Multi-Objective Reinforcement Learning for Cognitive Radio–Based Satellite Communications

Paulo Victor R. Ferreira,* Randy Paffenroth,† and Alexander M. Wyglinski‡
Worcester Polytechnic Institute, Worcester, MA 01609, USA

Timothy M. Hackett§ and Sven G. Bilén¶
The Pennsylvania State University, University Park, PA 16802, USA

Richard C. Reinhart‖ and Dale J. Mortensen‖
NASA John H. Glenn Research Center, Cleveland, OH 44135, USA

Previous research on cognitive radios has addressed the performance of various machine-learning and optimization techniques for decision making of terrestrial link properties. In this paper, we present our recent investigations with respect to reinforcement learning that potentially can be employed by future cognitive radios installed onboard satellite communications systems specifically tasked with radio resource management. This work analyzes the performance of learning, reasoning, and decision making while considering multiple objectives for time-varying communications channels, as well as different cross-layer requirements. Based on the urgent demand for increased bandwidth, which is being addressed by the next generation of high-throughput satellites, the performance of cognitive radio is assessed considering links between a geostationary satellite and a fixed ground station operating at Ka-band (26 GHz). Simulation results show multiple objective performance improvements of more than 3.5 times for clear sky conditions and 6.8 times for rain conditions.

I. Introduction

Software-defined radio (SDR) technology has enabled numerous advances in wireless communications technologies for both terrestrial and satellite-based communications. For example, several channel access and fading-mitigation technologies for spectrum sensing[1] and adaptive coding and modulation (ACM)[2] have been developed and tested using SDR prototypes. However, the next technological leap for SDR is expected to be the implementation of onboard cognition, which can potentially provide the SDR platform with environmental awareness across several Open Systems Interconnection (OSI) layers,[3] real-time knowledge of the current channel conditions, status of its communicating nodes and the network, and assessment of its own available resources. Such information is very important for optimizing communications link performance.

Within the past 15 years or so, several optimization algorithms such as genetic algorithms[4] have been applied to cognitive radios (CRs).[5] These optimization techniques do not always converge and might take several hundred or even thousands of iterations, e.g. algorithm runs, until a stable solution is found. Changes in the environmental conditions would require another series of iterations in order to search for a new solution, increasing latency. Also, if these same scenarios repeat, the system is unable to use previous solutions given that it cannot learn. Learning is considered to the be cornerstone of artificial intelligence (AI). Therefore,

*Graduate Research Assistant, Electrical and Computer Engineering Dept., 100 Institute Rd., Worcester, MA.
†Professor, Mathematical Sciences Dept. and Data Science Program, 100 Institute Rd., Worcester, MA.
‡Professor, Electrical and Computer Engineering Dept., 100 Institute Rd., Worcester, MA. alexw@wpi.edu
§Graduate Research Assistant, School of Electrical Engineering and Computer Science, 304 Electrical Engineering East, University Park, PA.
¶Professor, School of Electrical Engineering and Computer Science, 313 Electrical Engineering East, University Park, PA.
‖Senior Communications Engineer, 21000 Brookpark Rd., Cleveland, OH.

CR systems must somehow consider its principles, thorough technique implementations, in order to leverage true CR capabilities.

Machine learning (ML) techniques have been studied with respect to their approach to CR, thus addressing the learning issue.[6–9] Both approaches, optimization and ML, have been studied in order to determine how they can assist CR systems with respect to finding the best configuration parameter set, with the majority of research focused on spectrum management, including sensing techniques, for terrestrial links.[9] The scenario is the same for satellite links, with the majority of CR research focused on spectrum resource allocation,[10,11] and to the best of the authors' knowledge, almost none on radio resource management.

Consequently, the aim of this paper is to analyze the framework of a cognitive-engine that enables a satellite-based CR to learn, reason, and make decisions over multiple available resources and multiple goals based on its past experiences. Regarding a satellite communications channel, some additional factors play a significant role that are not considered in terrestrial links including orbital dynamics, such as spacecraft trajectory, velocity, and antenna elevation angle profile, space weather, spacecraft mission, and payload status.

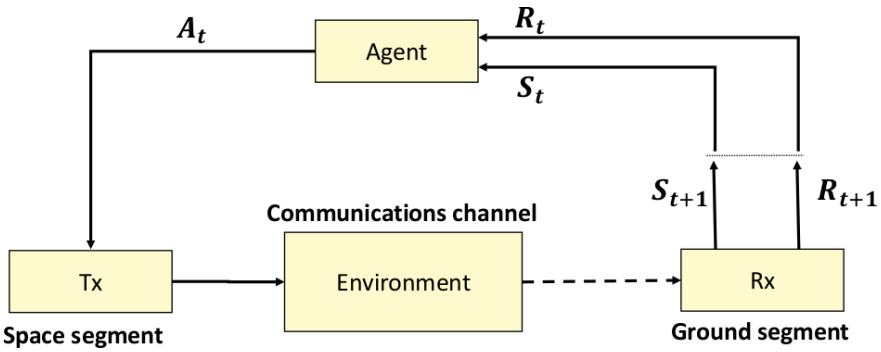

Figure 1. Diagram block of reinforcement learning elements within the feedback loop for a down-link pair with the receiver at ground and transmitter in space. A_t, R_t and S_t are the action, state, and reward at time t, respectively.

All of these factors magnify the complexity of the decision-making process being performed by the CR since the dimensionality of the multi-objective function, radio resources, and end-to-end link conditions are much more diverse than a point-to-point terrestrial link. Not only does the resource management complexity for satellite-based CR increase, but also the learning, reasoning, and decision-making algorithms that must be redesigned in order to cope with all these new challenges while operating in a dynamically changing and complex environment.

Based on the increasing demand for high-throughput satellite-based communications systems, the need for CR systems operating on board satellites possessing different orbits at the same time, with nodes spread throughout the Earth experiencing very different channel conditions, is of high importance and requires additional research.

In 2012, NASA Glenn Research Center (GRC) installed the Space Communications and Navigation (SCaN) Testbed[12] on-board the International Space Station (ISS) and continues to operates the testbed from a mission control center at GRC. The SCaN Testbed is an experimental communication system comprised of three different SDRs,[13,14] and is the first space-based testbed of its kind available to selected researchers to propose implementation solutions to address issues related to SDR-based communications systems to and from space. The radios operate at S-band and Ka-band with NASA's satellite relay infrastructure *i.e.*, Tracking and Data Relay Satellite System (TDRSS), and S-band direct-to-ground stations on Earth. The communication links to and from the SCaN Testbed provide real-world satellite dynamics between user spacecraft and relay satellites and user spacecraft directly with ground stations. These dynamics include time-varying Doppler changes, thermal variations, differences in waveform characteristics, real-time interference, significant range variation, ionospheric effects and scintillation, and other propagation impairments.

The radios onboard the testbed are flight-grade systems, fully compliant with NASA's SDR architecture, the Space Telecommunications Radio System (STRS).[15,16] The STRS Architecture provides abstraction interfaces between radio software and proprietary hardware (general-purpose processors and field-programmable gate array processors), allowing third-party software waveforms and applications such as the CR system software to interact and run on the radio. STRS also makes a library of waveforms available to developers to provide various modulation, coding, framing, and data rate options (the so-called radio "knobs") available to the decision-making algorithms.

Solutions such as adaptive communications based on cognitive decision making needs to be researched, not

only to solve communications issues on Earth, but also to allow the development of systems that will enable space exploration in the near future.[17] SDRs will become a larger part of the communications infrastructure as exploration continues over the next decade. These flexible and reconfigurable systems are generally more complex than traditional fixed-operation radios and CR systems offer a solution to reduce the complexity and risk associated with these new systems. The work presented in this paper is part of this effort, providing CR algorithm research and development for SDR systems in space.

This paper covers the performance analysis of a reinforcement learning (RL) technique for satellite-based CR at Ka-band (26 GHz). It starts with an overview of RL in Section II. The proposal framework design and a numerical evaluation are provided in Section III. The system environment is covered in Section IV. Simulation results and performance analysis are provided in Section V. Section VI provides an overview on the application of cognitive engines for NASA. Final comments and concluding remarks of the proposed approach are given in Section VII.

II. Reinforcement learning overview

The classical multi-armed bandit (MAB) problem[18,19] is one approach for modeling problems in which one must choose the best combination of options that will result in the largest reward. In MAB problems, it is assumed that K independent arms generate independent rewards from a family of probability distributions each time an arm is played. A policy chooses the arms to be played, known as the action, which results in a regret each time a non-optimal arm is played. Several publications have applied MAB to CR-related issues. For instance, references 20–22 modeled cognitive-medium access problems as a MAB problem.

A MAB problem can also be modeled as a state-transition problem. As such, a chosen action might end up causing a system to change its state, given a certain environment. Thus, the state–change itself can be modeled as a Markov decision process (MDP).[23] State transitions can be deterministic, *i.e.*, a given action taken while in a given state always brings the system to the same state or stochastic, *i.e.*, next state is a random variable. This paper considers the deterministic case only. MDPs are also categorized based on its model. If a state-transition model exists, the algorithm employed is known as Dynamic Programming (DP). If the MDP is model-free, it is known as a Reinforcement Learning (RL) approach. Within the RL class, the algorithm starts without a model, and either builds one online or does not use a model at all, in which case it is known as an RL Q-learning algorithm.

RL, thoroughly described in reference 24, is one of several techniques used to solve MAB problems. It is designed to learn and optimize while operating in an unknown environment without prior knowledge. Thus, RL is among the best options to be used when a system model is too difficult/complex to obtain, or inexistent. This technique aims to provide an agent, learning about an environment with the goal of building the "best model" of the "real-world" using noisy observations. After building the model, the system attempts to pick the best action that brings it closest to the operational goal, based on the set of actions available to it at the moment. Next, after choosing an action based on a certain policy, the system interacts with the environment and, as a result, a new state is observed together with a new reward. The reward is a metric of how close to the operational goal the chosen action brought the system.

As shown by Figure 1, a RL-based system has an agent element, which chooses an action using a certain policy at an attempt to optimize its rewards. While doing this, the employed policy might tackle the trade-off issue between exploration and exploitation. When trying new actions randomly, the system is exploring. When exploiting, it is choosing those actions previously attempted that resulted in the best rewards so far, given the expected state. The attractiveness of this approach is that, independent of being in either an exploiting or exploring mode, the system is always reinforcing its learning for a certain state in terms of rewards for actions taken while operating in that state.

The way RL records its performance is through the computation of a Q-function for a Q-learning algorithm:

$$Q_{k+1}(s_k, u_k) = Q_k(s_k, u_k) + \alpha_k[r_{k+1} + \gamma Q_k(s_{k+1}, u_{k+1}) - Q_k(s_k, u_k)], \tag{1}$$

which is derived from the well-known Bellman equations.[1] The Q-function computes the updated Q-values, $Q_{k+1}(s_k, u_k)$, for the state-action pair (s_k, u_k), where s_k and u_k are the state and action, respectively, at instant k. The reader may notice $u_{u'}$, which means any action. $\alpha_k \in [0, 1)$ is the learning rate at instant k, $\gamma \in [0, 1)$ is the discount factor, and r_{k+1} is the reward received after the transition from state s_k to s_{k+1}.

[1]The interested reader is referred to references 23, 24 for the derivations

γ weights how the agent accounts for accumulated rewards, *i.e.*, obtained after a large number of steps that can be taken in future given that a certain action is taken in the present. Choosing values for both α and γ involves a trade-off between the quality of the solution and the convergence rate and are left as designer parameters. In this paper it was assumed that α decreases every time an action is used recursively.

Within the context of decision-making of radio communication parameters, the time-difference portion of Eq. 1 can be taken for granted. Based on the idea of taking the best known action, based on accumulated knowledge over the past, in order to achieve the best performance on the next step, completely removes the need of a discount factor for possible actions that could be taken from the next one. In other words, the achievement of any performance level is independent of previous actions already taken, besides the knowledge they have provided. Thus, a modified version of Eq. 1 may be used, *i.e.*,

$$Q_{k+1}(s_k, u_k) = Q_k(s_k, u_k) + \alpha_k r_{k+1}. \tag{2}$$

Although a model for the environment is not necessary, it is required for the definition of some functions, for instance, a state–action policy h

$$u_k = h(s_k), \tag{3}$$

responsible for the trade-off between exploration and exploitation, a state-transition function g

$$s_{k+1} = g(s_k, u_k), \tag{4}$$

a reward function ρ

$$r_{k+1} = \rho(s_k, u_k), \tag{5}$$

and some parameters as described above. It is worth noting that for a model-free RL g is not required, since the s_{k+1} is simply observed.

There are several challenges regarding the practical implementation of RL. Among them are the choices of learning and discount parameters, as well as the definition of Eqs. (3)–(5). Regarding the policy function, there are three algorithms that define how the policy is defined: value iteration (searches for the maximum Q-value, which is used to compute the optimal policy), policy iteration (iterates through a set of policies, evaluates their results, and builds a new set to be evaluated in the next iteration), and policy search (computationally heavy and consists of using optimization techniques to search for the optimal policy directly).

Regarding policy iteration, policy evaluation can be performed off-line (off-policy), with the evaluation being done only after the Q-function converges, or on-line (on-policy), which is known as SARSA[2] and replaces $\max_{u'} Q_k(s_{k+1}, u_{u'})$ by $Q_k(s_{k+1}, u_{k+1})$, considering the action chosen for the next step. All the following work considers the class of deterministic RL Q-learning on-policy, *i.e.*, the SARSA category with the caveat that there is no time difference term nor discount factor, for the reasons mentioned above.

III. Framework Design

A. Goals and Rewards

Within a general communication system, the RL agent is located at the receiver and feeds the output control, *i.e.*, the action, back to the transmitter, as shown by Figure 1. A key requirement for any RL algorithm is to define its goal. In this case the authors propose a goal given by a decision function, based on the current communications mission-phase objectives, constraints, and the overall maximum performance possible to achieve, which is limited by hardware. Thus, the approach taken here consists of defining that state as a percentage of the current goal. Then, the RL agent might choose actions, *i.e.*, a new set of radio parameters that may bring this system closer to the "goal state", represented by 100% of the current goal. A threshold must be set so that any chosen action that brings the system performance above the threshold is recognized as part of the solution set by a reward. These actions are described in the following subsection.

In this paper communication channels are assumed to be a Markov process, with the current channel state being independent from the previous states. Thus, every time an RL agent chooses an action the system can be sent to any possible state. Besides choosing actions, another role of the RL agent is to build a knowledge base based on previous experiences, *i.e.*, to learn. In order to keep track of results of its chosen actions, as well as to support future action choices, the agent receives rewards, which measure how close to the goals the system was able to get using a certain action.

[2]Initials of the elements in the data tuples: state, action, reward, (next) state, (next) action (SARSA).

Only actions that bring the system from any previous state to a state above the defined threshold receive a reward, otherwise no reward is given. Doing so, the system "reinforces" actions that make the system achieve its goals, as illustrated by Figure 2.

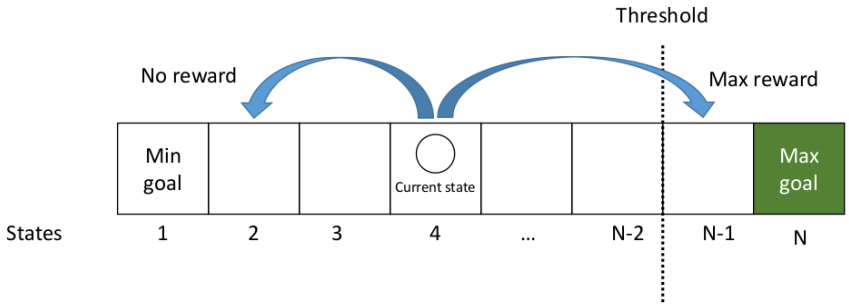

Figure 2. State change concept for communications system Markov process. Rewards are given for actions that sends the system to states that are above a certain threshold. No rewards are given otherwise. Even if the previous state was rewarded, if a different action results in an state above threshold it must be rewarded.

In this paper, the goals for a communications system are: maximizing throughput (R), minimizing bit error rate (BER), minimizing power consumption (P) to maximize on-board satellite battery life, and keeping the bandwidth W constant.[3] An analysis is provided based on the proposed approach for using RL to handle the cases when multiple goals are required during a certain satellite communications mission phase.

These goals are achieved individually by parameter adaptation, such as bit rate R, modulation scheme M (hereafter comprised of modulation type and index, and coding rate), available power $P_{\max}$ for transmission (then E_b is enabled to adapt as well). In the following, we provide relationships between the adaptable parameters, presented as equations for the individual goals.[25] The parameter W can be computed by

$$W = \frac{NR}{2\log_2(M)}, \tag{6}$$

where N is the number of orthogonal dimensions in the modulation constellation, e.g., $N = 2$ for QAM. M is the constellation size of the modulation scheme being used. P is given by

$$P = R\ E_b, \tag{7}$$

where E_b is the energy per bit. For two-dimensional modulation schemes, the parameter P can be rewritten as

$$P = W\ \log_2(M)\ E_b. \tag{8}$$

Based on the bit error probability equations provided in reference 25, BER relates to the previous equations by

$$BER \approx \frac{1}{E_b}, \tag{9}$$

for a fixed M, and by

$$BER \approx \log_2(M), \tag{10}$$

for fixed E_b. When both M and E_b are varied at the same time, there is a nonlinear relationship with BER.

Thus, the cognitive engine must learn how to tune these parameters while considering other objectives. Clearly, there are several trade-offs while adapting R, M, P, W, and BER in order to achieve a certain goal. Adaptation of these parameters for one goal might affect the achievement of other goals at the same time. These adaptation consequences are described in Table 1.

Attempting to achieve multiple goals at the same time can cause resource competition. An approach to minimize this dispute is to prioritize each goal individually using weights. Thus, communications mission phases are defined such that, for each mission, a predefined set of weight values, vector $\bar{w}$, is used during the decision-making process. Therefore, the goal of the approach presented is to maximize

$$f_{\text{decision}_{\max}}(\mathbf{x}) = w_1 f_{\max}(R) + w_2 f_{\min}(\text{BER}) + w_3 f_{\min}(P) + w_4 f_{\text{constant}}(W), \tag{11}$$

with $\bar{w_k} = (w_{1_k}, w_{2_k}, w_{3_k}, w_{4_k})$, $0 < w_{1_k}, w_{2_k}, w_{3_k}, w_{4_k} < 1$, $\sum_{i=1}^{n} \bar{w_{i_k}} = 1$, and $\bar{x_k} = (R_k, \text{BER}_k, P_k, W_k)$, which is a vector containing the observed values after choosing an action $\bar{u_k} = (R_k, M_k, E_{b_k})$ containing

[3] Although a decrease in W could be beneficial for cooperation, for critical missions it could end up attracting secondary users that could cause interference.

Table 1. Radio parameter adaptation inter-relationship

Adaptation	Goals	Consequences	Conflicts	Constants
$\downarrow M$	min(BER), min(P)	$\downarrow R$,P,BER	max(R)	E_b
$\uparrow M$	max(R)	$\uparrow R$,P,BER	min(BER), min(P)	E_b
$\uparrow R$	max(R), W const.*	$\uparrow W$,P	min(P), W const.*	M, E_b
$\downarrow R$	min(P), W const.*	$\downarrow W$,P	max(R), W const.*	M, E_b

* Keeping W constant can be a goal or a conflicting goal while adapting R_s.

the adaptable parameters. Among the observed values, values of R_k, P_k, and W_k are monitored at the transmitter and sent to the receiver, where BER is estimated in real time at the receiver, by using Eq. 17 in reference 26, which requires ${}^{E_b}/_{N_0}$ to be obtained by SNR measurements.

$f_{\max,\min}$ is a normalized value of each element of $\bar{x}$, based on the current system's available parameter ranges,

$$f_{\max}(R) = \frac{R}{R_{max}}, \tag{12}$$

$$f_{\min}(BER) = \frac{\text{BER}_{min}}{\text{BER}}, \tag{13}$$

and

$$f_{\min}(P) = \frac{P_{min}}{P}, \tag{14}$$

with the exception of $f_{\text{constant}}(W)$, which is computed as:

$$f_{\text{constant}}(W) = \begin{cases} 0, & \text{if } W \leq \text{BW}, \\ -1, & \text{if } W > \text{BW}, \end{cases} \tag{15}$$

where BW is the bandwidth allocated to the communication channel of concern.

In order to reward a chosen action $\bar{u_k} = (R_k, M_k, E_{bk})$, the same assumptions supporting the derivation of Eq. 2 are valid to consider Eq. 4 as

$$s_{k+1} = f_{\text{decision}_k}(\bar{x}_k), \tag{16}$$

computed using the same structure of Eq. 11, where $\bar{x}_k = (R_k, BER_k, P_k, W_k)$ is observed at the receiver, which results in $f_{\text{decision}_k}(\bar{x}_k)$ reflecting the effects of the combined uncertainty imposed by both channel impairments and spacecraft orbital dynamics. Therefore, values for r_{k+1} in Eq. 2 are given by

$$r_{k+1} = \begin{cases} f_{\text{decision}_k}(\bar{x}_k), & \text{if } f_{\text{decision}_k}(\bar{x}_k) > \text{tr}, \\ 0, & \text{otherwise}, \end{cases} \tag{17}$$

where tr is a threshold value that guarantees that only actions that brought the system performance above a certain level should be accounted for in the learning process. Rewards computed that have a value less than the threshold do not represent a successful action choice, given the current channel environment. Thus, it must not be reinforced and should be forgotten instead.

For simulation purposes, Table 2 shows the communication mission phases considered throughout this paper, with a suggestion for their respective weight vectors $\bar{w}$. Each mission phase is composed of some individual objectives, four in this case, demonstrating the multi-objective optimization facet of the proposed RL learning. Each objective is affected somehow by the combination of several impairment sources such as current weather conditions encountered by the line-of-sight (LOS), spacecraft orbital dynamics and on-board electronics status, etc., described in the next section.

Table 2. Communication mission phases, objectives and weights

Objectives / Mission	$w1 : \max(R)$	$w2 : \min(\text{BER})$	$w3 : \min(P)$	$w4 : \text{constant}(W)$
1 - Launch/Re-entry	0.1	0.6	0.2	0.1
2 - Multimedia	0.6	0.3	0.05	0.05
3 - Power saving	0.2	0.1	0.6	0.1
4 - Normal	0.25	0.25	0.25	0.25
5 - Cooperation	0.1	0.2	0.1	0.6

B. Actions

An action u, as mentioned in Section III-A, consists of a set of radio parameters that can be employed by an SDR platform. When choosing an action, the agent might first decide if it should explore new actions in order to expand its knowledge about the current environmental conditions, or if it should exploit already known actions that resulted in bringing the system above the state's threshold tr in the past. It is worth noting that, while following a certain algorithm to choose an action, the decision-making process can be informed by additional knowledge of certain communication channel conditions, reading from external sensors, databases, or predictors such as the Interacting Multiple Model (IMM) proposed by the authors in reference 27. The bias can be inserted directly into the decision function or indirectly as a modification of the main goal of the current mission. The following sections cover the mathematical portion where this term is implemented.

Several methods were proposed in reference 24 to solve the explore–exploit trade-off. They propose the classical ε_k-greedy algorithm with $\varepsilon \in (0, 1)$, which represents the exploration probability with uniform distribution that also picks a uniformly-distributed random action, with exploitation on the remaining portion of time. Since ε_k is function of k, it should decrease as the time goes by in order to reflect the system-learning capability, *i.e.*, explore more during the first steps.

One of the main challenges with ε_k is that it requires manually tuning. The adaptive ε_k, value-difference based exploration (VBDE), was proposed to adaptively change the ε_k value based on a temporal-difference error, making the ε_k value a function of state, *i.e.*, $\varepsilon_k(x_k)$.[28,29] Other common solutions for the exploration versus exploitation problem include the Boltzmann exploration,[24] probability matching,[30] contextual bandit,[31] on-line clustering,[32] etc., each requiring its own functions and parameters, usually a designer's choice. This work considers the classic ε_k-greedy algorithm, and analyzes the multi-objective performance with both fixed and varying ε values.

C. Algorithm

We begin with several general comments regarding the operation of Algorithm 1. During the mission, the range of radio parameters might change due to system upgrades or failures. When this is the case, $\bar{U}$, a matrix containing all the current possible adaptable combinations (for instance, R, M, and E_b) must be updated.

The proposed RL method requires knowledge of the observable $\mathbf{x}_k = (R_k, W_k, P_k, \text{BER}_k)$. Due to the impairments imposed by the current channel conditions, each different action might result in a specific overall performance, measured by $f_{\text{decision}_k}(\mathbf{x}_k)$, that has the same structure as Eq. 11.

While deciding its actions, the agent uses policy h to choose an action u_k, an entry from matrix $\bar{U}$. This policy is detailed in the Section V, and covers how the exploit–explore trade-off was treated. Based on the goals of the current communications mission, which are multi-objective, a vector of weights $\bar{w}$ is chosen. Next, the agent chooses an action, forwards the parameters set to the transmitter, which sends data back using that radio configuration. Finally, the receiver takes measurements, computes the next state and next reward, and updates the Q-matrix using Eq. 2.

Algorithm 1 Multi-objective reinforcement learning for cognitive radio-based satellite communications

Require: Initial parameters $(\bar{R}, \bar{M}, \bar{E}_b, \bar{P}, W, \mathrm{tr}, \varepsilon_k, \alpha_k)$

$\mathbf{U} \leftarrow$ all combinations of $(\bar{R}, \bar{M}, \bar{E}_b, \bar{P})$
$Q_0 \leftarrow 0$
Measure and compute initial state s_0:
    $u_0 \leftarrow \bar{U}$
    Apply u_0 and measure $\mathbf{x}_0 = (R_0, W_0, P_0, \mathrm{BER}_0)$
    Compute $f_{\mathrm{decision}}(\mathbf{x}_0)$
while termination condition is not met, $k = 0, 1, 2, \cdots$ **do**
    if $(\bar{R}, \bar{M}, \bar{E}_b, \bar{P})$ has changed **then**
        $\mathbf{U} \leftarrow$ all combinations of $(\bar{R}, \bar{M}, \bar{E}_b, \bar{P})$
    end if
    $z \leftarrow$ uniform random number $[0, 1]$
    if z$> \varepsilon_k$ **then** ▷ with probability $1 - \varepsilon_k$ (Exploit)
        $u_k \leftarrow u \in \arg\max_{\bar{u}} Q_k(s_k, \bar{u})$
    else ▷ with probability ε_k (Explore)
        $u \leftarrow \mathbf{U}$ randomly chosen action with uniform probability
    end if
    $\bar{w} \leftarrow$ Communications goals
    Apply u_k and measure $\mathbf{x}_k = (R_k, W_k, P_k, \mathrm{BER}_k)$
    Compute next state s_{k+1} and next reward r_{k+1}
    Update $Q_{k+1}(s_k, u_k)$
end while

IV. System Environment

A. Satellite Communications Channel

The method proposed in this work addresses the multi-objective reinforcement learning for satellite-based communications systems implemented via SDR in order to enable cognition. The dynamics of the channel between a satellite and a ground station (GS) play an interesting role in determining the required radio parameters that will allow those multi-objectives to be achieved. Different scenarios can be considered for combinations between the satellite orbit, such as geostationary (GEO), and the GS being fixed or mobile.[33]

Depending on the communication mission phase, additional elements need to be taken into account during the decision making process. For a space-to-ground point-to-point communication link, the channel is quite different from a terrestrial one. Besides the difficulty of knowing the on-board electronics status, orbital geometry elements increase the dynamics of the communications channel. Depending on the GS location, the LOS might cross different sections of Earth's atmosphere, experiencing different weather conditions if either the GS, the satellite, or both, are moving with respect to each other.[4] The scenario becomes more complex if more than one GS, at different locations, are considered with respect to one satellite and their relative motion, and vice versa. Furthermore, the ultimate complexity can be achieved by a scenario in which several GSs and satellites are moving with respect to each other. Thus, each different combination of orbital geometry and weather condition experienced by an LOS represents a whole different channel, and might require a very specific radio parameter set in order to achieve the multi-objective optimal performance by the transmitter–receiver pair.

Channel impairments can be derived from several different sources, including atmospheric effects, space weather, orbital motion, obscuration, and multipath. For Ka-band frequencies (26 GHz) atmosphere effects represent a large challenge concerning satellite communications since losses due to absorption by water vapor, oxygen, rain, snow/ice and clouds[34–36] result in slow and/or fast fading, depending on the elevation angle as well. Also, ionospheric scintillation[37] might be an issue for LOS crossing the ionosphere at low-to-mid latitudes or polar caps, or any other region of the ionosphere after a geomagnetic storm strikes that region of the Earth.[38,39] For the SCaN Testbed, Ka-band communications is performed via GEO relay and typically experience 2 dB of variation during a pass. However, several dB of attenuation can be experienced due to

[4]Varying weather conditions are also experienced for scenarios where both GS and satellite have fixed positions. In this case the channel conditions vary slowly compared to when there is movement associated with the communicating nodes.

multipath caused by structures surrounding the antennas onboard the ISS during clear sky conditions or tens of dB during rainy conditions.

Dynamic orbital elements include the spacecraft's trajectory and/or orbit, relative motion between the spacecraft and its communicating node on the ground, and changing the LOS elevation angle and Doppler frequency shift. Also, positioning of external on-board elements such as solar panels, robotic arms and/or antennas might result in different multipath profiles[40] affecting the reception of the signal on the space segment. Other channel elements exclusive to satellite communications account for the GS surroundings including its terrain profile and material composition.[41] It is assumed that the channel is slow fading, frequency non-selective, and has a constant noise density power.[33,42]

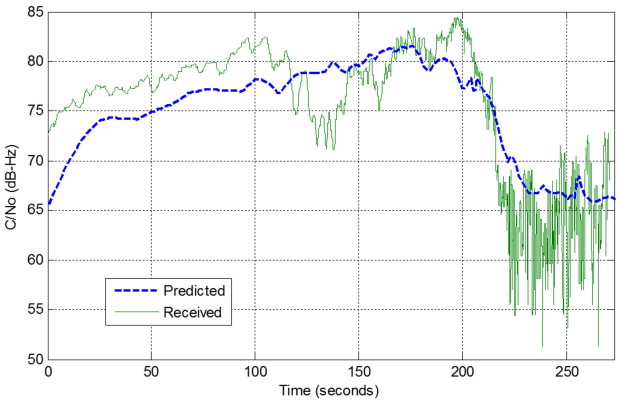

Figure 3. Typical SCaN Testbed direct-to-ground receive power profile.

After four years of operations in low-earth orbit, NASA's SCaN Testbed has characterized both GEO relay link environments as well as direct-to-ground link conditions. Currently, NASA GRC's ground station operates at S-band frequencies providing numerous link dynamics to challenge the communications system. Figure 3 shows a downlink signal strength profile for a typical ISS direct-to-ground pass. Some of the slowly changing variations can be predicted based on the antenna pattern and structural obscurations. However, the more rapidly changing variations, such as those due to multipath, are more difficult to accurately predict.

The combination of these impairments results in a very dynamic channel behavior that affects the multi-objective optimal performance achievement. Thus, an efficient cognition engine might be able to not only optimize the performance for multi-objectives, but also to learn which one achieves the minimum acceptable quality of service (QoS), given the communication mission phase. The following section presents results obtained for the proposed RL approach for different satellite channel conditions, for GEO communicating at Ka-band, with flexible-radio payloads onboard. Clear sky conditions are assumed through an additive white Gaussian noise (AWGN) channel. Rain attenuation conditions are simulated through the synthesis of attenuation time series based on ITU recommendations.[34] The attenuation profile shown in Figure 4 is the time-series used for the simulation considered in this paper for the cases of rain fading. Samples of the resultant channel time series are provided together with their learning performance within the results section.

B. SDR parameters

SDR waveform applications already available on the SCaN Testbed provide dynamic reconfiguration of parameters such as modulation, coding, and transmit power. An increasing number of NASA's STRS-compliant waveforms are being designed to support cognitive radio and cognitive networking applications. Currently, a DVB-S2[43] compliant waveform is being operated on the SCaN Testbed, which gives 27 modulation-coding combinations with which to adjust for link conditions.

The adaptive parameters R, M, and E_b are presented in Table 3. M is comprised of all combinations of the available QAM modulation constellation sizes M_s and encoding rates M_r. The action search space $\bar{U}$ for the experiments in this paper is comprised of all the combinations between all the values available with the ranges shown.

Table 3. SDR parameter ranges

Parameters	Bit rate (R)	Constellation size (M_s)	Encoding rate (M_r)	Tx gain (P)
Values	[128 kbps, 1 Mbps]*	[4, 16, 64]	[4/7, 11/15]	[1, 20]† dB

* Steps of 128 ksps.

† Transmitter gain range achievable depending on combinations of available values for E_b and R.

V. Simulation Results

Any study of RL algorithms involving simulation or experimentation requires the researcher to address the challenge of efficiently exploiting known actions while exploring available actions, looking for better ones. An ideal strategy would consist of long exploitation periods mixed with short exploration periods to identify suitable actions as quick as possible. For comparison reasons, four different scenarios of ε values were simulated. The first is called the "brute force" (BF) scenario, where all actions are evaluated once. In this case, $\varepsilon = 1$ and only exploration is performed, with its duration in time being equal to the total number of possible actions. The second and third scenarios consider fixed values for ε, 0.5 and 0.01, respectively.

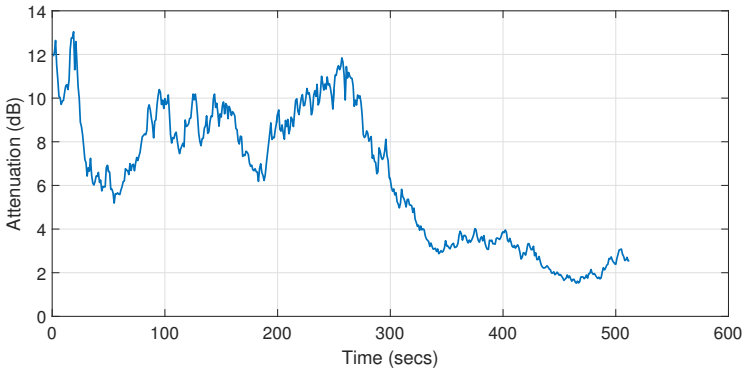

Figure 4. Sample of rain attenuation profile at Ka-band. The time-series was synthesized using ITU recommendations[34] and gives an example of how harmful fading alone can be at this frequency band.

The last scenario considers a varying ε that decreases with the iteration number and is reset when it reaches a value below a threshold, in this case a value of 10^{-4} was assumed. The analysis of the rate at which ε decreases and gets reset is beyond the scope of this paper and it is left as a future work.

Decisions of whether to explore or exploit consist of drawing a random number that is uniformly distributed before the transmission of each packet. The current ε has the role of serving as the threshold, since it represents the exploration probability.

In the case of exploitation, the chosen action is the one associated with the current maximum Q-value, computed by Eq. 2. When exploring, a random action is chosen from a uniform distribution, its corresponding radio parameters are applied at the transmitter, and its performance is measured by the receiver. Regarding the threshold shown in Figure 2, it is assumed zero in order to allow a more detailed study of the algorithm's behavior at any performance level. Also, the learning rate is initialized as 1 for each possible action and decreased as it gets reused, making sure that selecting repeated actions does not provide new knowledge.

Assuming an SNR reading rate of 1 reading/second, each action is used during 1 second. Given an universe of 540 available actions, combinations of the available tunable radio parameters, the BF scenario also lasts 540 seconds. In order to allow a fair comparison, all scenarios also have the same simulation duration as the BF.

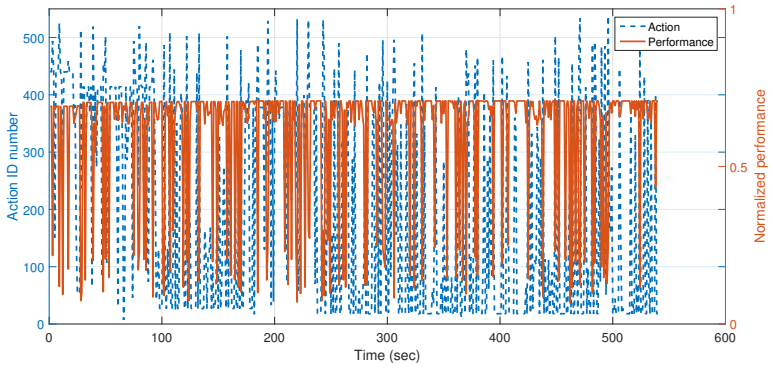

Figure 5. Time-series for chosen action and multi-objective communication performance for $\varepsilon = 0.5$. Knowledge of good performing actions is taken for granted due to the fixed exploration probability.

Examples of time-series for actions chosen and their respective multi-objective performance value normalized are provided in Figure 5 for a fixed ε, and in Figure 6 for varying ε. The communication channel assumed was a direct link from a GEO satellite during clear sky conditions to a fixed ground station as well as rain attenuation at Ka-band.

One can observe that for a fixed ε the system does not take full advantage of the knowledge previously acquired. Thus it keeps exploring with a constant pace even after finding sufficient rewarding actions. On the other hand, the varying ε the knowledge is used wisely with exploration being performed only sporadically.

In order to provide an overview of the multi-objective performance, each mission was simulated $1,000$ times, with a duration of 540 seconds each. The 25th, 50th and 75th percentiles for the mean multi-objective performance of the four scenarios, for each mission, are shown in Tables 4–8. Higher percentile values represent a better performance, since the percentiles themselves refer to multi-objective performance levels, with the 25th percentile considered a baseline for performance evaluation between different ε values. Having large 25th percentile values means that the lowest 25% of the multi-objective performance values are equal or less than that percentile. In addition to performance percentiles, integral values of histograms' areas are also provided.

Histograms for BF and varying ε scenarios, including standard error bars for these, for each mission are provided in Figure 7. For all missions, the effect of using the RL algorithm to manage exploration and exploitation periods while learning which actions were the best ones, results in a higher time spent on those high-rewarding actions when compared to a scenario without this reinforced learning capability, such as the BF.

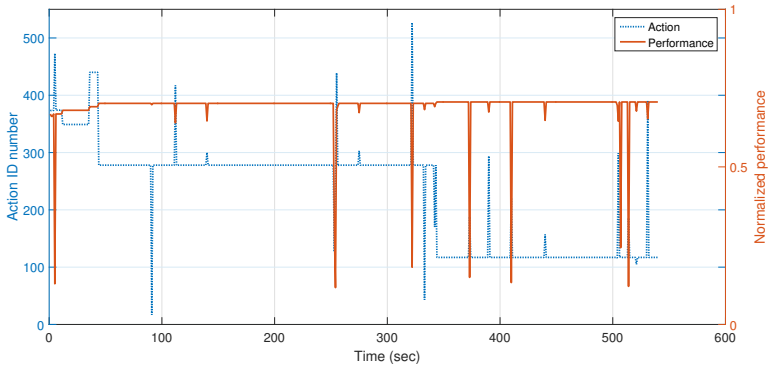

Figure 6. Time-series for chosen action and multi-objective communication performance for varying ε. The best known actions are exploited for a while, and eventually new ones are explored.

Since the multi-objective function weights are different for each mission, each one has a different performance profile. Independent of the profile, based on both results shown in Tables 4–8 and in Figure 7, the RL effect is to focus on the most rewarding actions, located on the rightmost side of the BF histograms, for the whole operation time duration.

The best performances for clear sky conditions are achieved for Missions 1 and 2, since the 25th percentile difference between BF and varying ε scenarios, as well as their respective integral values, are the largest ones. For instance, the improvement achieved for Missions 1 and 2 was sufficiently good that the 25th percentile difference between BF and varying ε scenarios was 0.49 and 0.44, respectively. Also, depending on the mission requirements multiple parameters need to be taken into consideration. Even though Mission 1 had a larger difference value, representing an improvement of more than 3.59 times between these two scenarios, the 25th percentile for Mission 2 was the highest among all the missions simulated.

The rain attenuation cases for all the missions had a worse multi-objective performance when compared to results obtained under clear sky conditions. This is expected given the deep fading values present on the attenuation profile time-series used. Mission 1 seems to be the most one affected by rain attenuation, with a difference of 0.54 in its mean for the BF scenario, and 0.24 difference for its first quartile when compared to results for clear sky condition. Mission 3 seems to be most immune to rain attenuation, showing the smallest difference value for integral, 0.05, followed by Missions 4 and 5.

When comparing percentiles only for scenarios affected by rain attenuation, the 25th percentile of the RL algorithm using the varying ε had an improvement of more than 6.8 times when compared to the system without the RL algorithm, BF scenario, for Mission 1. Mission 5 had an improvement of more than 4.1 times under these same conditions, followed by an improvement of more than 3.7 times for Mission 2, 3 times for Mission 4, and 2.1 for Mission 3.

Table 4. Multi-objective Communications Performance Distribution for Mission 1

Percentile \ Scenarios	BF	$\varepsilon = 0.5$	$\varepsilon = 0.01$	Varying ε	BF (rain)	Varying ε (rain)
25th	0.1917	0.6109	0.5923	0.6892	0.0646	0.4393
50th	0.6476	0.6203	0.6467	0.6937	0.0978	0.4937
75th	0.6769	0.6282	0.6754	0.7095	0.3791	0.6322
Integral	0.4863	0.6224	0.6294	0.7002	0.2355	0.5266

Table 5. Multi-objective Communications Performance Distribution for Mission 2

Percentile \ Scenarios	BF	$\varepsilon = 0.5$	$\varepsilon = 0.01$	Varying ε	BF (rain)	Varying ε (rain)
25th	0.3699	0.708	0.6559	0.8114	0.1835	0.6826
50th	0.5487	0.7151	0.7209	0.8289	0.3952	0.7148
75th	0.6903	0.7205	0.7665	0.8679	0.6024	0.8094
Integral	0.531	0.7235	0.7092	0.8428	0.4055	0.745

Table 6. Multi-objective Communications Performance Distribution for Mission 3

Percentile \ Scenarios	BF	$\varepsilon = 0.5$	$\varepsilon = 0.01$	Varying ε	BF (rain)	Varying ε (rain)
25th	0.1861	0.3824	0.2752	0.3524	0.1348	0.2912
50th	0.2285	0.4222	0.3023	0.3791	0.1789	0.3189
75th	0.2721	0.441	0.3211	0.403	0.225	0.354
Integral	0.2308	0.4094	0.2982	0.3789	0.1897	0.3249

Table 7. Multi-objective Communications Performance Distribution for Mission 4

Percentile \ Scenarios	BF	$\varepsilon = 0.5$	$\varepsilon = 0.01$	Varying ε	BF (rain)	Varying ε (rain)
25th	0.2379	0.4172	0.3996	0.4688	0.1164	0.3573
50th	0.3438	0.4203	0.4309	0.4758	0.1859	0.382
75th	0.4233	0.4236	0.4501	0.4934	0.3046	0.4484
Integral	0.3291	0.4251	0.4247	0.4851	0.2243	0.4023

Table 8. Multi-objective Communications Performance Distribution for Mission 5

Percentile \ Scenarios	BF	$\varepsilon = 0.5$	$\varepsilon = 0.01$	Varying ε	BF (rain)	Varying ε (rain)
25th	0.1139	0.2506	0.2427	0.2877	0.0486	0.2005
50th	0.2375	0.2526	0.2649	0.2904	0.0792	0.2189
75th	0.2693	0.2544	0.2771	0.2979	0.1621	0.2721
Integral	0.2007	0.2613	0.2598	0.2998	0.116	0.2383

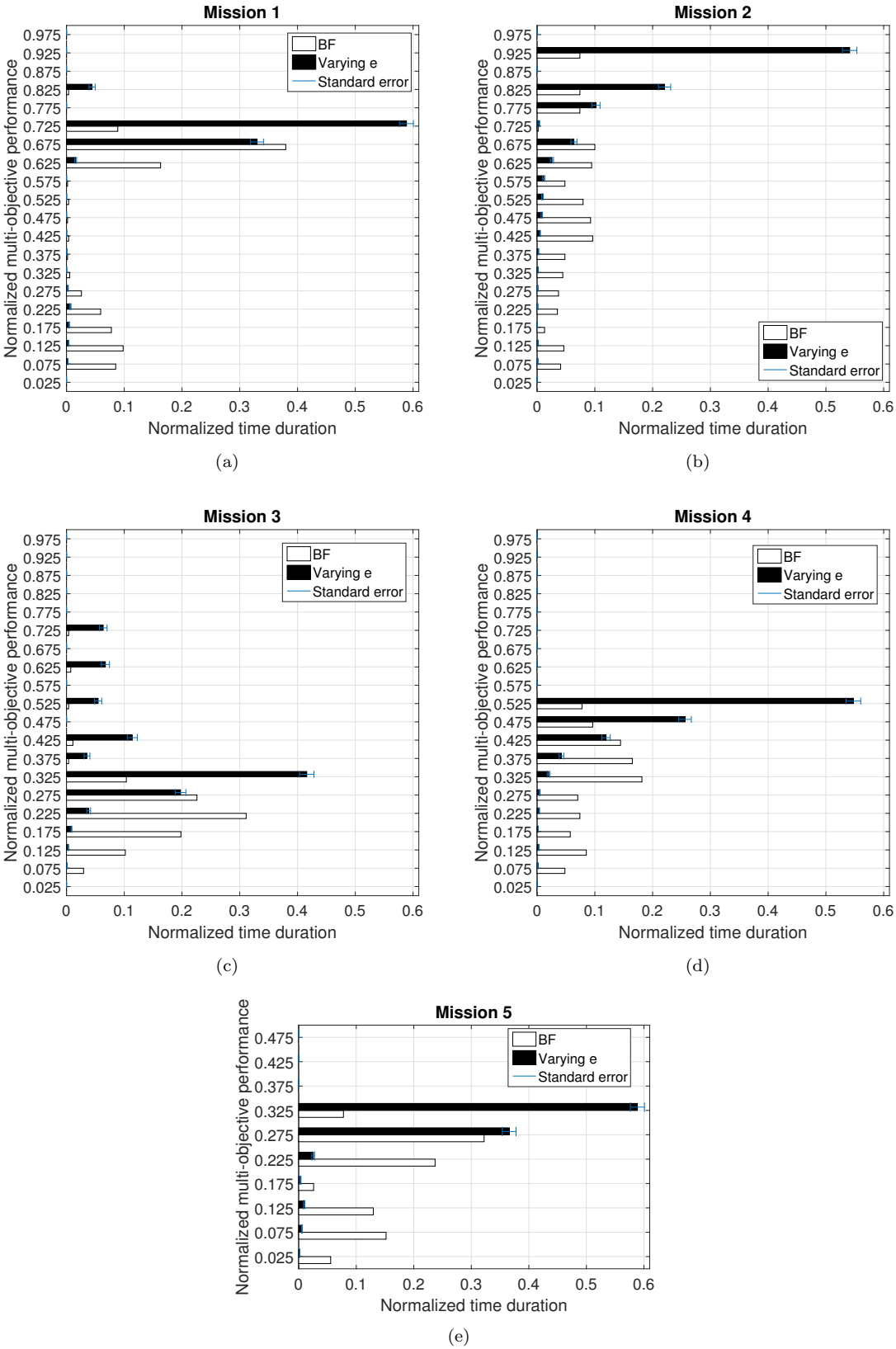

Figure 7. Normalized histograms for average time spent on a certain multi-objective normalized performance level for a GEO satellite-based link under clear sky conditions. For varying ε scenario, in all missions, more time was spent at higher performance levels when compared to the BF scenario.

VI. Application of Cognitive Engines for NASA

Applying cognitive applications to space poses a number of challenges due to the risk-adverse culture of space missions coupled with the resource-constrained environment of space platforms. These two driving conditions will heavily influence where and when CR systems are used within mission spacecraft and throughout the communication system.

Three areas have emerged as candidate application areas for CR systems. The different areas are node-to-node communications, system-wide intelligence, and intelligent internetworking. The first application entails the radio-to-radio link between mission spacecraft and ground terminal (either through relay or direct-to-ground). Cognitive decision making may improve (increase) throughput across a communication link by consuming otherwise unused designed link margin or mitigating impairments. Algorithms that sense performance and understand the entire link capacity could adjust waveform settings to maximize user data and symbol rate by minimizing coding or other overhead. Taking advantage of significant range changes during a ground station pass or operating at reduced data rates at low elevation angles (normally outside the traditional link design) or through weather events offers additional opportunities for additional science data return. Signal recognition among nodes may alleviate missed opportunities due to configuration errors or mitigate unexpected interference.

The second application of CR systems is system-wide intelligence where CR systems make operational decisions normally performed by operators or data-intensive aspects not currently done. For example, CR systems could be applied to relay and ground station scheduling, asset utilization (proper asset loading and accommodating mission priority), optimum link configuration and access times, infrastructure fault monitoring, and failure prediction, among others. Many of the applications will help reduce operation oversight (and cost) and help reduce operational complexity due to the large number of possible configurations. Large data analysis opens a new area to discover performance and operational benefits from all aspects of data collected including: link performance, platform environment (*e.g.* radiation, thermal, and mechanical/vibration), asset availability, and system performance.

Finally, as communications infrastructure becomes more network-based using commercial and international standard protocols, CR systems may benefit the control and data functions of the communications network. Optimizing data throughput according to QoS metrics such as bit error rate, loss packet rate, routing decisions, store-and-forward protocols, and publish-and-subscribe techniques may benefit from cognitive control. Allowing algorithms to learn network behavior, especially small networks with repeatable data flows, may yield throughput and reliability benefits.

One notable aspect regarding CR systems for space is the need for verification or ground testing of all operational conditions before launch. To minimize risk on orbit, missions generally test each mode of operation prior to flight. This helps provide confidence in the on-orbit operation. Having CR systems make unplanned and at times unpredictable changes to flight systems on-orbit will take considerable research and technology demonstrations such as those described in this paper.

VII. Conclusion

In order to support the development of future cognitive satellite-based communication systems, a reinforcement learning-based algorithm design for multi-objective radio communication systems was proposed and had its performance analyzed through simulations conducted for different communication mission goals.

For clear-sky GEO satellite channel, results demonstrated improvements of 3.59 times in the 25th percentile, while under rain conditions the improvement achieved was more than 6.8 times, compared to systems without a learning algorithm under the same weather conditions. The results also demonstrated that, even though the multi-objective performance can be improved by the proposed algorithm, depending on the communication mission requirements certain choices of the exploration probability ε can be made to achieve a specific goal, such as minimum acceptable performance level, or overall performance within a time window period, by imposing conditions on the percentiles and integral values.

For scenarios when rain attenuation was present, all missions showed inferior performance compared to clear-sky conditions, given the deep fading values observed at Ka-band. Simulation results showed that Mission 1 is the most susceptible to have its performance affected in the presence of rain, whereas Mission 3 had the minimum decrease in its multi-objective performance.

Based on the improvements shown in this paper, there is a potential for RL to be further developed

to assist cognitive satellite-based systems dealing with varying communication channels due to dynamics originated by atmospheric and space weather events as well as the dynamics of mechanical orbit conditions imposed to the spacecraft.

Future work is expected to further analyze the behavior of the exploration probabilities while facing dynamic channel conditions for spacecrafts communicating in harsh environments. We also anticipate that on-orbit tests using SCaN testbed are expected to be performed in order to evaluate the improvements achieved by the proposed algorithm under real-world conditions.

Acknowledgments

This work was partially supported by: NASA John H. Glenn Research Center, grant number NNC14AA01A; NASA Space Technology Research Fellowship, grant number NNX15AQ41H; and CAPES Science without Borders scholarship, grant number BEX 18701/12-4.

References

[1]Yucek, T. and Arslan, H., "A survey of Spectrum Sensing Algorithms for Cognitive Radio Applications," *IEEE Communications Surveys and Tutorials*, Vol. 11, No. 1, 2009, pp. 116–130.

[2]Morello, A. and Mignone, V., "DVB-S2: The Second Generation Standard for Satellite Broad-Band Services," *Proceedings of the IEEE*, Vol. 94, No. 1, 2006, pp. 210–227.

[3]Kurose, J. F. and Ross, K. W., *Computer Networking A Top-Down Approach*, Pearson, 2013.

[4]Chen, S., Newman, T. R., Evans, J. B., and Wyglinski, A. M., "Genetic Algorithm-Based Optimization for Cognitive Radio Networks," *IEEE Sarnoff Symposium*, 2010.

[5]Hossain, E., Niyato, D., and Kim, D. I., "Evolution and Future Trends of Research in Cognitive Radio: a contemporary Survey," *Wiley Wireless Communications and Mobile Computing*, Vol. 15, 2013, pp. 1530–1564.

[6]He, A., Bae, K. K., Newman, T. R., Gaeddert, J., Kim, K., Menon, R., Morales-Tirado, L., Neel, J., Zhao, Y., Reed, J. H., and Tranter, W. H., "A Survey of Artificial Intelligence for Cognitive Radios," *IEEE Transactions on Vehicular Technology*, Vol. 59, No. 4, 2010.

[7]Roijers, D. M., Vamplew, P., Whiteson, S., and Dazeley, R., "A Survey of Multi-Objective Sequential Decision-Making," *Journal of Artificial Intelligence Research*, Vol. 48, 2013, pp. 67–113.

[8]Abbas, N., Nasser, Y., and Ahmad, K. E., "Recent Advances on Artificial Intelligence and Learning Techniques in Cognitive Radio Networks," *EURASIP Journal on Wireless Communications and Networking*, 2015.

[9]Bkassiny, M., Li, Y., and Jayaweera, S. K., "A Survey on Machine-Learning Techniques in Cognitive Radios," *IEEE Communications Survey and Tutorials*, Vol. 15, No. 3, 2013, pp. 1136–1159.

[10]Chatzinotas, S., Ottersten, B., and Gaudenzi, R. D., *Cooperative and Cognitive Satellite Systems*, Elsevier, Academic Press, 2015.

[11]Sharma, S. K., Maleki, S., Chatzinotas, S., Grotz, J., and Ottersten, B., "Implementation Issues of Cognitive Radio Techniques for Ka-band (17.7-19.7 GHz) SatComs," *7th Advanced Satellite Multimedia Systems Conference and the 13th Signal Processing for Space Communications Workshop*, 2014.

[12]*SCaN Testbed*, NASA Glenn Research Center. Available at http://spaceflightsystems.grc.nasa.gov/SOPO/SCO/SCaNTestbed/.

[13]Johnson, S. K., Reinhart, R. C., and Kacpura, T. J., "CoNNeCTs Approach for the Development of Three Software-Defined Radios for Space Application," *Proceedings of IEEE Aerospace Conference*, 2012, pp. 1–13.

[14]Reinhart, R., Kacpura, T. J., Johnson, S. K., and Lux, J. P., "NASA's Space Communications and Navigation Testbed Aboard the International Space Station," *IEEE Aerospace and Electronic Systems Magazine*, Vol. 28, No. 4, 2013.

[15]"Space Telecommunications Radio Systems Architecture Standard," Available at https://standards.nasa.gov/standard/nasa/nasa-std-4009.

[16]"Space Telecommunications Radio Systems Architecture Standard Rationale," Available at https://standards.nasa.gov/standard/nasa/nasa-hdbk-4009.

[17]Reinhart, R. C., "Using International Space Station for Cognitive System Research and Technology with Space-Based Reconfigurable Software-Defined Radios," *66th International Astronautical Congress*, 2015.

[18]Berry, D. A. and Fristedt, B., *Bandit Problems: Sequential Allocation of Experiments*, Springer, 1985.

[19]Lai, T. and Robbins, H., "Asymptotically Efficient Adaptive Allocation Rules," *Advances in Applied Mathematics*, Vol. 6, No. 1, 1985.

[20]Motamedi, A. and Bahai, A., "Dynamic Channel Selection for Spectrum Sharing in Unlicensed Bands," *European Transactions on Telecommunications and Related Technologies*, 2007.

[21]Lai, L., Jiang, H., and Poor, H. V., "Medium Access in Cognitive Radio Networks: a Competitive Multi-Armed Bandit Framework," *Proceedings of IEEE Asilomar Conference on Signals, Systems, and Computers*, 2008.

[22]Lai, L., Gamal, H. E., Jiang, H., and Poor, H. V., "Cognitive Medium Access: Exploration, Exploitation and Competition," *IEEE Transactions on Mobile Computing*, Vol. 10, No. 2, 2011.

[23]Busoniu, L., Babuska, R., Schutter, B. D., and Ernst, D., *Reinforcement Learning and Dynamic Programming Using Function Approximators*, CRC Press, 2010.

[24]Barto, A. and Sutton, R. S., *Reinforcement Learning: An Introduction*, MIT Press, 1988.

[25]Proakis, J. G. and Salehi, M., *Digital Communications*, McGraw Hill, 2007.

[26]Cho, K. and Yoon, D., "On the General BER Expression of One- and Two-Dimensional Amplitude Modulations," *IEEE Transactions on Communications*, Vol. 50, No. 7, 2002.

[27]Ferreira, P., Paffenroth, R., and Wyglinski, A. M., "IMM Performance Analysis for Variable Measurement Delay for Satellite Communications at Ka-Band for LMS Channel Under Rain Fading," *Wiley International Journal of Satellite Communications and Networking*, Submitted on April 29 2016.

[28]Tokic, M. and Palm, G., "Value-Difference Based Exploration: Adaptive Control Between Epsilon-Greedy and Softmax," *KI 2011: Advances in Artificial Intelligence: 34th Annual German Conference on AI*, 2011, pp. 335–346.

[29]Tokic, M., "Adaptive E-Greedy Exploration in Reinforcement Learning Based on Value Differences," *KI 2010: Advances in Artificial Intelligence: 33rd Annual German Conference on AI*, 2010, pp. 203–210.

[30]Scott, S. L., "A modern Bayesian Look at the Multi-Armed Bandit," *Applied Stochastic Models in Business and Industry*, Vol. 26, No. 6, 2010, pp. 639–658.

[31]Langford, J. and Zhang, T., "The Epoch-Greedy Algorithm for Multi-Armed Bandits with Side Information," *Advances in Neural Information Processing Systems*, Vol. 20, 2007.

[32]Gentile, C., Li, S., and Zappella, G., "Online Clustering of Bandits," *Proceedings of the 31st International Conference on Machine Learning*, Vol. 20, 2014.

[33]Corazza, G. E., *Digital Satellite Communications*, Springer, 2007.

[34]Union, I. T., "Rec. ITU-R P.618-12 - Propagation Data and Prediction Methods Required for the Design of Earth-Space Telecommunication Systems." Tech. rep., ITU, 2015.

[35]Liu, J., *Spacecraft TTC and Information Transmission Theory and Technologies*, Springer Berlin Heidelberg, 2015.

[36]Richharia, M., *Mobile Satellite Communications Principles and Trends*, Wiley, 2014.

[37]Union, I. T., "Rec. ITU-R P.531-12 - Ionospheric Propagation Data and Prediction Methods Required for the Design of Satellite Services and Systems," Tech. rep., ITU, 2013.

[38]Huba, J. and Joyce, G., "Global Modeling of Equatorial Plasma Bubbles," *Geophysical Research Letters*, 2010.

[39]"Sun unleashes 1st monster solar flare of 2015," Available at http://www.space.com/28797-sun-unleashes-monster-solarflare-x2.html.

[40]Fehse, W., *Automated Rendezvous and Docking of Spacecraft*, Cambridge University Press, 2003.

[41]Ferreira, P. and Wyglinski, A. M., "Performance Analysis of UHF Mobile Satellite Communication System Experiencing Ionospheric Scintillation and Terrestrial Multipath," *82nd IEEE Vehicular Technology Conference*, 2013.

[42]Patzold, M., *Mobile Radio Channels*, John Wiley and Sons, 2012.

[43]"DVB-S2 Standard," Available at https://www.dvb.org/standards/dvb-s2.