1}, \cdots, a_{j,1,T}), \cdots, (a_{j,C,1}, \cdots, a_{j,C,T})]$ and $\boldsymbol{a}_{-j}$ correspond, respectively to the action vector of SBS $j$ and all other SBSs, over all the channels $\mathcal{C}$ during $T$. The goal of each SBS $j$ is to maximize its own utility:

$$\max_{a_j \in \mathcal{A}_j} u_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j}) \;\; \forall j \in \mathcal{J}, \quad (7)$$

$$\text{s.t.} \qquad \alpha_{j,c,t} \leq x_{j,c,t} \quad \forall c, t, \quad (8)$$

$$\sum_{c=1}^{C} x_{j,c,t} \leq \min(M_c, C) \quad \forall t, \quad (9)$$

$$\sum_{t_T=1}^{t}\sum_{c=1}^{C} \alpha_{j,c,t_T} B_c \leq \sum_{t_T=1}^{t} f(L_{j,t_T}) \quad \forall t, \quad (10)$$

$$\alpha_{w,c,t} + \alpha_{j,c,t} + \sum_{i=1,i\neq j}^{J} \alpha_{i,c,t} \leq t_{\max} \quad \forall c, t, \quad (11)$$

$$x_{j,c,t} \in \{0,1\}, \;\; \alpha_{j,c,t} \in [0,1] \quad \forall c, t. \quad (12)$$

where $M_c$ denotes the total number of unlicensed channels which an SBS can aggregate. Constraint (8) allows the allocation of a channel access proportion for SBS $j$ on channel $c$ during $t$ only if

SBS $j$ transmits on channel $c$ at time $t$. Constraint (9) guarantees that each SBS can aggregate a maximum of $M_c$ channels at a given time $t$. Constraint (10) limits the amount of allocated bandwidth to the required demand where $f(L_{j,t})$ captures the relationship between bandwidth requirement and offered load. (11) captures coupling constraints which limit the proportion of time used by SBSs and WLAN on a given unlicensed band to the maximum fraction of time an unlicensed channel can be used, $t_{\max}$. (12) represents the feasibility constraints.

Given the fact that different operators and technologies have equal priorities on the unlicensed spectrum, we incorporate the Homo Egualis (HE) anthropological model, an inequity-averse based fairness model, into the strategy design of the agents [9].

**Definition 1.** *Inequity aversion* is the preference for fairness and resistance to incidental inequalities. In other words, it refers to the willingness of giving up some material payoff in order to move in the direction of more equitable outcomes.

In an HE society, agents focus not only on maximizing their own payoffs, but also become aware of how their payoffs are compared to other agents' payoffs [9]. The HE concept comes from the anthropological literature in which Homo sapiens evolved in small hunter-gatherer groups without a centralized governance [9]. To model our players as HE agents, we consider the following two coupling constraints for the allocated airtime fraction on each channel $c$ for each SBS $j$:

$$\frac{1}{w_{j,c}}\frac{1}{T}\frac{\sum_{t=1}^{T}\alpha_{j,c,t}}{\sum_{t=1}^{T}\bar{L}_{j,t}}=\frac{1}{w_{i,c}}\frac{1}{T}\frac{\sum_{t=1}^{T}\alpha_{i,c,t}}{\sum_{t=1}^{T}\bar{L}_{i,t}}\ \forall c \in \widehat{\mathcal{C}}_j, i \in \widehat{\mathcal{S}}_{j,c}(i \neq j), \tag{13}$$

$$\frac{1}{T}\frac{\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\alpha_{n,c,t}}{P_{\text{LTE}}\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\bar{L}_{n,t}}=\frac{1}{T}\frac{\sum_{t=1}^{T}\alpha_{w,c,t}}{P_{\text{WiFi}}\sum_{t=1}^{T}L_{w,c,t}}\ \ \forall c \in \widehat{\mathcal{C}}_j, \tag{14}$$

where $\widehat{\mathcal{C}}_j$ is the subset of channels used by SBS $j$ during $T$. $\mathcal{S}_{c,t}$ is the subset of SBSs that are transmitting over channel $c$, $c \in \widehat{\mathcal{C}}_j$, during time $t$ and $\widehat{\mathcal{S}}_{j,c}$ is the subset of other neighboring SBSs, $i \neq j$, that are using the same channel $c \in \widehat{\mathcal{C}}_j$ as SBS $j$ during $T$. $\bar{L}_{j,t}$ corresponds to the remaining traffic that needs to be served by SBSs $j$ and can be expressed as $\bar{L}_{j,t} = L_{j,t} - \sum_{c'} f(\alpha_{j,c',t})$ where $L_{j,t}$ is the *total* aggregate traffic demand of SBS $j$ on channel $c$ during time epoch $t$. $f(.)$ corresponds to the served traffic load as a function of the airtime allocation. $c'$ in that case represents all the set of channels except channel $c$. $\alpha_{w,c,t}$= $\min(f(L_{w,c,t}),\ t_{\max} - \alpha_{j,c,t} - \sum_{i\in\mathcal{S}_{j,c,t}}\alpha_{i,c,t})$ is the airtime allocated for WLAN transmissions over channel $c$ during time $t$. $P_{\text{WiFi}}$ and $P_{\text{LTE}}$ correspond to the priority metric defined for each technology when operating on the unlicensed band. These parameters allow adaptation of the level of fairness between LTE-U and WLAN.

Constraint (13) represents inter-operator fairness which guarantees an equal weighted airtime allocation among SBSs belonging to different operators on a given channel $c$. The adopted notion of fairness is based on a long-term weighted equality over $T$, as opposed to instantaneous weighted equality. $w_{j,c} = \sum_{t=1}^{T} x_{j,c,t}$ is the weight of SBS $j$ on channel $c$ during $T$ and thus different SBSs are assigned different weights on each channel $c$ based on the number of time epochs $t$ a given SBS $j$ is active on that particular channel. (14) defines an inter-technology fairness metric to guarantee a long-term equal weighted airtime allocation over $T$ for both LTE-U and WiFi. Therefore, constraints (13) and (14) reflect the inequity aversion property of the SBSs. Here, we consider the allocated airtime as a metric for fairness in order to overcome the rate anomaly problem that arises when different nodes use distinct data rates [3], which leads to the slowest link limiting the system performance.

Our game $\mathcal{G}$ belongs to the family of generalized Nash equilibrium problems (GNEPs) in which both the objective functions and the action spaces are coupled. To solve the GNEP, we incorporate the Lagrangian penalty method into the utility functions thus reducing it to a simpler Nash equilibrium problem (NEP). The penalized utility function will be given by the following, $\forall j \in \mathcal{J}$:

$$\begin{aligned}
&\widehat{u}_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j}) = \sum_{t=1}^{T}\sum_{c=1}^{C}\alpha_{j,c,t}r_{j,c,t}\\
&-\rho_{1,j}\sum_{c=1}^{C}\sum_{t=1}^{T}\Big(\min(0, t_{\max} - \alpha_{w,c,t} - \alpha_{j,c,t} - \sum_{i=1,i\neq j}^{J}\alpha_{i,c,t})\Big)^2\\
&-\rho_{2,j}\sum_{c\in\widehat{\mathcal{C}}_j}\ \sum_{i\in\widehat{\mathcal{S}}_{j,c}(i\neq j)}\frac{1}{T^2}\left(\frac{1}{w_{j,c}}\frac{\sum_{t=1}^{T}\alpha_{j,c,t}}{\sum_{t=1}^{T}\bar{L}_{j,t}}-\frac{1}{w_{i,c}}\frac{\sum_{t=1}^{T}\alpha_{i,c,t}}{\sum_{t=1}^{T}\bar{L}_{i,t}}\right)^2\\
&-\rho_{3,j}\sum_{c\in\widehat{\mathcal{C}}_j}\frac{1}{T^2}\left(\frac{\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\alpha_{n,c,t}}{P_{\text{LTE}}\sum_{t=1}^{T}\sum_{n\in\mathcal{S}_{c,t}}\bar{L}_{n,t}}-\frac{\sum_{t=1}^{T}\alpha_{w,c,t}}{P_{\text{WiFi}}\sum_{t=1}^{T}L_{w,c,t}}\right)^2,
\end{aligned}$$

where $\rho_{1,j}$, $\rho_{2,j}$ and $\rho_{3,j}$ are positive penalty coefficients corresponding to constraints (11), (13), and (14) respectively. For our reformulation, we consider equal penalty coefficients for all players for each coupled constraint, $\rho_{1,j}$=$\rho_1$, $\rho_{2,j}$=$\rho_2$ and $\rho_{3,j}$=$\rho_3$. This allows all SBSs to have equal incentives to give up some payoff in order to satisfy the coupled constraints. Moreover, to determine the values of $\rho_1$, $\rho_2$ and $\rho_3$, we adopt the incremental penalty algorithm in [10] where it has been shown that there exists some penalty parameters $\boldsymbol{\rho}_l^*$=$[\rho_1^*, \rho_2^*, \rho_3^*]$ at which the coupled constraints can be satisfied.

In our game model $\mathcal{G}$, $\alpha_{j,c,t}$ is a continuous variable bounded between 0 and 1, however, for a particular network state, we are interested only in a certain region of the continuous space where the optimal actions are expected to be. Therefore, we will propose a sampling-based approach to discretize $\alpha_{j,c,t}$ in Section IV. Given that the action space becomes discrete, we turn our attention to mixed strategies in which players choose their strategies probabilistically. Such a mixed strategy approach enables us to analyze the frequency with which players choose different channels and channel access combinations. Let $\Delta(\mathcal{A})$ be the set of all probability distributions over the action space $\mathcal{A}$ and $\boldsymbol{p}_j$=$[p_{j,\boldsymbol{a}_1}\cdots, p_{j,\boldsymbol{a}_{|\mathcal{A}_j|}}]$ be a probability distribution with which SBS $j$ selects a particular action from $\mathcal{A}_j$. Therefore, our objective is to maximize $\overline{u}_j(\boldsymbol{p}_j, \boldsymbol{p}_{-j})$, the expected value of the utility function where $\overline{u}_j(\boldsymbol{p}_j, \boldsymbol{p}_{-j}) = \mathbb{E}_{\boldsymbol{p}_j}[\widehat{u}_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j})]$=$\sum_{\boldsymbol{a}\in\mathcal{A}}\widehat{u}_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j})\prod_{j=1}^{J}p_{j,\boldsymbol{a}_j}$.

**Definition 2.** A mixed strategy $\boldsymbol{p}^*$=$(\boldsymbol{p}_1^*, \cdots, \boldsymbol{p}_J^*)$=$(\boldsymbol{p}_j^*, \boldsymbol{p}_{-j}^*)$ is a *mixed-strategy Nash equilibrium* if, $\forall j \in \mathcal{J}$ and $\forall \boldsymbol{p}_j \in \Delta(\mathcal{A}_j)$, $\overline{u}_j(\boldsymbol{p}_j^*, \boldsymbol{p}_{-j}^*) \geq \overline{u}_j(\boldsymbol{p}_j, \boldsymbol{p}_{-j}^*)$.

Here, we note that any finite noncooperative game will admit at least one mixed-strategy Nash equilibrium [11]. However, solving for this equilibrium in our proposed game model is challenging due to the proposed long-term fairness notion which necessitates the prediction of a sequence of actions for each SBS over $T$ at $t = 0$ as well as a temporal dependence among these predicted actions. Therefore, next, we develop a novel deep

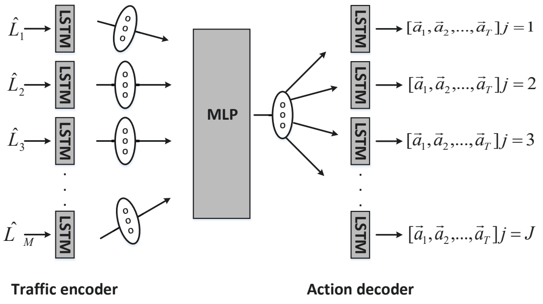
$\hat{L}_1$
$\hat{L}_2$
$\hat{L}_3$
$\hat{L}_M$
LSTM
MLP
$[\vec{a}_1, \vec{a}_2, ..., \vec{a}_T]j = 1$
$[\vec{a}_1, \vec{a}_2, ..., \vec{a}_T]j = 2$
$[\vec{a}_1, \vec{a}_2, ..., \vec{a}_T]j = 3$
$[\vec{a}_1, \vec{a}_2, ..., \vec{a}_T]j = J$
Traffic encoder
Action decoder

Fig. 1. Proposed framework.

learning algorithm for solving for the mixed-strategy NE of our game.

## IV. RL-LSTM for Resource Allocation

To proactively allocate resources on the unlicensed band, we propose a novel RL-LSTM algorithm that allows SBSs to learn a sequence of future actions based on a sequence of historic traffic load and thus maximizing the sum of their future rewards.

LSTMs are a special kind of "deep" recurrent neural networks (RNNs) that are capable of storing information for long periods of time and hence learning the long-term dependency within a given sequence [12]. In essence, LSTMs process a variable-length sequence $\boldsymbol{y} = (y_1,\, y_2,\, ...,\, y_m)$ by incrementally adding new content into a single memory slot, with gates controlling the extent to which new content should be memorized, old content should be erased, and current content should be exposed. Predictions at a given time step are influenced by the network activations at previous time steps thus making LSTMs suitable for our application in which an action at time $t$ depends on all previous and future actions within the current window $T$.

Consequently, we consider an end-to-end RL-LSTM based approach to train the network to find a mixed-strategy NE of the game $\mathcal{G}$ without any prior knowledge. Fig. 1 summarizes the proposed approach. The traffic encoder learns a vector representation of the input time-series (i.e., historical traffic loads), the multi-layer perceptron (MLP) summarizes the input vectors into one vector and the action decoder uses the summarized vector to reconstruct the predicted action sequence. In our scheme, an MLP is required to encode all the vectors together since a particular action at time $t$ depends on the values of all other input vectors (i.e., traffic values of all SBSs and WLAN on all the unlicensed channels). All SBSs share the same traffic encoders while different SBSs use different action decoders.

During the training phase, the parameters of the algorithm are learned from a given training data set. For our proposed framework, we train the weights of our neural network based on RL. We consider a policy gradient approach that is an RL technique that aims at maximizing the expected return of a policy. This is achieved by representing the policy by its own function approximator and updating it according to the gradient of the expected reward with respect to the policy parameters. Consider the set $\mathcal{M}$ of $M$ history traffic sequences corresponding to either an SBS or WiFi on each unlicensed channel, where $M = J + C$. Let $\boldsymbol{h}_{m,t} \in \mathbb{R}^n$ and $\boldsymbol{h}_{j,t} \in \mathbb{R}^n$ denote the hidden vectors of the traffic encoder $m$ and action decoder of SBS $j$, respectively, at time $t$. $\boldsymbol{h}_{m,t}$ and $\boldsymbol{h}_{j,t}$ are then computed by:

$$\boldsymbol{h}_{m,t}{=}\phi\left(\boldsymbol{v}_{m,t}, \boldsymbol{h}_{m,t-1}\right), \quad \boldsymbol{h}_{j,t}{=}\phi\left(\boldsymbol{v}_{j,t}, \boldsymbol{h}_{j,t-1}\right), \tag{15}$$

where $\phi$ refers to the LSTM cell function [12] being used, and $\boldsymbol{v}_{m,t}$ is the input vector. For the encoder, $\boldsymbol{v}_{m,t} = \left[\widehat{L}_{m,t}\right]$ is the history traffic value. For the decoder, $\boldsymbol{v}_{j,t} = [\boldsymbol{W}_d \boldsymbol{e}(\boldsymbol{x}_{j,t-1}) || \alpha_{j,c,t-1}]$ is the vector of the previous predicted action where $\boldsymbol{e}()$ maps discrete value to a one-hot vector, $\boldsymbol{W}_d \in \mathbb{R}^{n \times N_x}$ is a matrix that is used to transform the discrete actions $\boldsymbol{x}_{j,t-1}$ to a vector, and $N_x$ is the number of discrete actions. In our implementation, we learn the channel selection vector for all the channels simultaneously and thus $\boldsymbol{x}_{j,t} = [x_{j,1,t}, \cdots, x_{j,C,t}]$.

We use a softmax classifier to predict the distribution for the discrete variable $\boldsymbol{x}_{j,t}$ and a Gaussian policy for the distribution of the continuous variable $\alpha_{j,c,t}$. For the Gaussian policy, the probability of an action is proportional to a Gaussian distribution with a parameterized mean and a fixed value for the variance in our implementation. The variance of the Gaussian distribution defines the area around the mean from which we explore the action space. For our implementation, the initial value of the variance is set to $0.06$ in order to increase exploration and then is decreased linearly towards $0.02$. Therefore, defining probability distributions for our variables allows the initialization of the action space $\mathcal{A}_j$ by sampling $Z$ actions from the proposed distributions. This enables the SBSs to learn more accurate transmission probabilities for $\alpha_{j,c,t}$, as opposed to fixed discretization, thus satisfying the fairness constraints. The hidden vector $\boldsymbol{h}_{j,t}$ in the decoder is used to predict the $t$-th output actions $\boldsymbol{x}_{j,t}$ and $\alpha_{j,c,t}$. The probability vector over $\boldsymbol{x}_{j,t}$ and $\alpha_{j,c,t}$ can be defined, respectively, as:

$$\boldsymbol{x}_{j,t} | \boldsymbol{x}_{j,<t}, \alpha_{j,c,<t}, \widehat{\boldsymbol{L}}_t \sim \sigma\left(\boldsymbol{W}_x \boldsymbol{h}_{j,t}\right), \tag{16}$$

$$\mu_{j,c,t} = S\left(\boldsymbol{W}_\mu \boldsymbol{h}_{j,t}\right), \;\; \alpha_{j,c,t} \sim \mathcal{N}(\mu_{j,c,t}, \mathrm{Var}(\alpha_{j,c,t})), \tag{17}$$

where $\mu_{j,c,t}$ and $\mathrm{Var}(\alpha_{j,c,t})$ correspond to the mean value and variance of the Gaussian policy respectively, $\boldsymbol{W}_x \in \mathbb{R}^{|V_a| \times n}$, $\boldsymbol{W}_\mu \in \mathbb{R}^n$ are parameters, $\sigma(.)$ is the softmax function $\sigma(\boldsymbol{b})_q = \frac{e^{b_q}}{\sum_{o=1}^{O} e^{b_o}}$ for $q = 1, \cdots, O$, and $S(.)$ is the sigmoid function where $S(b) = \frac{1}{1+e^{-b}}$ and is used to normalize the value to $(0, 1)$. Note that $\alpha_{j,c,t}$ is computed only in the case when $x_{j,c,t} = 1$. The probability of the whole action sequence for SBS $j$, given a historic traffic sequence $\widehat{\boldsymbol{L}}$, $p_{j,\boldsymbol{a}_j|\widehat{\boldsymbol{L}}}$, is given by:

$$p_{j,\boldsymbol{a}_j|\widehat{\boldsymbol{L}}} = \prod_{t=1}^{T} p\left((\boldsymbol{x}_{j,t}, \alpha_{j,c,t}) | \boldsymbol{x}_{j,<t}, \alpha_{j,c,<t}, \widehat{\boldsymbol{L}}_t\right), \tag{18}$$

where $\widehat{\boldsymbol{L}}_t{=}(\widehat{L}_{1,t}, \cdots, \widehat{L}_{M,t})$, $\boldsymbol{x}_{j,<t}{=}[\boldsymbol{x}_{j,1}, \cdots, \boldsymbol{x}_{j,t-1}]$, and $\mu_{j,c,<t}{=}[\mu_{j,c,1}, \cdots, \mu_{j,c,t-1}]$.

Our goal is to maximize the exact expectation of the reward $\widehat{u}_j(\boldsymbol{a}_j, \boldsymbol{a}_{-j})$ over the action space for the training dataset. Therefore, the objective function can be defined as:

$$\max_{a_j \in \mathcal{A}_j} \sum_{\mathcal{D}} \mathbb{E}_{\boldsymbol{p}_j|\widehat{\boldsymbol{L}}} \left[\widehat{u}_j\left(\boldsymbol{a}_j, \boldsymbol{a}_{-j}\right)\right], \tag{19}$$

where $\mathcal{D}$ is the training dataset. For this objective function, the REINFORCE algorithm [13] can be used to compute the gradient, and then standard gradient descent optimization algorithms [14] can be adopted to allow the model to generate optimal action sequences for input history traffic values. Specifically, Monte Carlo sampling is used to compute the expectation.

On the other hand, the testing phase corresponds to the actual execution of the algorithm on each SBS. Based on history traffic values, each SBS learns the future sequence of actions based on the learned parameters from the training phase. Therefore, for applicability, we assume knowledge of historical measurements of the WiFi activity on each of the unlicensed channels through long-term channel sensing [4] and of other SBSs by exchanging history traffic information via the X2 interface. Consequently, the

Algorithm 1: Training phase of the proposed approach.

```
Input: $\mathcal{J}; \mathcal{W}; \mathcal{C}; \widehat{L}_{j,t} \forall j \in \mathcal{J}, t; \widehat{L}_{w,c,t} \forall c \in \mathcal{C}$ , $t$.
Initialization: The weights of all LSTMs are initialized following a uniform distribution
with arbitrarily small values.
Training: Each SBS $j$ is modeled as an LSTM network.
while Any of the coupled constraints is not satisfied do
    for Number of training epochs do
        for Size of the training dataset do
            Step 1. Run Algorithm 2 to compute the best actions for all SBSs.
            for $j$=1:$J$ do
                Step 2. Sample actions for SBS $j$ based on the best expected actions of
                other SBSs.
                Step 3. Use REINFORCE [13] to update rule and compute the gradient
                of the expected value of the reward function.
                Step 4. Update model parameters with back-propagation algorithm [15].
            end for
        end for
    end for
    Step 5. Using the incremental penalty algorithm, check the feasibility of the coupled
    constraints and update the values of $\boldsymbol{\rho}_l$ accordingly.
end while
```

Algorithm 2: Testing phase of the proposed approach.

```
Input: $\mathcal{J}; \mathcal{W}; \mathcal{C}; \widehat{L}_{j,t} \forall j \in \mathcal{J}, t; \widehat{L}_{w,c,t} \forall c \in \mathcal{C}$ , $t$.
for For each SBS $j$ do
    Step 1. Traffic history encoding: The history traffic of each SBS and WLAN activity
    on each channel is fed into each of the $M$ LSTM traffic encoders.
    Step 2. Vector summarization: The encoded vectors are transformed to initialize
    action decoders.
    Step 3. Action decoding: Action sequence is decoded for each SBS $j$.
end for
```

TABLE I
SYSTEM PARAMETERS

| Parameters | Values | Parameters | Values |
|---|---|---|---|
| Transmit power ($P_t$) | 20 dBm | BW (channel) | 20 MHz |
| CCA threshold | -80 dBm | Noise variance | 92 dBm/Hz |
| Path loss | $15.3 + 50\log_{10}(m)$ | SIFS | 16 $\mu$s |
| Hidden size (encoder) | 70 | DIFS | 34 $\mu$s |
| Hidden size (decoder) | 70 | $\text{CW}_{\text{min}}$ | 15 slots |
| time epoch ($t$) | 5 min | $\text{CW}_{\text{max}}$ | 1023 slots |
| Action sampling ($Z$) | 100 samples | ACK | 256 bits |
| History traffic size | 7 time epochs | $P_{\text{LTE}}$, $P_{\text{WiFi}}$ | 1, 1 |
| Learning rate | 0.01 | LSTM layers | 1 |
| Learning rate decay | 0.95 | $t_{\max}$ | 0.9 |

proposed algorithm offers a practical solution that is amenable to implementation. The training and the testing phases are given in Algorithms 1 and 2 respectively. In what follows, we characterize the convergence point of our proposed algorithm.

**Proposition 1.** If Algorithm 1 converges, then the convergence strategy profile corresponds to a mixed-strategy NE of game $\mathcal{G}$.

*Proof.* The resulting penalized utility function is an affine combination of convex functions, and hence is convex. Therefore, a gradient-based learning algorithm for our game $\mathcal{G}$ allows the convergence to an equilibrium point of that game [16]. Moreover, following the penalized reformulation of our game $\mathcal{G}$, one can easily show that a strategy that violates the coupled constraints cannot be a best response strategy. From [10], there exists $\boldsymbol{\rho}_l^*$ such that the incremental penalty algorithm terminates. Therefore, there exists a mixed strategy for which the coupled constraints are satisfied at $\boldsymbol{\rho}_l^*$. In that case, there is no incentive for an SBS to violate any of the coupled constraints, otherwise, its reward function would be penalized by the corresponding penalty function. Hence, all strategies that violate the coupled constraints are dominated by the alternative of complying with these constraints. Since in the proposed algorithm, the optimal strategy profile results in maximizing $\mathbb{E}_{\boldsymbol{p}_j}\left[\widehat{u}_j\left(\boldsymbol{a}_j, \boldsymbol{a}_{-j}\right)\right]$, we can conclude that the converged mixed-strategy NE is guaranteed not to violate the coupled constraints and hence it corresponds to a mixed-strategy NE for the game $\mathcal{G}$.

□

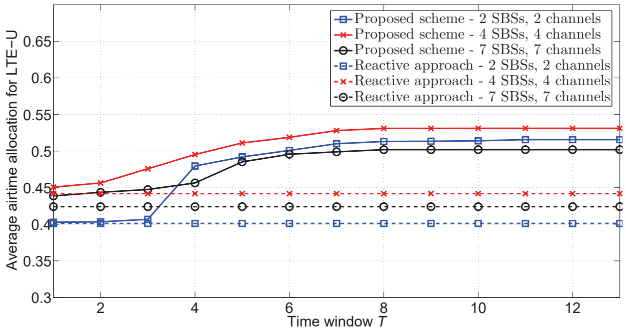

Fig. 2. Comparison of the average airtime allocated for LTE-U (with varying $T$) resulting from our proposed scheme as well as from a conventional reactive approach.

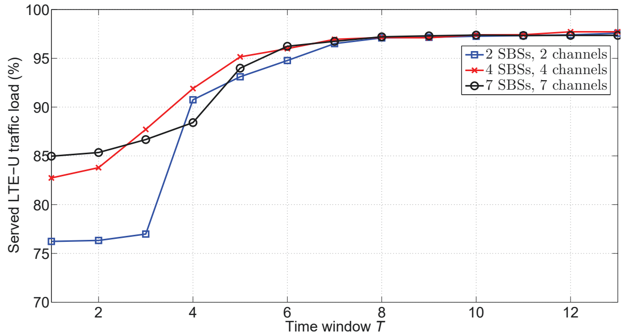

Fig. 3. The proportion of load served over LTE-U as a function of $T$.

## V. Simulation Results and Analysis

For our simulations, we consider a 300 m × 300 m square area in which we randomly deploy a number of SBSs and WAPs that share 7 unlicensed channels. We use real data for traffic loads from a dataset provided at [17] and divide it as 80% for training and 20% for testing. We consider WAPs to be passive such that their channel selection action is fixed and thus we characterize the activity on a given channel by the level of activity of WAPs. Table I summarizes the main simulation parameters. All statistical results are averaged over a large number of independent runs.

Fig. 2 shows the average airtime allocated for LTE-U as a function of $T$ for our proposed scheme as well as the conventional reactive approach under three different network scenarios. Intuitively, a larger $T$ provides the framework additional opportunities to benefit over the reactive approach, which does not account for future traffic loads. First, evidently, for very small $T$, the proactive approach does not yield any significant gains. However, as $T$ increases the gains start to become more pronounced. For example, for the case of 4 SBSs and 4 channels, the average airtime allocated for LTE-U increases from 0.45 to 0.52 as $T$ increases from 2 to 5, respectively, as opposed to 0.44 for the reactive approach. Eventually, as $T$ grows, LTE-U transmission opportunities of our proposed framework remains almost constant at the maximum achievable value.

In Fig. 3, we evaluate the proportion of LTE-U served load for different values of $T$. Fig. 3 shows that as $T$ increases, the proportion of LTE-U served traffic increases. For example, the proportion of served load increases from 82% to 97% for the case of 4 SBSs and 4 channels. Clearly, the gain of the LTE-U network stems from the flexibility of choosing actions over a large time horizon $T$. Instantaneous actions are taken based on the current traffic and future predictions of the traffic as opposed to a reactive approach that considers the current network state only. Therefore, the optimal policy will balance the instantaneous reward and the available information for future use and thus

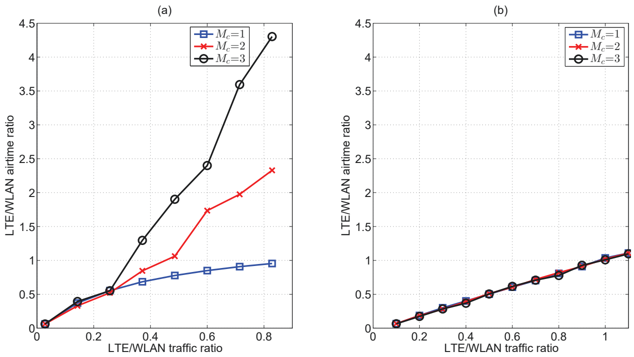

Fig. 4. LTE/WLAN airtime ratio as a function of the LTE/WLAN traffic ratio for 3 different values of $M_c$ ($M_c$=1, 2 and 3). The number of unlicensed channels is fixed to 7 and the number of SBSs is equal to 2 and 7 in (a) and (b) respectively.

maximizing the total load served over time. Based on the results given in Fig. 2 and Fig. 3, a suitable value of $T$ for the studied dataset is 8.

Fig. 4 shows the value of the LTE/WLAN airtime ratio under varying LTE/WLAN traffic ratio and for different values of $M_c$. We consider two different scenarios with varying number of SBSs (2 and 7 SBSs for scenarios (a) and (b) respectively), while the number of unlicensed channels is fixed to 7. Fig. 4 shows that inter-technology fairness is satisfied. This can be clearly seen in scenario (b) for the case of $M_c$=1. For instance, when the traffic ratio is 1, LTE/WLAN airtime ratio is 1 and thus equal weighted airtime is allocated for each technology (given that $P_{\text{LTE}}$=1 and $P_{\text{WiFi}}$=1). From Fig. 4, we can also see that enabling carrier aggregation impacts the resource allocation outcome. In fact, we can see that a considerable gain in terms of spectrum access time can be achieved with carrier aggregation. For instance, in the case of 2 SBSs and 2 channels, the LTE/WLAN airtime ratio increases from 0.84 for $M_c$=1 to 1.7 and 2.4 for $M_c$=2 and 3 respectively for the value of 0.6 for LTE/WLAN traffic ratio. On the other hand, this gain decreases as more SBSs are deployed and for a densely deployed LTE-U network, there is no need to aggregate more channels. This can be seen from (b) where the LTE-U network gets the same airtime share for $M_c$=1, 2 and 3.

Moreover, Fig. 4 shows that deploying more SBSs does not necessarily allow more airtime fraction for the LTE-U network. For example, LTE/WLAN airtime ratio of scenarios (a) and (b) corresponding to 0.6 LTE/WLAN traffic ratio is equal to 0.84 and 0.6 respectively for $M_c$=1. Consequently, the proposed scheme can avoid causing performance degradation to WLAN in the case LTE operators selfishly deploy a high number of SBSs.

Fig. 5 investigates the proportion of served LTE-U traffic for different network parameters. From Fig. 5, we can see that, as the number of SBSs increases, the proportion of LTE-U served traffic, relative to its corresponding offered load decreases thus avoiding degradation in the WLAN performance in the case of a densely deployed LTE-U network. Moreover, reducing the number of unlicensed channels leads to a decrease in the proportion of LTE-U served traffic. Although the number of available unlicensed channels are not players in the game, they affect spectrum allocation action selection for each SBS. As the number of channels increases, the action space for the channel selection vector increases, thus giving more opportunities for an SBS to serve more of its offered load.

## VI. Conclusion

In this paper, we have proposed a novel resource allocation framework for the coexistence of LTE-U and WiFi in the unlicensed band. We have formulated a game model where each SBS seeks to maximize its rate over a given time horizon while achieving long-term equal weighted fairness with WLAN and other LTE-U operators transmitting on the same channel. To solve this problem, we have developed a novel algorithm based on LSTMs. The proposed algorithm enables each SBS to decide on its spectrum allocation scheme autonomously with limited information on the network state. Simulation results have shown that the proposed approach yields significant performance gains in terms of rate compared to a conventional approach that considers only instantaneous network parameters.

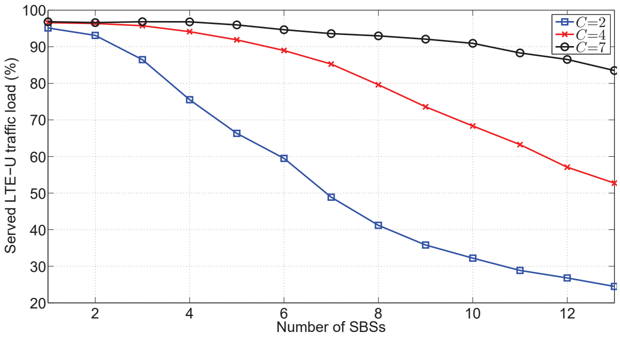

Fig. 5. The proportion of LTE-U served traffic load as a function of the number of SBSs and for different number of unlicensed channels ($C$=2, 4, and 7).

## References

[1] R. Zhang, M. Wang, L. Cai, Z. Zheng, X. Shen, and L. Xie, "LTE-unlicensed: the future of spectrum aggregation for cellular networks," *IEEE Wireless Communications*, vol. 22, pp. 150–159, June 2015.

[2] A. Mukherjee, J.-F. Cheng, S. Falahati, L. Falconetti, A. Furuskr, B. Godana, D. H. Kang, H. Koorapaty, D. Larsson, and Y. Yang, "System architecture and coexistence evaluation of licensed-assisted access LTE with IEEE 802.11," in *Proc. of IEEE International Conference on Communications (ICC)*, London, UK, June 2015.

[3] C. Hasan, M. K. Marina, and U. Challita, "On LTE-WiFi coexistence and inter-operator spectrum sharing in unlicensed bands: Altruism, cooperation and fairness," in *Proc. of ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, Paderborn, Germany, July 2016.

[4] U. Challita and M. K. Marina, "Holistic small cell traffic balancing across licensed and unlicensed bands," in *Proc. of the 19th ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWiM)*, Malta, Nov. 2016.

[5] M. Chen, W. Saad, and C. Yin, "Echo state networks for self-organizing resource allocation in LTE-U with uplink-downlink decoupling," *IEEE Transactions on Wireless Communications*, vol. 16, pp. 3–16, Jan. 2017.

[6] Y. Gu, Y. Zhang, L. X. Cai, M. Pan, L. Song, and Z. Han, "Exploiting student-project allocation matching for spectrum sharing in LTE-Unlicensed," in *in Proc. of IEEE Global Communications Conference (GLOBECOM)*, San Diego, CA, USA, Dec. 2015.

[7] Z. Guan and T. Melodia, "CU-LTE: Spectrally-efficient and fair coexistencebetween LTE and Wi-Fi in unlicensed bands," in *Proc. of IEEE Conference on Computer Communications (INFOCOM)*, San Francisco, CA, USA, Apr. 2016.

[8] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[9] H. Gintis, *Game theory evolving: A problem-centered introduction to modeling strategic behavior.* Princeton University Press, 2000.

[10] M. Fukushima, "Restricted generalized Nash equilibria and controlled penalty algorithm," *Computational Management Science (CMS)*, vol. 8, no. 3, pp. 201–218, Aug. 2011.

[11] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, pp. 286–295, 1951.

[12] W. Zaremba, I. Sutskever, and O. Vinyals, "Recurrent neural network regularization," in *Proceedings of International Conference on Learning Representations (ICLR)*, San Diego, CA, May 2015.

[13] R. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Machine Learning*, vol. 8, pp. 229–256, May 1992.

[14] R. Sutton, D. McAllester, S. Singh, and Y. Mansour, "Policy gradient methods for reinforcement learning with function approximation," *Advances in Neural Information Processing Systems*, vol. 12, pp. 1057–1063, 2000.

[15] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, Oct. 1986.

[16] K. J. Arrow and L. Hurwicz, "Stability of the gradient process in n-person games.," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, pp. 280–295, June 1960.

[17] M. Balazinska and P. Castro, "IBM Watson Research Center." CRAWCAD, Feb. 2003.