Cognitive Engine Design for Link Adaptation: An Application to Multi-Antenna Systems

Haris I. Volos, *Graduate Student Member, IEEE,* and R. Michael Buehrer, *Senior Member, IEEE*

***Abstract*—In this paper, we present a Cognitive Engine (CE) design for link adaptation and apply it to a system which can adapt its use of multiple antennas in addition to modulation and coding. Our design moves forward the state of the art in several ways while having a simple structure. Specifically, the CE only needs to observe the number of successes and failures associated with each set of channel conditions and communication method. From these two numbers, the CE can derive all of its functionality. First, it can estimate confidence intervals of the packet success rate (PSR) using the Beta distribution. A low computational approximation to the CDF of the Beta distribution is also presented. Second, the designed CE balances the tradeoff between learning and short-term performance (exploration *vs.* exploitation) by applying the Gittins index. Third, the CE learns the radio abilities independently of the operation objectives. Thus, if an objective changes, information regarding the radio's abilities is not lost. Finally, prior knowledge such as capacity, BER curves, and basic communication principles are used to both initialize the CE's knowledge and maximize the learning rate across different channel conditions. The proposed CE is demonstrated to have the ability to learn in a dynamic scenario and quickly approach maximal performance.**

***Index Terms*—Cognitive Engine, cognitive radio, learning, optimization, link adaptation, Gittins index, Bayes' Rule, beta distribution.**

I. Introduction

A radio has traditionally used a fixed set of communication methods selected by its operator. Today, however, a radio is expected not only to use a large number of the plethora of current communication methods, but also to be able to select the method that best meets its objective under the current operating environment.

Typically, the radio designer will analyze each communication method in terms of the desired objective under assumed channel models. Then, the designer will use the analysis to arrive at a set of adaptation rules that best meets the objective for the radio. Performing the required analysis is usually a lengthy task requiring a significant amount of effort. The latter is especially true when a large number of different methods are involved. Furthermore, if the channel models do not hold or the communication methods perform in a way not expected, the design becomes irrelevant. The same applies if the desired objective changes.

The question then arises: What if a radio could be designed in such a way that it could determine on its own the best communication method to use to meet its objectives? The scope of our work[1]–[3] is to propose methods that can make such a design possible. This work is based on the pioneering work of Mitola, who first described the "cognitive radio" (CR). Mitola's ideal CR is not only capable of optimally utilizing its own wireless capabilities, but also of self-determining its goals by observing its human operator's behavior [4]. In this work, we are only interested in optimizing the wireless capabilities of the radio by observing the environment and its own performance.

A. Previous work

In the last few years communication engineers have borrowed ideas from the fields of machine learning and Artificial Intelligence (AI) in an effort to make CR possible [5]. The agent that implements the methods that enable the radio to have the desired functionality is referred to as a Cognitive Engine (CE). This paragraph provides a brief overview of the state-of the art in CE design related work. The works of Rieser, Rondeau, and Le [6]–[8] attempt to approach Mitola's vision by proposing a CE that deals with the user, policy and radio domains. Their designs are similar and based on the Genetic Algorithm (GA), Case Based Reasoning (CBR), and multi-objective optimization principles. The work of [9] designed a CBR based CE for IEEE 802.22 WRAN applications also looking into both the radio and policy domains. Other works are more focused only on the radio domain. The work in [10] and [11] applied a GA and particle swarm optimization respectively to multi-channel links. Finally, [12] and [13] applied an Artificial Neural Network (ANN) and predicate logic respectively for learning and optimization of a wireless link.

CE design is only one aspect of CR; significant work has been done in other CR areas, mostly related to Dynamic Spectrum Access (DSA). For example, fundamental CR studies have been done on achievable rates [14], limits [15], fundamental issues [16], and design tradeoffs [17]. References [18] and [19] derive theoretical capacities for MIMO CR systems and [20] proposes some techniques for operating in a MIMO spectrum sharing setting. Work in other CR aspects includes spectrum sensing [21], [22], cognitive networks [23], [24], security [25], and minimizing system power consumption [26].

B. Research Challenges

The field of CR was advanced by the contributions of the aforementioned and other works. However, there are still several challenges that are not sufficiently addressed which we examine in this work. First, predictable performance might be critical for some applications. For example, if the radio

Manuscript received November 5, 2009; revised April 15, 2010; accepted June 15, 2010. The associate editor coordinating the review of this paper and approving it for publication was S. Affes.

The authors are with Wireless at Virginia Tech, Blacksburg, VA 24061, USA (e-mail: {hvolos, buehrer}@vt.edu).

This work was supported by the National Science Foundation under Grant No. 0520418

Digital Object Identifier 10.1109/TWC.2010.070910.091651

is part of a network, performance indicators can be used to select a configuration that meets the network's needs. Unfortunately, none of the current works present a direct way to provide confidence intervals of the radio's expected performance. In this work, we show how the observations are ideally translated into confidence intervals, and we also provide an approximation to the ideal, albeit computationally demanding, method.

Second, once the CE learns a few methods that meet its objective, there arises a question of whether it should continue seeking better methods or should it use what it already knows? This is the classic problem of exploration *vs.* exploitation. *Exploration* refers to trying options with unknown but potentially beneficial outcomes. On the other hand, *exploitation* refers to using what is already known to have the highest performance metric. Lai *et al.* [27] addressed this problem in the context of finding unused channels for DSA applications, but we are not aware of any work addressing this problem in the context of link adaptation. In the current work, we apply the Gittins index [28] to optimally balance exploration *vs.* exploitation.

Third, most works (for example, [7], [9], [10]) propose techniques that learn what serves the radio's current objective. Should the objective change, the capabilities of the radio are only taken into account indirectly by analyzing past cases. We believe that both learning and optimization can be significantly improved by learning the capabilities of the radio independently of the current objective. The proposed approach allows this to occur.

Fourth, we would like the CE to utilize fundamental communication principles and prior information such as Bit Error Rate (BER) and capacity curves to improve the learning rate. For example, [7], [10] use BER curves to initialize their GA. In this work, we generalize this idea by showing how any prior knowledge can be converted to prior observations and how to propagate observations from the radio's operation across channel conditions and configuration options.

Finally, although in principle all the CE's mentioned can work with multi-antenna systems, none is directly applied to such systems. Multi-antenna systems pose additional challenges to the CE because of the expanded set of possible configurations. Thus, this work includes multi-antenna configurations as a base capability of the radio.

We believe it is important to have a CE that utilizes prior information, past experiences and wireless communication principles to perform efficient joint learning and optimization of a radio's resources. This work provides a design that meets these requirements based on classic statistical learning and communication concepts.

C. Paper Organization

This paper is organized as follows: Section II provides an overview of the proposed CE. Section III describes the methods used for confidence interval estimation. Section IV demonstrates how we can achieve a balance in the tradeoff between exploration and exploitation. In Section V, we explain how prior knowledge is translated to observations in our system. Section VI presents results from an operation example demonstrating the abilities of the CE, and Section VII provides some concluding remarks.

II. Overview of the Proposed Cognitive Engine

A CE is an intelligent *agent* that enables the radio to have the desired learning and adaptation abilities. This AI agent[29] *senses* its *environment* (the wireless channel), *acts* by using a communication method based on its past experience, and *observes* its own performance to learn its capabilities.

The CE is designed with respect to the Packet Success Rate (PSR) of the transmitted data packets. The PSR is easier to observe than BER and its usage simplifies the design process. Using the PSR allows our design to be centered around only two numbers: the successes α and failures β at each set of channel conditions and configuration pair. In the proposed CE, any knowledge is translated to $\{\alpha, \beta\}$ pairs from which decisions are made by using statistical inference. The $\{\alpha, \beta\}$ pairs, for a discrete set of channel conditions, are stored in an observations database[1]. The channel conditions are properly discretized to provide a reasonable resolution of the channel metrics' range. The observations database indexes the CE's observations for each channel condition and communications method set. The size of the observations database is generally fixed, unless more advanced methods are used.

Figure 1 provides an overview of the CE's operation. First, any available prior information, such as BER and capacity curves, is converted to $\{\alpha, \beta\}$ pairs (see Section V) and added to the observations database. This operation is typically done *once* per communication method in the CE's lifetime. Therefore, computational complexity is not an issue for this step.

Second, the channel is observed[2] and the decision section determines the configuration to be used based on the current channel conditions, the radio's objective, and the current contents of the observations database. If the CE only considers ensuring short-term performance (exploitation), Bayes' rule is applied on the $\{\alpha, \beta\}$ pairs to estimate confidence intervals of the PSR. The latter are used to calculate a performance metric and the method that has the highest metric is used. On the other hand, if optimal exploration *vs.* exploitation balance is desired, the Gittin's index is estimated by using the $\{\alpha, \beta\}$ pairs. The best exploration *vs.* exploitation balance is achieved when the method with the highest Gittin's index is used. All the aforementioned operations are discussed in detail in the sections to follow.

Third, the selected communication method is used to transmit data packets over the radio channel. Finally, at the receiver the channel metrics and performance data, $\{\alpha, \beta\}$, are estimated, fed back to the transmitter, and added to the observations database. The latter is updated by adding the new observations to the respective entries; then the process is repeated from the second step.

III. Performance Estimation

Deciding which communication method to use, given the channel conditions, requires an estimate of the expected performance. This section provides a method for bounding

[1]We will discuss the observations database in more detail in Section V-D.

[2]Channel observation is assumed to be done either via feedback or if the link is time-division duplexed, by direct observation. The impact of imperfect observations is the subject of current work.

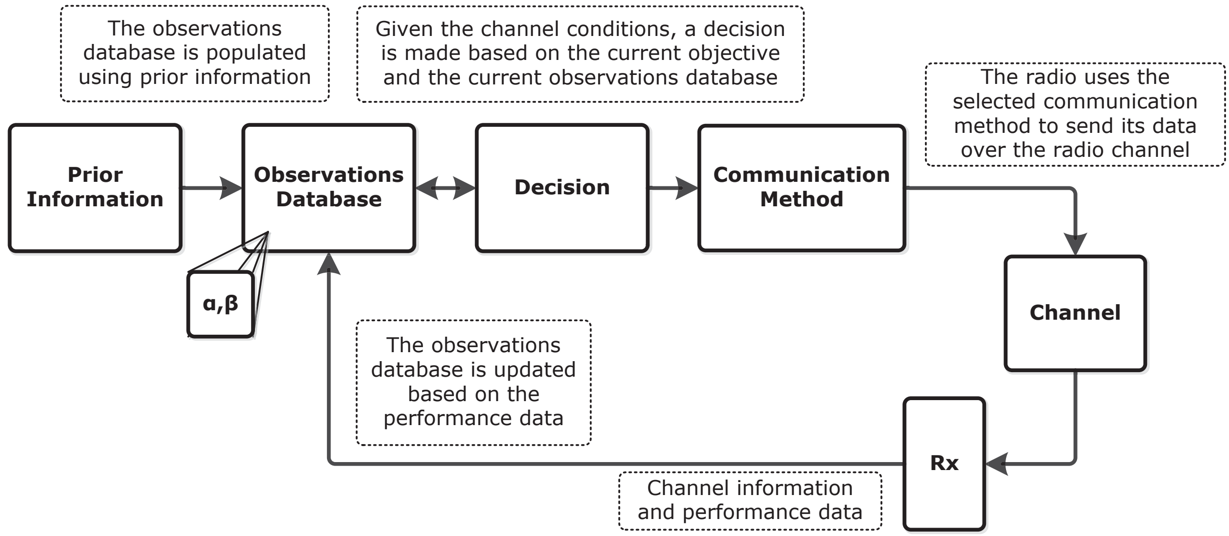

Fig. 1. Cognitive Engine operation diagram.

performance given past experiences for the same (quantized) channel conditions. The CE typically uses this information when it is focused on short-term performance, *i.e.*, using the best performing method (exploitation).

To facilitate exploitation, we adopt a confidence interval method based on Bayes' rule. This method estimates a confidence interval for the PSR of a communication method based on the prior observations of successes and failures at certain channel conditions. However, the proposed method for estimating the confidence intervals can be computationally demanding. Therefore, we also propose to use the significantly less demanding normal approximation of the original method.

A. Bayes' Rule and Confidence Intervals

Bayes' Rule simply states the following [30]:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

which means that our posterior belief that an event is going to happen is proportional to the prior belief that it would occur times the likelihood of a particular observation related to the event. Generally, the prior is what we believe before any observations and the likelihood is based on observations associated with the occurrence of the event. Assuming that we want to estimate the posterior density of θ, after observing the parameter vector $\mathbf{y}$, Bayes' Rule states:

$$p(\theta|\mathbf{y}) \propto p(\mathbf{y}|\theta)p(\theta) \tag{1}$$

where $p(\mathbf{y}|\theta)$ is the likelihood of $\mathbf{y}$ as a function of θ, and $p(\theta)$ is a prior density that reflects our beliefs about θ. In the context of our application, θ is the PSR and $\mathbf{y}$ represents observed successes and failures for a specific set of channel conditions.

We assume that the process of sending packets over a radio channel is Bernoulli. Specifically, each packet is a Bernoulli trial that has two outcomes: a successful reception of the packet or an unsuccessful reception of the packet. A successful reception is defined as the case where the packet is either error-free or has a correctable number of errors. In $n = \alpha + \beta$ Bernoulli trials let α be the number of successful trials, β be the number of unsuccessful trials, and θ be the probability of a successful trial. We are going to apply Bayes' Rule to estimate the posterior density $p(\theta|\alpha)$. The likelihood of α, given θ, after n observations, is given by the binomial distribution:

$$p(\alpha|\theta) = \binom{n}{\alpha}\theta^{\alpha}(1-\theta)^{n-\alpha} \quad (\alpha = 0, 1, \ldots, n) \tag{2}$$

Initially let's assume that we have no prior information, *i.e.*, every PSR is equally likely:

$$p(\theta) = 1 \quad (0 \leq \theta \leq 1) \tag{3}$$

By substituting (2) and (3) into (1) the posterior density is found to be proportional to [30]:

$$p(\theta|\alpha) \propto p(\alpha|\theta)p(\theta) \propto \binom{n}{\alpha}\theta^{\alpha}(1-\theta)^{n-\alpha} \propto \theta^{\alpha}(1-\theta)^{n-\alpha} \tag{4}$$

The normalization constant required for $p(\theta|\alpha)$ to be a proper density function is $1/B(\alpha+1, n-\alpha+1)$, *i.e.*,

$$p(\theta|\alpha) = \frac{1}{B(\alpha+1, n-\alpha+1)}\theta^{\alpha}(1-\theta)^{n-\alpha} \tag{5}$$

where $B(j,k)$, is the Beta function given by:

$$B(j,k) = \int_0^1 u^{j-1}(1-u)^{k-1}du \tag{6}$$

where j and k are parameters of the Beta function.

Comparing (5) with the Beta distribution:

$$Be(\theta; j,k) = \frac{1}{B(j,k)}\theta^{j-1}(1-\theta)^{k-1} \tag{7}$$

it follows that θ given α is Beta distributed:

$$p(\theta|\alpha) = Be(\theta; \alpha+1, n-\alpha+1) \tag{8}$$

The result above assumes no prior information. Now, let us proceed by deriving the result accounting for prior information. Let this prior information be α_0 and β_0 successes and failures respectively. In this case, the prior distribution of θ is taken to be $Be(\theta;\alpha_0+1,\beta_0+1)$[31]:

$$p(\theta) \propto \theta^{\alpha_0}(1-\theta)^{\beta_0} \tag{9}$$

Note: If $\alpha_0=\beta_0=0$ (*i.e.,* we have no prior information) the above formulation reverts to the uniform prior distribution.

Similar to the uniform prior case, combining (9) with (2) and the result of (8), the posterior density of θ given α can be also shown to be Beta distributed:

$$p(\theta|\alpha) \propto \theta^{\alpha+\alpha_0}(1-\theta)^{n-\alpha+\beta_0} \tag{10}$$

$$p(\theta|\alpha) = Be(\theta;\alpha+\alpha_0+1, n-\alpha+\beta_0+1) \tag{11}$$

A property of the Beta distribution that is being utilized here is that it can be used as a *conjugate prior*. When a conjugate prior distribution is combined with the likelihood, the resulting posterior distribution belongs to the same family as the prior distribution. Therefore, a Beta prior will give a Beta posterior.

The CDF of the Beta distribution is given by the regularized incomplete Beta function:

$$I_x(\alpha,\beta) = \int_0^x \frac{1}{B(\alpha,\beta)} u^{\alpha-1}(1-u)^{\beta-1}du \tag{12}$$

Using the CDF, confidence intervals for the parameter θ can be estimated. Specifically, we seek an interval $\{\theta_l,\theta_u\}$ such that with probability $1-\delta$ the parameter[3] $\theta \in (\theta_l,\theta_u)$. The lower and upper bound θ_l and θ_u respectively are given by finding the values of θ_l and θ_u that satisfy the following equations:

$$I_{\theta_l}(\alpha_0+\alpha,\beta_0+\beta) = \frac{\delta}{2} \tag{13}$$

and

$$I_{\theta_u}(\alpha_0+\alpha,\beta_0+\beta) = 1-\frac{\delta}{2} \tag{14}$$

There are many ways that the CE can use θ_l and θ_u to serve its operational goals. For example, if the CE wants to exploit (rather than explore), it will use the best method with $\theta_l \geq$ the target PSR. In such a case, the target PSR will be met or exceeded with a probability of $1-\delta/2$. A low value of δ can be used when predictable performance is required and a higher δ can be used to allow for a more opportunistic behavior. On the other hand, when the CE is exploring, methods with a $\theta_u <$ the target PSR need not be explored. This is because there is only a $\delta/2$ probability that those methods can meet the target PSR.

Solving (13) and (14) is a computationally intensive task which might be prohibitive in some applications. However, a normal approximation to the Beta distribution can be used as an alternative.

[3]The interval chosen its the most common found in classic statistics literature. Other intervals are possible, and if there is a specific need by a certain application, a different interval can be chosen.

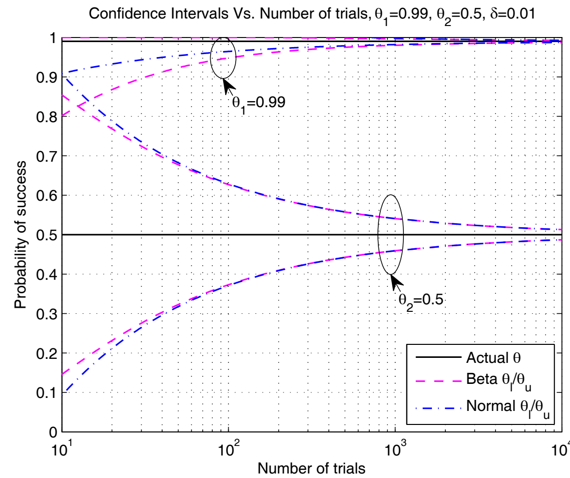

Fig. 2. Confidence interval estimation comparison.

B. Normal approximation

The confidence intervals of θ can also be estimated by approximating the Beta distribution using a normal distribution with a mean $\hat{\theta}=\frac{\alpha}{n}$ and variance $\sigma^2=\frac{\hat{\theta}(1-\hat{\theta})}{n}$. The resulting confidence interval is given by [32]:

$$\hat{\theta} - z_{\delta/2}\sigma < \theta < \hat{\theta} + z_{\delta/2}\sigma \tag{15}$$

where $z_{\delta/2}$ is the $1-\delta/2$ percentile of the standard normal distribution. This approximation is valid when $1<\alpha<n$ and when both $n\theta$ and $n(1-\theta)$ are greater than five. Figure 2 compares the approximation of the 99% confidence intervals estimated using the normal approximation for two key θ values: 0.5 and 0.99. In the former case, the normal approximation was closer to the actual bounds in comparison to the latter case. Furthermore, the normal approximation was found to be very close to the actual bounds after a few hundred trials. Based on these results, one can use the actual bounds given by the Beta distribution for the first hundred samples, and then use the normal approximation.

IV. Exploration Vs. Exploitation

A primary function of the CE is to learn the capabilities of the radio. This is generally done by trial and error. When the radio performs this function it is said to be *exploring*. On the other hand, when the radio is choosing methods with the best known performance it is said to be *exploiting*. However, the radio might not have the latter option if it does not know any satisfactory methods (*i.e.*, it has not learned its performance for the situation that it is currently facing) and will have no choice but to explore. This operation consumes valuable resources such as time and energy and might significantly impact the link performance (*i.e.*, dropped packets). One option to avoid these negative effects during the radio's operation is to put the CE through prolonged training sessions covering most expected operating conditions. However, even if the CE is assumed to go through prolonged learning (exploration) sessions, it is practically impossible to expose it to all possible channel conditions *a priori*. Consequently, it is reasonable to expect that the CE sooner or later will face unknown conditions. In such a case, if the radio is operating in a critical mission it may not have the luxury of time to learn what is best before operating; it has to establish a connection and learn at the same time. Optimally balancing exploration *vs.* exploitation ensures that the negative effects of learning will be kept to a minimum. Therefore, we need to evaluate the performance of the radio controlled by the CE during learning.

The goal of this section is to investigate and apply exploration *vs.* exploitation balancing techniques. Namely, we examine the simple, yet effective, ϵ-greedy strategy and the more complex, but optimal Gittins [28] index method.

A. *Background and Problem Formulation*

We have K communication methods. For each method k, we have a belief state $\pi_k(n)$ which represents our knowledge about the underlying reward distribution at a time step n. $\boldsymbol{\pi}(n)$ is a vector of all K belief states at time step n: $\boldsymbol{\pi}(n) = [\pi_1(n), \pi_2(n), \ldots, \pi_K(n)]^T$. The belief state is the number of observed successes α and failures β for a Bernoulli reward process, and for a Normal reward process the belief state is $(\bar{\mu}_k(n), \bar{\sigma}^2_{\mu_k}(n), n')$ the estimates of the mean μ_k and the variance of the mean σ_{μ_k}, using n' samples, of the underlying reward process.

The reward in our problem is the number of bits/s/Hz that were successfully received. If we use method k, at a time step n, we receive a reward $R_k(n)$. Our belief about the reward distribution changes from $\pi_k(n)$ to $\pi_k(n+1)$, as given by a state transition probability $p_k(\pi_k(n+1)|\pi_k(n), R_k(n))$. A method that was not used at time step n remains in the same belief state, *i.e.*, $\pi_k(n+1) = \pi_k(n)$. Given the belief state vector $\boldsymbol{\pi}(n)$ we want to select a method that will maximize the expected reward. One might suggest to use the following strategy:

$$k = \underset{k\in[1,K]}{\arg\max}\, E\{\, R_k(n)|\, \pi_k(n)\} \tag{16}$$

however, this strategy is myopic *i.e.*, it will just use the method with the maximum expected reward and ignore the confidence in our beliefs about the expected reward. This confidence is generally derived from the number of observations that represents each belief state $\pi_k(n)$. This is especially true when we start from a blank or a partial state of knowledge. Let $P_s\,(k = x|\boldsymbol{\pi}(n))$ be the probability of a strategy s choosing the method x given the belief state vector $\boldsymbol{\pi}(n)$. Moreover, let $f(R_k(n)|\pi_k(n))$ be the probability density function (PDF) of receiving a reward $R_k(n)$ given the belief state $\pi_k(n)$. The expected reward, using strategy s, at time step n is given by:

$$\begin{aligned} E\left\{R(n)|\boldsymbol{\pi}(n)\right\} \;=\; &\sum_{x=1}^{K}\left(P_s\left(k=x|\,\boldsymbol{\pi}(n)\right)\int_0^{R_x^{mx}} R_x(n)\cdot\right. \\ &\left. f\left(R_x(n)|\,\pi_x(n)\right)dR_x(n)\right) \end{aligned} \tag{17}$$

where R_x^{mx} is the maximum possible reward given by method x.

Our goal is to maximize the expected total discounted reward $R(1)|\boldsymbol{\pi}(1) + \gamma R(2)|\boldsymbol{\pi}(2) + \gamma^2 R(3)|\boldsymbol{\pi}(3) + \cdots$ where γ is a discount factor. In other words, given an initial belief state vector $\boldsymbol{\pi}_0$ we wish to maximize the expected total discounted reward $V(\boldsymbol{\pi}_0)$

$$V\left(\boldsymbol{\pi}_0\right) = E\left\{\left.\sum_{n=0}^{\infty}\gamma^n R\left(n\right)\right|\boldsymbol{\pi}(1) = \boldsymbol{\pi}_0\right\} \tag{18}$$

To find the strategy that maximizes (18) is a K-dimensional problem. At each time step a method k needs to be selected. To accomplish this task the following need to be considered: all the possible reward values, all the possible belief state transitions, and all possible selections for the subsequent steps until the end. This problem becomes computationally intractable even for small K. Two methods that simplify the process are the Gittins index and the ϵ-greedy strategy. The latter may not necessarily maximize return in a finite time period, however, it is considered because of its simplicity, and because it is generally found to provide acceptable solutions for some problems.

B. *The Gittins Index Strategy*

Gittins and Jones [28] proved that our K-dimensional problem can be reduced to K *one*-dimensional problems by using a dynamic allocation index based strategy. Specifically, the optimal strategy that maximizes (18) is simply selecting the method k that has the highest index ν_k at each time step[33]:

$$\nu_k\left(\pi_0\right) = \sup_{N>0}\frac{E\left\{\left.\sum_{n=0}^{N-1}\gamma^n R_k\left(n\right)\right|\pi_k(1) = \pi_0\right\}}{E\left\{\left.\sum_{n=0}^{N-1}\gamma^n\right|\pi_k(1) = \pi_0\right\}} \tag{19}$$

which is the expected total discounted reward normalized by the expected total discount time up to the stopping time N. The stopping time N may be different for each method k and depends on the initial belief state π_0. For example, after a few trials using a promising method k, the newly acquired knowledge may suggest that it is not worthwhile to continue using method k and the process needs to stop. This ratio is taken over all possible stopping times N and $\nu_k\left(\pi_0\right)$ is the maximum value. The optimal strategy is simply to use the option k with the highest ν_k. This method has been proven to be optimal by Gittins [28], [33] and others [34].

The Gittins index is dependent upon the underlying distribution of R_k. In this work, we consider the Gittins index for the Normal Reward Process (NRP) and the Bernoulli Reward Process (BRP). It may be noted that in the application examined in this work the underlying process is Bernoulli - either a packet is successful or it is not. In our application, if a transmitted packet is successfully received, then we assume a return equal to the rate of the communication option used, otherwise the return is zero. Therefore, a BRP is in theory more suitable than a NRP. However, as will be shown in the results, assuming a NRP has some practical performance advantages over assuming a BRP.

For a NRP the Gittins index is equal to:

$$\nu(\bar{\mu}, \bar{\sigma}^2, n', \gamma) \equiv \bar{\mu} + \bar{\sigma}\nu(0, 1, n', \gamma) \tag{20}$$

where $\nu(0, 1, n', \gamma)$ is the Gittins index for a zero mean, unit variance distributed process.

For a BRP the Gittins index is equal to [35]:

$$\nu(\alpha,\beta,\gamma,R_k) \equiv R_k \nu(\alpha,\beta,\gamma,1) \quad (21)$$

where $\nu(\alpha,\beta,\gamma,1)$ is the Gittins index for a Bernoulli process with α successes and β failures, with a reward of 1, if successful. R_k is the reward received when option k is successful. In the BRP case, the belief state is represented by $\{\alpha,\beta\}$.

Although easier than solving the original problem, calculating the Gittins indices is still not a trivial task. For the interested reader, [35] provides a concise description of the method for calculating these indices. However, for most practical purposes the indices tabulated in [33] are sufficient, and thus we use a table-lookup approach.

C. The ϵ-greedy Strategy:

The ϵ-greedy strategy[36] is a simple strategy that uses (*i.e.*, exploits) the best method ($k_{greedy} = \arg\max_k \bar{\mu}_k(n)$) 1-$\epsilon$ ($\epsilon \in [0,1]$) of the time (greedy). However, with probability ϵ it explores by using a random method,k, uniformly selected. As $n \to \infty$ by the law of large numbers $\bar{\mu}_k(n)$ is going to converge to the true mean. The ϵ-greedy methods guarantee that all the options are explored as the horizon tends to infinity. The ϵ parameter controls how fast exploration is performed. A higher ϵ will cause a faster exploration and arrive more quickly at an optimal or near-optimal option. However, the high exploration rate may cause reduced overall returns because of the higher exploration cost.

In our case, we do have some prior information about the communication system that should be used: we know the maximum potential return of each option (capacity) and we also know the upper bound of the capacity that can be achieved under the current channel. Therefore, we restrict the exploration to machines that potentially can outperform the current k_{greedy}. We also prohibit exploration of methods whose rate exceeds the current channel capacity as given in Section V-B. The latter restriction was also applied to the Gittins index method.

D. Evaluation Description

We evaluated the proposed methods by implementing the system using the MATLAB simulation software package.

Configuration: Three MIMO Techniques were considered: Beamforming, Transmit Diversity using a space-time block code (STBC), and Spatial Multiplexing using V-BLAST. The three options had rate multipliers of 1, 3/4, and 4 respectively. The MIMO system was assumed to have four transmit and four receive antennas (4×4). The CE had the choice of the following modulation schemes: QPSK, 8-PSK, 16, 32, 64, 128, and 256 QAM. Furthermore, it could vary the coding rate of a convolutional codec with a constraint length $K = 8$ [37]. The available coding rates were: 1, 7/8, 3/4, 2/3, 1/2, 1/4, 1/6, and 1/8. Not all modulation/coding combinations were allowed. The allowable modulation/coding combinations were chosen such that the combined distance metric [37] and spectral efficiency monotonically decreased and increased respectively. The final combinations had 22 different spectral efficiencies from 0.25 bits/s/Hz (QPSK, $1/8$ codec) to 8 bits/s/Hz (256-QAM, uncoded). *Methods Tested:* We tested the Gittins index for both NRP and BRP for γ equal to[4] 0.5, 0.7, and 0.99. The $\epsilon-$greedy strategy was evaluated with ϵ equal to 0.01, and 0.1. Values around 0.1 are commonly used.

The methods were tested in an SNR range between 5 and 50 dB at 5 dB intervals at a maximum pairwise antenna correlation $\rho = 0.1$. We also tested the case of $\rho = 0.5$ for a few values of γ and ϵ. Specifically, for $\rho = 0.5$, we only considered $\gamma = 0.7$ and $\epsilon = 0.1$. A high correlation, ρ, between the antenna elements negatively affects the use of spatial multiplexing. The reader is reminded that spatial multiplexing exploits the availability of multiple spatial modes, which are reduced to one as $\rho \to 1$.

It may be noted that the different channel metrics such as SNR and ρ represent different sets of working options. Higher SNR levels have more options available and a low ρ value allows the use of more spatial multiplexing options. The availability of working options will affect the total return as options that do not work will adversely affect the return, if they are tried.

Evaluation Metrics: In the results to follow, we examine the average total return as a fraction of the average optimal total return. We also examine the average instantaneous return as a fraction of the average optimal instantaneous return. The total return is the sum of all instantaneous returns after a number of trials. Since the return depends on the specific channel and noise realizations, for each channel realization we averaged the performance over multiple noise realizations. Further, we averaged over multiple channel realizations for a given channel condition (i.e., average SNR and channel correlation). The average instantaneous return is the average (over noise and channel realizations) return experienced after a specific number of trials. The optimal return for each case is the best possible return averaged over noise and specific channel conditions. The optimal returns were estimated by employing a brute force search over the mean return of all the available options. Given the SNR and ρ, the mean return was estimated by using 400 different channel realizations for *each option*.

E. Results

Table I presents the average total return as a fraction of the optimal total return (*i.e.*, the total return achieved by an omniscient agent). The average total return was calculated by taking the average of the total return for all the SNR levels tested. This applies for all the methods and parameters tested. There are two sets of results, the first measured after 50 trials per channel condition (i.e., average SNR and correlation) and the second measured after 500 trials. The results for 50 trials represent the performance over a relatively short time frame while the results for 500 trials represent a significantly longer time frame. Multiple trends can be seen here. First, the overall return is improved when we increase the number of trials. This is because the overall cost of exploration is higher for a short time horizon in comparison to a long time horizon. Second, after a small number of trials, the total return is negatively

[4]Gittins [33] provides tables for γ equal to 0.5, 0.6, 0.7, 0.8, 0.9 and 0.99 for both NRP and BRP

TABLE I
AVERAGE TOTAL RETURN OVER OPTIMAL TOTAL RETURN

	Number of Trials Performed					
	50	50	50	500	500	500
	Discount Factor, γ					
Method	.5	.7	.99	.5	.7	.99
Gittins Index, NRP	.73	.73, .70[a]	.72	.93	.93, .93[a]	.94
Gittins Index, BRP	.60	.59, .65[a]	.56	.89	.89, .87[a]	.85
	Exploration Parameter, ϵ					
Method	.01	.1	.2	.01	.1	.2
ϵ-greedy strategy	.53	.65, .50[a]	.70	.74	.87, .87[a]	.87

[a]The maximum pairwise antenna correlation, ρ, equal to .5 for these scenarios. For all others $\rho = 0.1$.

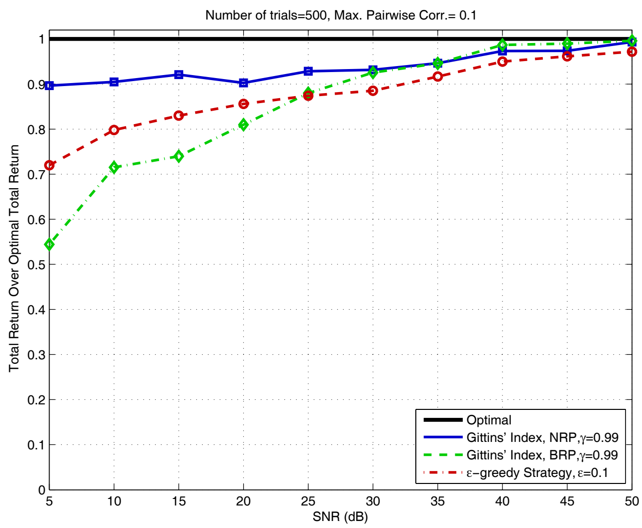

Fig. 3. Total return vs. SNR, 500 trials.

impacted by large discount factors. Large discount factors encourage exploration due to the anticipated long time horizon (i.e., large discount factors emphasize future rewards) which hurts short term performance. Third, we see that the use of a NRP as opposed to a BRP improves performance. We will discuss this shortly. Finally, we see that the Gittins index methods both outperform the ϵ-greedy approach in almost every case. The Gittins index method not only learns the performance of the radio, but it adapts the amount of exploration, whereas the ϵ-greedy approach maintains a constant exploration rate, despite what it knows about the performance.

The last paragraph raises an important point worthy of further discussion. The discount factor γ adjusts our time horizon. A small value of γ translates to a short time horizon, and a γ close to 1 translates to a horizon approaching infinity. The goal of the Gittins index method is to optimally balance exploration vs. exploitation subject to our time horizon. If we have a short time horizon, more value will be given to methods that were found to work than to methods with limited information about their actual performance. On the other hand, if we have a long time horizon, methods with limited information will be assigned higher values, therefore, making them more attractive for exploration. In other words, the risk taken (exploration) is proportional to the value of the discount factor. A higher risk may impair short term operation

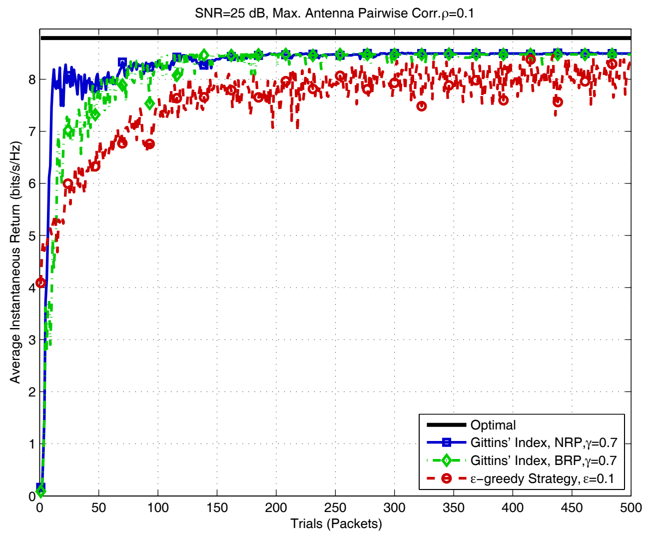

Fig. 4. Average instantaneous return vs. trials, SNR = 25 dB.

(trying methods that do not work), but will benefit long term operation because it is more likely for the optimal method to be found. To use an analogy to human behavior, when we have a small amount of time, we go with what works. Only when we have the luxury of time do we explore riskier but potentially rewarding options.

The results in Table I were mostly obtained assuming a low antenna correlation, $\rho = 0.1$. However, there is a small set for the case of $\rho = 0.5$ (Section IV-D). In general, increasing correlation reduces the number of options that achieve the desired goal and thus there is higher exploration cost, albeit minor depending on the correlation difference. Again, this isn't universal, but is a fairly consistent trend.

Figure 3 plots the total achieved reward over the optimal reward value by using the Gittins index methods for $\gamma = 0.99$ and the ϵ-greedy strategy for $\epsilon = 0.1$ at different average SNR values. The results of Figure 3 show: first, that all the methods have degraded performance at low SNRs. Performance for the Gittins methods is not guaranteed to be monotonic with SNR, but the general trend is clearly monotonic. In general lower SNR values result in fewer viable communications methods and correspondingly higher exploration costs. The reason for the lack of strictly monotonic behavior is that at each level of SNR, a different set of options will be useful. Consequently, it will take a different amount of time for the CE to learn which options are reliable. For low values of SNR learning is particularly slow for a BRP, and as a result it provides the worst efficiency at low SNR. However, at very high SNR, learning is very fast and therefore, it is the most efficient. This is because failures are treated more critically in the NRP as in comparison to the BRP thus rejecting bad results faster. Considering the totality of results, we have found that the NRP model is the better option.

A look at the instantaneous return is provided by Figure 4 for an SNR equal to 25. Specifically, we plot the average instantaneous return for 100 separate runs. Several things may be seen here. First, it may be observed that after approximately 150 trials all the methods achieve a return very close to the return achieved after 500 trials. Second, the similar trends in the relative performance are observed. The NRP approach is superior to the BRP approach at a low number of trials due to the higher exploration costs associated with the BRP, but eventually the instantaneous returns are very similar. Further, the ϵ-greedy approach performs the worst (except at very low numbers of trials) because it has a fixed exploration probability and thus a non-decreasing exploration cost.

It should be noted that the results obtained by all the methods tested are always subject to the parameters used and are also dependent on the number of available options and the underlying conditions. For example, the ϵ-greedy strategy is known to suffer when the number of options is very large [38]. However, in the given results and other results not shown here we have seen some general trends. For example, the Gittins index approach with the NRP tends to perform the best verifying the *long-term* optimality of the Gittins indices. Additionally, the NRP has a much simpler table and extrapolation method than the BRP table which has significantly more values and requires a more sophisticated extrapolation method. The work of this section is presented with more detail in [3].

V. Applying Prior Knowledge

Analytically predicting the best configuration for a set of channel conditions is a complex task. This is especially true for complex systems, such as the ones that use multiple antennas. This challenge was one of the motivations for developing a CE. Even though it is not always easy to predict performance analytically, many useful analytical results do exist; thus it would not be wise for the CE not to benefit from them. For example, BER and capacity curves exist for most communication methods in an AWGN channel. In this section, we demonstrate how we can convert analytical data to observations.

A. Converting knowledge to Observations

In Section III-A we showed how the parameters α_0, β_0 of the Beta distribution can be used to take into account prior beliefs (knowledge) about the posterior distribution which is also Beta distributed. A typical way to convert knowledge to α_0, β_0 is by using the mean $\hat{\theta}$ and the variance σ^2 of our prior beliefs:

$$n = \frac{\hat{\theta}(1-\hat{\theta})}{\sigma^2} \tag{22}$$

$$\alpha_0 = \hat{\theta} n \tag{23}$$

$$\beta_0 = n - \alpha_0 \tag{24}$$

Although we do not typically have the exact mean and variance of our previous knowledge, this method can be still used as a guideline to generate α_0, β_0 from our prior knowledge by using an estimate of $\hat{\theta}$ and σ^2. The latter can be estimated, with some assumptions, using the normal approximation method of Section III-B. Given an estimate of $\theta \in (\theta_l, \theta_u)$ with an assumed δ we estimate $\hat{\theta}$ and σ using:

$$\hat{\theta} = \frac{\theta_l + \theta_u}{2} \tag{25}$$

$$\sigma = \frac{\hat{\theta} - \theta_l}{z_{\delta/2}} = \frac{\theta_u - \hat{\theta}}{z_{\delta/2}} \tag{26}$$

where $z_{\delta/2}$ is the $1-\delta/2$ percentile of the standard normal distribution. As $\sigma \rightarrow 0$, $n \rightarrow \infty$. This is equivalent to stating that we are extremely confident about $\hat{\theta}$. This is not a valid statement, because the θ_l & θ_u estimates are generated by models and may or may not represent the actual performance. An arbitrarily high n might cause the CE to focus on under-performing methods. In this case, the CE will require a significant number of trials to overcome the prior beliefs. For this reason, either n should be limited or a σ_{min} should be set. In this case, should the prior observations not reflect reality, the CE can quickly revise its beliefs based on the actual observations.

B. Capacity Curves

Capacity curves can be used to disqualify methods that have a capacity higher than the capacity predicted by the capacity curves. Such methods are assigned an arbitrarily high number of failures and as a result the CE ignores them.

If no spatial multiplexing is used, the link capacity (spectral efficiency) is given by the well known Shannon formula:

$$C = log_2\left(1+\gamma_S\right) \quad (bits/s/Hz) \tag{27}$$

where γ_S is the SNR at the receiver.

When spatial multiplexing is used and the channel is unknown to the transmitter the capacity is given by[39]:

$$C = \sum_{i=1}^{\min\{N_t, M_r\}} \log_2\left(1+\frac{\gamma_S}{N_t}\lambda_i\right) \quad (bits/s/Hz) \tag{28}$$

where N_t and M_r are number of transmit and receive antennas respectively. λ_i is the ith eigenvalue of the HH^H matrix, where H is the $M_r \times N_t$ channel matrix.

C. SER & BER Curves

The most typical performance indicators are BER or Symbol-Error-Rate (SER) curves. Most of the available fading channel SER curves, for multi-antenna methods, are upper bounds. For this reason, we use the AWGN SER curves as a lower bound. The SER curves used in this work are shown in Table II. The first three curves listed are for an AWGN channel, and the last three curves for a fading channel. The coding gain g_c for a convolutional codec using a soft decision decoder, at a BER=10^{-3}, is given by [40]:

$$g_c \approx \frac{r d_{free}}{2} \tag{29}$$

where r and d_{free} are the rate and the free distance of the code respectively. A hard decision decoder reduces the gain by an additional factor of two [37]. A hard decision decoder is used in the VBLAST decoder because the received symbols need to be hard estimated as a part of the successive interference cancellation employed. Assuming Gray coding, the BER can be approximated from the SER by [37]:

$$P_b \approx \frac{1}{log_2(M)} P_s \tag{30}$$

TABLE II
SYMBOL ERROR RATE BOUNDS AND APPROXIMATIONS[a]

M-PSK[b]	$P_s = 2Q\left(\sqrt{2\frac{\gamma_S M_r}{N_t} g_c} sin\left(\frac{\pi}{M}\right)\right)$
M-QAM	$P_s = 1 - \left[1 - 2\left(1 - \frac{1}{\sqrt{M}}\right) Q\left(\sqrt{\frac{3}{M-1}\frac{\gamma_S M_r}{N_t} g_c}\right)\right]^2$
M-QAM[c]	$P_s \leq 1 - \left[1 - 2Q\left(\sqrt{\frac{3}{M-1}\frac{\gamma_S M_r}{N_t} g_c}\right)\right]^2$
MRC[d]	$P_s \leq M_n \prod_{i=1}^{N_t M_r} \frac{1}{1 + \left(\gamma_S g_c \frac{d_{min}^2}{4N_t} \lambda'_i\left(R\right)\right)}$
STBC[e]	Same as MRC with $\gamma_S \equiv r\gamma_S$
V-BLAST[f]	$P_s \leq M_n \left(\frac{\gamma_S g_c d_{min}^2 \lambda_{min}}{2M_T}\right)^{-(M_r - N_t + 1)}$

[a]adapted from [37] & [39]. [b]g_c is the coding gain. [c]odd $log_2 M$. [d]M_n is the max. # of neighboring symbols within d_{min}. λ'_i is the ith eigen-value of $R = E\left\{vec\left(H\right) vec\left(H\right)^H\right\}$. [e]$r$ is the rate of the STBC code. [f]λ_{min} is the smallest eigen-value of HH^H.

Assuming a packet with N_b bits that can tolerate a maximum of n_e errors, then the PSR θ can be calculated using:

$$\theta = 1 - P[\#\text{errors} \leq n_e] = \sum_{i=0}^{n_e} \binom{N_b}{N_b - i} P_b^i (1 - P_b)^{N_b - i} \tag{31}$$

Using (30) & (31), θ_u can be estimated using the AWGN SER curves from Table II. Likewise, θ_l can be estimated using the fading channel SER curves from Table II. Next, θ_l & θ_u can be converted to observations as described in Section V-A. Finally, the resulting observations are used to initialize the observations database.

D. Observations Database

The methods discussed are applied subject to the channel conditions. Our proposed method is to keep a database of α, β pairs for a discrete set of channel conditions. Specifically, we represent the channel conditions by the average SNR γ_S and the eigen-spread κ (also known as the Demmel condition number) of the channel. In this paper, we assume that the channel stays constant during the time required to decide and send the packet. It is well established that the SNR is an important performance predictor for any system. We augment this with the eigen-spread of the channel since it can represent the quality of the available spatial channels (important for multi-antenna system performance). The eigen-spread is defined as $\kappa = \frac{\lambda_{max}}{\lambda_{min}}$, where λ_{max} and λ_{min} are the maximum and minimum eigen-values of the HH^H matrix. A $\kappa \approx 1$ means that the spatial channels have minimum to no correlation and spatial multiplexing can be readily used. A high κ means that the spatial channels are highly correlated and spatial multiplexing is not feasible. A similar argument applies to diversity techniques. The observations database is built by discretizing the SNR and eigen-spread values.

In an actual implementation, a balance between the observations database's granularity and the system's performance needs to be obtained by applying more advanced discretization methods such as clustering (which we are not considering at this stage). Generally, finer granularity will improve performance because it will allow the CE to identify maximal operating regions with higher precision. However, extremely fine granularity may increase the database size and the operations needed to unmanageable levels. On the other hand, extremely coarse granularity will not allow the CE to precisely identify maximal performance points, and as a result it will only be able to reach a fraction of the maximal performance. For the purposes of this work, we found that 1 dB steps of the SNR and 0.5 steps of $log_{10}(\kappa)$ give a reasonable balance between performance and the computing resources needed.

E. Propagating information

At each γ_S, κ, and communication method set the observations database is updated by adding the respective observed successes and failures. However, as the CE works in varying channel conditions, even the smallest change in channel conditions would appear to the CE as a completely different environment. As a result there is no benefit from neighboring (*i.e.*, similar) observations. Some simple facts from communication fundamentals can be used to minimize this problem. For example, we know that: first, increasing SNR increases PSR. Second, greater minimum distance d_{min} between adjacent symbols also improves PSR. Finally, an increased eigen-spread reduces spatial diversity and the PSR decreases. The reverse relationships also apply. Based on the above, each successful trial is propagated (added) to higher SNR, lower eigen-spread and higher d_{min} entries and each failed trial is propagated to lower SNR, higher eigen-spread and lower d_{min} entries. With these simple rules, each observation receives maximum utilization and directly contributes to the overall learning rate.

VI. OPERATION EXAMPLE

In this section, we apply the methods we developed in a scenario with dynamically varying channel conditions. The CE was found to be able to operate by balancing exploration *vs.* exploitation and reach near optimal performance in a relatively short time. Moreover, in a subsequent scenario, the CE successfully used the experiences from the previous scenario to optimize short term performance when revisiting the same channel conditions.

We assumed that there is a stationary and a mobile radio that need to communicate with each other. The stationary radio is in the center of a 1000 square foot area and that the mobile radio moves within that area. We created a signal level map (Figure 5) that applies fading, shadowing, and path loss to the transmitted signal. We simulated the performance of the link controlled by the CE assuming that the radio randomly moves within the map area.

A. Operation Parameters

In this example, the objective of the link is to maximize throughput. The CE applies all the techniques developed: prior knowledge is applied, the Gittins index method is used, and experiences are propagated. Furthermore, the CE utilizes the following ($M_r \times N_t$) multi-antenna techniques: 4×1 MRC, 4×2 STBC, 4×4 STBC, 4×2 VBLAST, and 4×4 VBLAST.

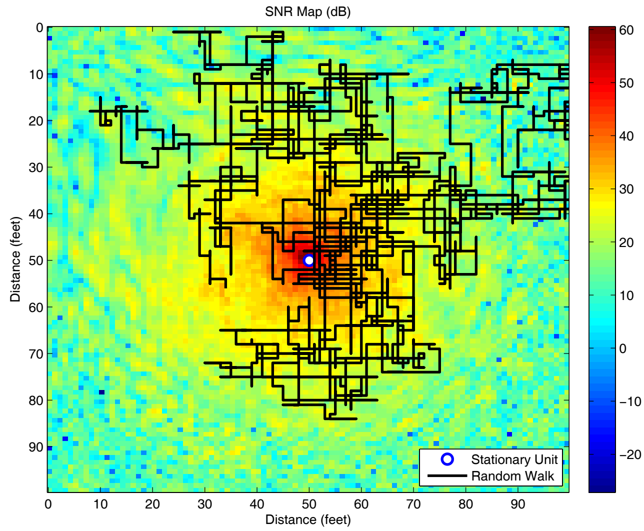

Fig. 5. SNR level map.

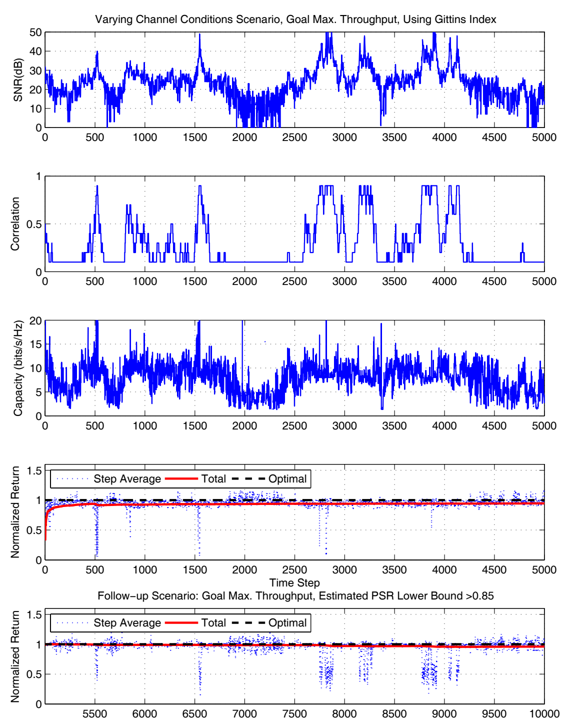

Fig. 6. Operation time line for section VI.

The five options had rate multipliers of 1, 1, 3/4, 2, and 4 respectively. The modulation and coding options were the same as Section IV-D. Thus, the total options per channel condition were 110 *vs.* 66 that were used in Section IV-D.

B. Results

Goal: Balancing Exploration vs. Exploitation: Before the CE's operation, the observation database is pre-populated with prior knowledge (Section V). The CE begins by attempting to balance exploration *vs.* exploitation by using the Gittins index method. Figure 6 (time steps 1 to 5000) shows a timeline of the channel conditions and the normalized performance. Each time step represents 20 attempted packets. Optimal refers to the best possible using the available options found by a brute force search. Beginning with just prior information in the observations database, it can be observed that: First, the CE reaches 88% of the normalized return in just 100 time steps. Second, the performance suffers more at the channel condition extremes: High/low SNR and high/low spatial correlation. However, as the CE revisits similar conditions, the performance is smoother than the last attempt. The reason that the performance suffers in the extremes is because of some limitations in the knowledge propagation process, *e.g.*, the highest SNR point will receive propagated successes but no failures. If the corresponding point has some failures associated with it, the CE has no choice but to determine the failures by trial and error. Similar situations apply to the other extreme points. Third, the optimal capacity at both the low SNR and high correlation points is relatively low. A small variation in the achievable capacity *vs.* the optimal will result in a high normalized difference. For example, if the optimal capacity is 1 bits/s/Hz (MRC, QPSK, $1/2$ codec), the remaining suboptimal options have a capacity of $1/2$ and $1/4$ bits/s/Hz, that is, MRC & QPSK with a $1/4$ and a $1/8$ codec respectively. If the former suboptimal options is used, the system will achieve only 50% of the optimal capacity, and just 25% if the latter suboptimal option is used. The performance variation in the extremities of the channel conditions is a result of the design, and it is not limited to this example.

Goal: Short Term Performance: In this scenario, the goal is to ensure short term performance of maximum throughput, subject to $\theta_l > 0.85$. The CE is using the observations database from the previous scenario. This observations database includes all the experiences gained up to time step 5000. Using the previous experiences, throughput is maximized for the next 5000 time steps. It is assumed that the radio repeats the exact same walk, and it gets the same average channel conditions at each step. However, at each step, every channel realization is distinct. The different channel realizations may cause some performance fluctuations in operating regions with limited knowledge. Even though the CE is not actively exploring, it still learns based on the packet successes or failures on the various channel conditions.

The bottom plot in Figure 6 (time steps 5001 to 10000) demonstrates that the CE achieves on average 97% of the optimal throughput even though the performance suffers at some extreme channel conditions for the same reasons explained in the previous goal discussion. The actual performance numbers may vary in different scenarios, but the trend of being able to learn and quickly reach near maximal performance should always be expected. This claim was validated by running similar simulations over various operating scenarios.

VII. Conclusion

We have developed a concise CE for link adaptation in multi-antenna systems. We demonstrated that all the operations of the CE can be based on just the number of successful and failed packets associated with a given communication method and channel condition. This was made possible by using a classic Bayes' Rule-based statistical learning method.

The latter provided an elegant way to combine prior knowledge and new observations to obtain a posterior belief about the system's performance. The result is an efficient and straightforward fully functional design, which can be potentially implemented in any radio (even resource constrained). A potential implementation can be based on table lookups (*e.g.*, normal distribution, Gittin's indices) and elementary operations.

The efficiency of the design does not limit the CE's functionality: from the number of successes and failures the CE can estimate confidence intervals of the packet success rate using the Beta distribution, or simple approximations. The confidence intervals allow the CE to quantify the expected performance of its operations both to itself and to other devices, if part of a network. The expected performance can be then used to make appropriate configuration decisions. Second, the designed CE balances the tradeoff between learning and ensuring short term performance (*i.e.*, exploration *vs.* exploitation) by applying the Gittin's dynamic allocation indices method. This minimizes the cost of learning while providing mostly operable links. Third, it learns the radio's abilities independently of the operation objectives, so that if an objective changes, any information regarding the radio's abilities is not lost. Finally, prior knowledge such as capacity, BER curves, and basic communication principles are used to initialize the CE's knowledge and maximize the learning rate across different channel conditions. The proposed CE is demonstrated to be able to learn in a dynamic scenario and quickly reach near-maximal performance.

Our CE does not consider MAC issues; nevertheless, it is expected to operate as designed on its assigned band and/or transmission slot. However, in this case some mechanism of collision detection needs to be implemented so the CE will know the true cause of a failed packet.

Furthermore, the CE should be able to perform both under the presence of noise and interference. However, they have slightly different properties and the exact effect of interference to the CE's operation will be studied as part of future work. We are also looking into how to enable the CE to cope with outdated and noisy channel estimates.

Acknowledgment

This work was supported by the National Science Foundation under Grant No. 0520418. Also the authors would like to thank Christopher I. Phelps for sharing his MIMO MATLAB code.

References

[1] H. I. Volos, C. I. Phelps, and R. M. Buehrer, "Initial design of a cognitive engine for MIMO systems," in *SDR Forum Tech. Conf.*, Nov. 2007.

[2] ——, "Physical layer cognitive engine for multi-antenna systems," in *Proc. IEEE Military Commun. Conf.*, Nov. 2008, pp. 1–7.

[3] H. I. Volos and R. M. Buehrer, "On balancing exploration vs. exploitation on a cognitive engine for multi-antenna systems," in *Proc. IEEE Global Telecommun. Conf.*, Nov. 2009, pp. 1-6.

[4] J. Mitola III, "Cognitive radio: an integrated agent architecture for software defined radio," Ph.D. dissertation, KTH, Stockholm, Sweden, May 2000.

[5] A. He, K. Bae, T. Newman, J. Gaeddert, K. Kim, R. Menon, L. Morales, J. Neel, Y. Zhao, J. H. Reed, and W. H. Tranter, "A survey of artificial intelligence for cognitive radios," *IEEE Trans. Veh. Technol.*, to appear, 2010.

[6] C. J. Rieser, "Biologically inspired cognitive radio engine model utilizing distributed genetic algorithms for secure and robust wireless communications and networking," Ph.D. dissertation, Virginia Tech, 2004.

[7] T. W. Rondeau, "Application of artificial intelligence to wireless communications," Ph.D. dissertation, Virginia Tech, 2007.

[8] B. Le, T. W. Rondeau, and C. W. Bostian, "Cognitive radio realities," *Wiley J. Wireless Commun. Mobile Comput.*, vol. 7, no. 9, pp. 1037-1048, 2007.

[9] A. He, J. Gaeddert, K. Bae, T. R. Newman, J. H. Reed, L. Morales, and C. Park, "Development of a case-based reasoning cognitive engine for IEEE 802.22 WRAN applications," *ACM Mobile Comput. Commun. Rev.*, vol. 13, no. 2, pp. 37-48, 2009.

[10] T. R. Newman, B. A. Barker, A. M. Wyglinski, A. Agah, J. B. Evans, and G. J. Minden, " Cognitive engine implementation for wireless multicarrier transceivers," *Wiley J. Wireless Commun. Mobile Comput.*, vol. 7, no. 9, pp. 1129-1142, 2007.

[11] Z. Zhao, S. Xu, S. Zheng, and J. Shang, " Cognitive radio adaptation using particle swarm optimization," *Wiley J. Wireless Commun. Mobile Comput.*, 2008, published online.

[12] N. Baldo and M. Zorzi, "Learning and adaptation in cognitive radios using neural networks," in *Proc. 5th IEEE Consumer Commun. Netw. Conf.*, Jan. 2008, pp. 998-1003.

[13] C. Clancy, J. Hecker, E. Stuntebeck, and T. O'Shea, "Applications of machine learning to cognitive radio networks," *IEEE Wireless Commun.*, vol. 14, no. 4, pp. 47-52, Aug. 2007.

[14] N. Devroye, P. Mitran, and V. Tarokh, "Achievable rates in cognitive radio channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 1813-1827, May 2006.

[15] ——, "Limits on communications in a cognitive radio channel," *IEEE Commun. Mag.*, vol. 44, no. 6, pp. 44-49, June 2006.

[16] S. Haykin, "Fundamental issues in cognitive radio," in *Cognitive Wireless Communication Networks*, E. Hossain and V. K. Bhargava, editors. New York: Springer, 2007, pp. 1-43.

[17] A. Sahai, R. Tandra, S. M. Mishra, and N. Hoven, "Fundamental design tradeoffs in cognitive radio systems," in *ACM First International Workshop Technol. Policy Accessing Spectrum*, Aug. 2006.

[18] S. Sridharan and S. Vishwanath, "On the capacity of a class of MIMO cognitive radios," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 103-117, Feb. 2008.

[19] S. Jafar and S. Shamai, "Degrees of freedom of the MIMO X channel," in *IEEE Global Telecommun. Conf.*, Nov. 2007, pp. 1632-1636.

[20] G. Scutari, D. P. Palomar, and S. Barbarossa, "Cognitive MIMO radio: a competitive optimality design based on subspace projections," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 49-59, 2008.

[21] S. Haykin, D. Thomson, and J. Reed, "Spectrum sensing for cognitive radio," *Proc. IEEE*, vol. 97, no. 5, pp. 849-877, May 2009.

[22] D. Datla, R. Rajbanshi, A. Wyglinski, and G. Minden, "An adaptive spectrum sensing architecture for dynamic spectrum access networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4211-4219, Aug. 2009.

[23] R. W. Thomas, D. H. Friend, L. A. Dasilva, and A. B. Mackenzie, "Cognitive networks: adaptation and learning to achieve end-to-end performance objectives," *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 51-57, Dec. 2006.

[24] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "Next generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Comput. Netw.*, vol. 50, no. 13, pp. 2127-2159, 2006.

[25] T. Clancy and N. Goergen, "Security in cognitive radio networks: threats and mitigation," in *Proc. 3rd International Conf. Cognitive Radio Oriented Wireless Netw. Commun.*, May 2008, pp. 1-8.

[26] A. He, S. Srikanteswara, K. K. Bae, T. R. Newman, J. H. Reed, W. H. Tranter, M. Sajadieh, and M. Verhelst, "System power consumption minimization for multichannel communications using cognitive radio," in *Proc. IEEE International Conf. Microwaves, Commun., Antennas Electron. Syst.*, Nov. 2009, pp. 1-5.

[27] L. Lai, H. E. Gamal, H. Jiang, and H. V. Poor, "Cognitive medium access: exploration, exploitation and competition," *IEEE/ACM Trans. Networking*, Oct. 2007, submitted.

[28] J. Gittins and D. Jones, "A dynamic allocation index for the sequential design of experiments," *Progress Statistics*, pp. 241-266, 1974.

[29] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Pearson Education, 2003.

[30] P. M. Lee, *Bayesian Statistics: An Introduction*, 2nd edition. Oxford, England: Arnold, 1997.

[31] T. Leonard and J. S. J. Hsu, *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers*. Cambridge University Press, June 1999.

[32] L. M. Leemis and K. S. Trivedi, "A comparison of approximate interval estimators for the Bernoulli parameter," *American Statistician*, vol. 50, no. 1, pp. 63-68, 1996.
[33] J. C. Gittins, *Multi-Armed Bandit Allocation Indices*. Wiley, 1989.
[34] J. N. Tsitsiklis, "A short proof of the Gittins index theorem," *Annals Applied Probability*, vol. 4, no. 1, pp. 194-199, 1994.
[35] D. Acuña and P. Schrater, "Bayesian modeling of human sequential decision-making on the multi-armed bandit problem," in *Proc. 30th Annual Conf. Cognitive Science Society*, V. Sloutsky, B. Love, and K. McRae, editors. Washington, DC: Cognitive Science Society, 2008.
[36] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, Mar. 1998.
[37] J. Proakis, *Digital Communications*, 4th edition. New York: McGraw-Hill, 2001.
[38] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics)*. Wiley-Interscience, 2007.
[39] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.
[40] C. I. Phelps, "Adaptation for multi-antenna systems," Master's thesis, Virginia Tech, 2009.

Haris I. Volos graduated in 2002 with a Higher National Diploma (HND) in Electrical Engineering and was awarded Best Performance in both Electronic and Power Engineering subjects from the Higher Technical Institute (HTI) in Nicosia, Cyprus. Also in 2002, Volos was awarded the International Student Scholarship (ISS) and attended Old Dominion University, where he was a graduate of the honors college with a BS (summa cum laude) in Electrical Engineering and a minor in Computer Engineering.

In 2004 Volos was awarded the Cyprus America Scholarship Program (CASP) Scholarship for MS studies at Virginia Polytechnic Institute and State University. He subsequently was awarded in 2006 an MS degree for his thesis titled "Ultra Wideband Ranging and Link Budget Design for Naval Crane Applications." He is now a Ph.D. candidate at Wireless @ Virginia Tech researching cognitive engine design issues.

In addition, he is working with OSSIE, Virginia Tech's open source SCA implementation. He was part of the SDR forum smart radio challenge 2007 best design award winning team. In 2008 he was awarded a best paper award for the SDR forum paper "Initial Design of a Cognitive Engine for MIMO Systems." Finally, he is participating in other interesting projects such as Ultra Wideband (UWB) channel measurements in unique enviroments.

Dr. R. Michael Buehrer joined Virginia Tech from Bell Labs as an Assistant Professor with the Bradley Department of Electrical Engineering in 2001. He is currently an Associate Professor and is member of Wireless @ Virginia Tech, a comprehensive research group focusing on wireless communications. During 2009 Dr. Buehrer was a visiting researcher at the Laboratory for Telecommunication Sciences (LTS) a federal research lab which focuses on telecommunication challenges for national defense. While at LTS, his research focus was in the area of cognitive radio with a particular emphasis on statistical learning techniques (viz. classification approaches) applied to forward error correction coding.

His current research interests include dynamic spectrum sharing, cognitive radio, communication theory, Multiple Input Multiple Output (MIMO) communications, intelligent antenna techniques, position location networks, Ultra Wideband, spread spectrum, interference avoidance, and propagation modeling. His work has been funded by the National Science Foundation, the Defense Advanced Research Projects Agency, Office of Naval Research, and several industrial sponsors.

Dr. Buehrer has authored or co-authored over 35 journal and approximately 100 conference papers and holds 11 patents in the area of wireless communications. He is currently a Senior Member of IEEE, and an Associate Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and IEEE TRANSACTIONS ON COMMUNICATIONS. He was formerly an associate editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGIES, IEEE TRANSACTIONS ON SIGNAL PROCESSING, and IEEE TRANSACTIONS ON EDUCATION. In 2003 he was named Outstanding New Assistant Professor by the Virginia Tech College of Engineering.