

# Stochastic Recruitment: A Limited-Feedback Control Policy for Large Ensemble Systems

Lael Odhner and Harry Asada

**Abstract**—This paper is about stochastic recruitment, a control architecture for centrally controlling the ensemble behavior of many identical agents, in a manner similar to motor recruitment in skeletal muscles. Each agent has a finite set of behaviors, or states, which can be switched based on a broadcast command. By switching randomly between states with a centrally determined probability, it is possible to designate the number of agents in each state. This paper covers stochastic recruitment policies for the case when little or no feedback is available from the system. Feed-forward control policies based on rate equilibria are presented, with an analysis of the performance trade-offs inherent in the problem. Minimal feedback control laws are also discussed, and a policy is presented which minimizes the expected convergence time of the system given only the ability to halt the system when the desired output has been achieved.

## I. INTRODUCTION

Finding methods for centrally controlling the collective behavior of many identical agents is an active research topic across diverse communities, including robotics, bioengineering, and control. Many different systems can be modeled as a swarm of identical agents, such as bacteria, cells, robots in a swarm, or computers in a network. In this paper, we will explore control of a class of finite state or hybrid state agents that we believe to be applicable to a variety of these problems. In particular, we wish to regulate the ensemble distribution of many agents over their discrete states, or the fraction of agents having a particular discrete state. The discrete states in question could be tasks performed by swarm robots, cell migration behaviors, or switched modes governing the time evolution of some continuous state variable. One of the key components of these regulatory mechanisms is stochasticity. In biological systems, stochasticity is a ubiquitous phenomenon that can be observed in mechanochemical cell and molecular behaviors. For example, angiogenesis cell migration and tissue development are often treated as directed random walks, where stochastic behaviors play key roles for developing multi-cellular structures that meet morphological and functional requirements [1], [2]. Natural systems are built upon random processes, and biological systems in particular exploit the stochastic nature of their building blocks. In control engineering, randomness has been treated as unwanted behavior that should be filtered out or avoided. Except for communication technology and some of the system identification techniques, randomness has not yet been fully exploited as a useful concept. In general, we assume that the transition between states can be modeled as a Markov state transition graph whose state transition probabilities are determined by

some input given to all agents in the system. The authors are actively working on applying this control framework to the problems of endothelial cell migration and artificial muscle actuators. Figure 1 illustrates the basic system architecture.

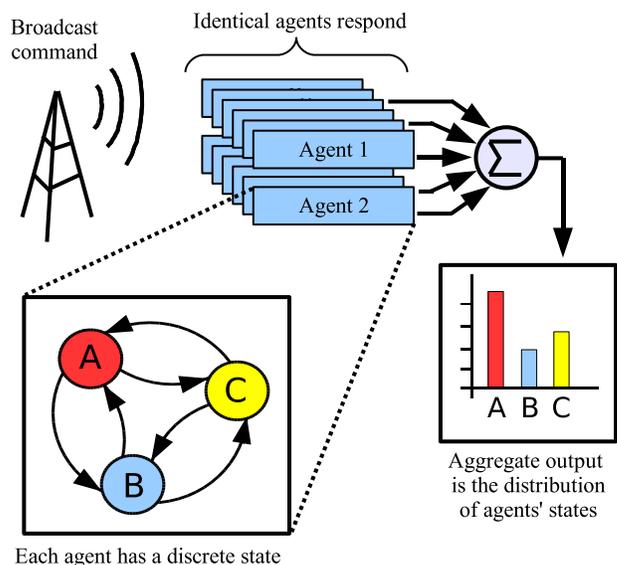


Fig. 1. Many systems in nature can be thought of as an ensemble of functionally similar discrete-state agents that each respond independently to some global stimulus. The ensemble output behavior of many such agents is the number of agents in each discrete state.

In this paper, we discuss the performance of feed-forward control policies for controlling the number of agents in each discrete state. If rich information usable for feedback control is available, then many control techniques that presume full-state knowledge can be applied. The authors have shown that simple feedback policies can be formulated for controlling the ensemble state distribution based on linear feedback laws [3], expectation-based control laws [4], and dynamic programming [5]. However, the output measurements from real systems often provide only uncertain estimates of system state distribution. For example, a robot swarm might have limited communication with a central control station. An artificial muscle composed of many individual sub-units may (and usually will) exhibit some output delay or hysteresis, so that some estimating filter must be used to predict the individual units' states. Uncertainty of this form will lead to increased uncertainty in the closed-loop behavior of the system. We will show that we can formulate open-loop control policies that cause the state distribution to reach a stochastic equilibrium that is close to any desired state distribution, having well-characterized variance. The performance of policies depending

o minimal or no feedback provide a good bounding case guidelines for examining the performance of closed-loop laws. Depending on the degree of uncertainty in the state distribution estimated from measurements, it may actually be advantageous to use an open-loop policy instead of a closed-loop policy.

In this paper a stochastic cellular system with limited output feedback will be formulated in its simplest form. Basic convergence properties and control policy will be discussed to gain insights into collective behaviors of stochastic cellular systems. The framing of the problem is based on cell and systems biology and motivated by needs in robotics and control. The paper, however, does not aim to apply the results directly to a specific area; instead the objective is to better understand the possible regulatory mechanism that may govern a large population of cellular agents.

## II. INTRODUCING STOCHASTIC RECRUITMENT

In an attempt to investigate a regulatory mechanism that works with limited or no feedback, we look at biological control systems, extract key features from there, and formulate a simple, abstract problem for detailed analysis. One striking difference from traditional engineered systems is that biological cells are living in a wet environment, where signals propagate through diffusion. Stimuli to the process pervasively affect all the cells involved in the wet environment. It is not likely that each cell receives a specific control signal from a central controller. Rather, the control signal, if it exists, is broadcast in nature.

In response to stimuli, an important property of biological systems is that each cell's behavior is stochastic in nature. Although receiving the same stimuli, the cell's response is randomly chosen, conditioned on the environmental factors. Endothelial cell migration, for example, each cell's movement is a random process, switching directions stochastically [1]. Furthermore, the state transition probabilities often vary depending on the stimuli that individual cells receive. It is known that biochemical kinetics highly depend on temperature and other factors, resulting in changes to state transition probabilities. In the case of endothelial cell migration, state transition probabilities are modulated by the stimuli each agent receives as well as the conditions of the wet environment to which it is exposed [2].

These stochastic behaviors of biological cells suggest a new control methodology for engineered systems; a central controller broadcasts signals that modulate state transition probabilities to be used at individual agents, or directly broadcasts state transition probabilities across the cell population. Whichever the case, the central controller does not dictate each agent to take a specific transition, but only specifies the probabilities of transition. The actual control action is up to each cellular agent, but the collective behavior may be effectively controlled with the central controller. This broadcast control protocol will be useful for engineered systems, since it requires very little bandwidth, and allows each cellular agent to be completely anonymous.

The authors have developed several types of stochastic broadcast control [3], [5], [6]. The major difference is that the

previous control architecture exploits feedback of aggregate output, while this paper addresses control issues with no feedback or limited feedback. We call this control "Stochastic Recruitment", which will be formally described next.

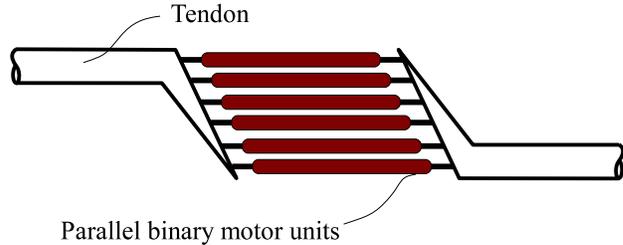


Fig. 2. Muscles are composed of many small motor units, which are either relaxed or activated, producing force. The net power and stiffness of the muscle depend on the number of recruited motor units.

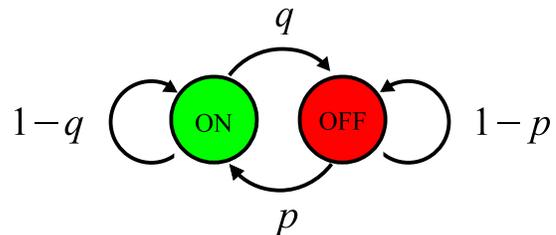


Fig. 3. A small automaton having two states, *ON* and *OFF*, can be commanded to transition between states with probabilities  $p$  and  $q$ .

The concept of recruitment is best explained by examining the function of skeletal muscle. A muscle is not a homogenous collection of individual cells; rather it is organized as a network of many small motor units, which respond autonomously to stimulus from the spinal column, as shown in Figure 2. A command from the nervous system produces a varied level of force by affecting the percentage of motor units which contract. [7]. Each motor unit has a threshold for response to the nervous excitation, so that varying the excitation being sent to the muscle acts to modulate the force produced [8]. This architecture can be replicated in engineered systems composed of a similar arrangement of subunits, like an artificial muscle [3]. A central controller broadcasts some excitation to all of the subunits, which change their behavior at a rate based on the excitation. The authors have been using two-state agents of this kind to control shape memory alloy muscle actuators made up of many binary units [5]. Collectively,  $N$  small actuators produce binary outputs which are modulated so that varied force and displacement are produced as a function of the number of *ON* (force-producing) agents,  $N_t^{on}$ . The probability per unit time of state transitions between the *ON* and *OFF* states are described by parameters  $p$  and  $q$ , as shown in Figure 3.

### A. The Dynamics of Randomized Recruitment

A key appealing feature of this state transition framework is its similarity to statistical models for many real-world

phenomena, including the dynamics of gene regulation [9], chemical reactions [10], and even swarms of insects [11]. It is well understood how kinetic chemical reaction models yield well-described transient and steady-state behaviors. These dynamics can be exploited in engineered systems, since an artificial system can be designed to arbitrarily vary the state transition probabilities of its agents. To demonstrate this, we will consider the simplest recruitment problem, a system made up of  $N$  small subsystems or agents having just two states,  $ON$  and  $OFF$ . The control task is to recruit a specific number of agents,  $N^{ref}$ , into the  $ON$  state. These states could represent two different modalities or behaviors exhibited by each individual agent. To keep track of the state evolution of the system, we will introduce a discrete distribution variable,  $x_t$ , describing the probability with which each agent is  $ON$  at time  $t$ ,

$$x_t = P(\text{state} = ON) \quad (1)$$

Of course, using  $x_t$  as a sufficient statistic for predicting system behavior does not guarantee that  $N_t^{ON}$  can be known exactly. Instead, the likelihood that  $k$  units are on is calculated as a function of  $x_t$  using a binomial distribution,

$$P(N_t^{on} = k|x_t) = \binom{N}{k} x_t^k (1 - x_t)^{N-k} \quad (2)$$

There are multiple reasons why using  $x_t$  makes more sense than considering  $N_t^{on}$ , if little feedback information is available. First and foremost, it is quite simple to predict the future behavior of  $x$  given an initial condition and a broadcast command. In the absence of other information, such as a measured ensemble output, this will provide a good guess of long-term behavior. Second, as  $N$  becomes large, the central limit theorem will guarantee that  $N_t^{on}$  will approach its expected value,

$$E(N_t^{on}|x_t) = Nx_t \quad (3)$$

The evolution of  $x_t$  can be written as a recursive sequence based on the Markov graph parameters  $p$  and  $q$ ,

$$x_{t+1} = (1 - q)x_t + p(1 - x_t) = (1 - p - q)x_t + p \quad (4)$$

The time evolution of  $x_t$  can also be written as a deterministic sequence which satisfies (4),

$$x_{t+1} = \frac{p}{p+q} + \left(x_0 - \frac{p}{p+q}\right) (1 - p - q)^t \quad (5)$$

Here  $x_0$  is the initial likelihood that an arbitrarily selected unit is in the  $ON$  state. Some physical meaning can be gleaned from (5). The time-independent term of the sequence corresponds to the fraction of  $ON$  agents at steady state. This is equal to the probabilistic rate at which agents transition from  $OFF$  to  $ON$ , normalized by the sum of all transition rates between states,

$$x_{ss} = \frac{p}{p+q} \quad (6)$$

The rate at which  $x_t$  exponentially approaches  $x_{ss}$  from an initial condition  $x_0$  is represented in the transient term,

$$\lambda = 1 - p - q \quad (7)$$

These observations are important in formulating an open-loop control policy.

### III. A NO-KNOWLEDGE CONTROL POLICY

In previous work, the authors formulated closed-loop control laws by finding the state transition graph parameters that minimize the expected future error,

$$E(N_{t+1}^{on}|p, q, N_t^{on}) = N^{ref} \quad (8)$$

If the central controller has no knowledge of the number of agents in each state, then the control policy must produce feed-forward dynamics that move the state distribution toward the desired goal. Equation (6) demonstrated that the time evolution of  $x_t$  has a steady-state component. Instead of choosing  $p$  and  $q$  to minimize the one-step ahead error conditioned on the present state distribution,  $p$  and  $q$  could be chosen so that some desired number of cells  $N^{ref}$  is expected in the steady state according to (3),

$$E(N_{ss}^{on}|p, q) = Nx_{ss} = N \frac{p}{p+q} = N^{ref} \quad (9)$$

Many policies satisfy this constraint for any given  $N^{ref}$ . For example, setting  $p = 0.1$  and  $q = 0.1$  will drive the agents to a 50% likelihood of being in either state. So will setting  $p = q = 0.3$ . In order to distinguish between these cases, a scaling analysis can be used to weigh the performance tradeoffs between these policies.

#### A. Convergence Rate and Steady State Distribution are Independent.

The control policy parameters  $p$  and  $q$  can be rewritten as  $\beta p_0$  and  $\beta q_0$ , where  $p_0 + q_0 = 1$  and  $\beta$  is a scaling factor that varies subject to the constraint that  $0 < p < 1$  and  $0 < q < 1$ . When the transition probabilities are scaled in this way, The steady-state distribution  $x_{ss}$  is independent of  $\beta$ ,

$$x_{ss} = \frac{\beta p_0}{\beta(p_0 + q_0)} = \frac{p_0}{p_0 + q_0} \quad (10)$$

However, the rate of convergence still depends on  $\beta$ ,

$$1 - p - q = 1 - \beta(p_0 + q_0) = 1 - \beta \quad (11)$$

This means that  $\beta$  is a free parameter with which the convergence time can be arbitrarily varied while still satisfying the condition imposed in (9). In the most extreme case,  $\beta$  is chosen to be 1, so that  $\lambda = 0$ . In this case,  $x_t$  converges to  $p/(p+q)$  after only one time interval. Figure 4 shows  $x_t$  converging to the same steady-state behavior from the same initial conditions, for several values of  $\beta$ .

### B. Accuracy Varies Only as $x_{ss}$ and $N$ .

The other point of concern for this control system is the accuracy of the control system once it reached steady state. Specifying  $x_{ss}$  does not guarantee that the number of recruited cells  $N_{ss}^{on}$  will converge. Instead, the distribution from (2) will have some variance. When expressed as a variance normalized by the total number of units, this yields a measure of accuracy for recruitment,

$$Var(N_{ss}^{on}/N|x_{ss}) = x_{ss}(1-x_{ss})N = \frac{pq}{(p+q)^2N} \quad (12)$$

The  $\beta$  scaling argument from (10) and (11) can be applied to the variance calculation. The numerator and denominator of (12) both vary by a factor of  $\beta^2$ , so the variance is independent of the rate at which the actuator converges to its steady state probability distribution,

$$\frac{\beta^2 p_0 q_0}{\beta^2 (p_0 + q_0)^2 N} = \frac{p_0 q_0}{(p_0 + q_0)^2 N} \quad (13)$$

This is an important observation; it means that nothing is to be gained by taking ‘‘baby steps’’, that is, selecting very small values of  $p$  and  $q$  in hopes of improving the accuracy of recruitment in exchange for a slower rate of response. The only parameter that can be varied to improve accuracy is  $N$ .

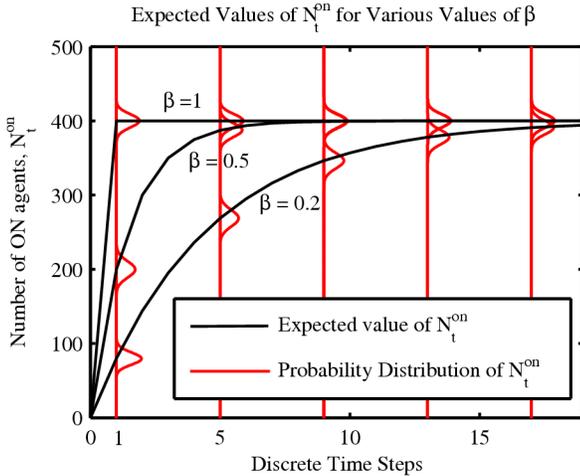


Fig. 4. The expected value of  $N_t^{on}$  and probability distribution of  $N_t^{on}$  at several points in time are shown for  $N = 500$ ,  $p_0 = 0.8$ ,  $q_0 = 0.2$ , and  $\beta = 0.2, 0.5$  and  $1$ . This plot illustrates the fact that the variance of  $N_t^{on}$  is independent of the rate of convergence. All three cases approach the same probability distribution in  $N_t^{on}$ .

### C. The Number of Transitions per Unit Time

In a physical system, there is often a significant energy cost associated with switching agents from one behavior to another. For example, a mobile robot switching between patrolling two different areas will expend energy in driving from place to place. A shape memory alloy actuator has significant latent heat associated with the phase transition used for actuation, so spurious phase transitions are costly. As a consequence, it may be useful to consider the expected number of state transitions per unit time when formulating a control policy. The expected

number of transitions can be calculated conditioned on  $x_t$ ,  $p$  and  $q$ ,

$$E(N_t^{trans}|x_t, p, q) = N(qx_t + p(1-x_t)) \quad (14)$$

In the steady state (9) can be substituted in, so that (14) is a function of  $N$ ,  $p$  and  $q$ ,

$$E(N_{ss}^{trans}|p, q) = \frac{2Npq}{p+q} \quad (15)$$

Using the scaling argument again, (15) can be rewritten in terms of  $\beta p_0$  and  $\beta q_0$ . This implies that an increase in  $\beta$  implies more expected transitions per unit time in the steady state,

$$E(N_{ss}^{trans}|\beta p_0, \beta q_0) = \frac{\beta^2 2N p_0 q_0}{\beta(p_0 + q_0)} = \beta(2N p_0 q_0) \quad (16)$$

The value of  $\beta$  minimizing the number of expected transitions is, naturally, 0, corresponding to the control policy that allows no random transitions between state.

### D. Generalization to Many States

Both closed-loop and open-loop control laws of this form are not limited to the two-state case. In general, the centrally-specified state transition probabilities for a  $k$ -state graph could be represented as a matrix  $\mathbf{M}$ ,

$$\mathbf{M} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{k1} \\ p_{12} & p_{22} & \cdots & p_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1k} & p_{2k} & \cdots & p_{kk} \end{bmatrix} \quad (17)$$

Each column of  $\mathbf{M}$  must add up to 1. These  $k^2 - k$  independent parameters could be chosen to minimize the expected error conditioned on the present knowledge for the number of agents in states  $N_{t+1}^1 \dots N_{t+1}^{k-1}$ , as was done in (8),

$$E(N_{t+1}^i|\mathbf{M}, N_t^1, \dots, N_t^k) = N^{i,ref} \quad (18)$$

Similarly, (9) can be extended to multiple states so that the steady-state expected number of agents in each state is equal to the reference,

$$E(N_{ss}^i|\mathbf{M}) = N v_1(\mathbf{M}) = N^{i,ref} \quad (19)$$

Here  $v_1(\mathbf{M})$  is the eigenvector of  $\mathbf{M}$  having eigenvalue 1. The second-largest eigenvalue will dominate the rate at which the feed-forward policy converges corresponding to the rate found in (7). Solving for these two conditions (the steady-state distribution and the rate of convergence) can be done algebraically or numerically. The other performance criteria also translate nicely to the multi-state case. The variance of the number of agents in state  $i$  is still determined only by the steady state probability of an agent being in that state,  $v_{1i}(\mathbf{M})$ , and by the total number of agents,  $N$ ,

$$Var(N_{ss}^i|\mathbf{M}) = v_{1i}(\mathbf{M})(1-v_{1i}(\mathbf{M}))N \quad (20)$$

Finally, the expected number of transitioning agents from (15) can be written for the  $k$ -state case in terms of the steady-state distribution,

$$E(N_{ss}^{trans}) = N \sum_{i=1}^k v_{1i}(1 - p_{ii}) \quad (21)$$

The trade-offs for the multi-state case are not as easy to quantify nicely. However, the general observation is that as the second-largest eigenvalue of  $\mathbf{M}$  decreases, the convergence time decreases and the expected number of transitions increases.

#### E. The One-Shot Policy

So far we have observed that it is impossible to minimize both convergence time and the number of transitions at steady state. This performance trade-offs in the no-knowledge, constant policy recruitment problem can be addressed by varying  $p$  and  $q$  with time. Suppose that we want to recruit, as accurately as possible, a specific number of agents, given no knowledge of  $N_t^{on}$  or  $x_0$ , in minimal time with minimal steady-state cost. It was demonstrated above that the accuracy of recruitment depends only on  $N$  and  $x_{ss}$ . As a consequence, the accuracy obtained by the  $\beta = 1$  policy will not improve for more than one round of state transitions. This motivates the one-shot, time-varying recruitment policy:

- 1) Compute  $p$  and  $q$  that satisfies the steady state condition from (9), and sets the second-largest eigenvalue  $\lambda$  equal to 0.
- 2) Broadcast  $p$  and  $q$  to all agents.
- 3) After one round of stochastic state transitions, broadcast a command to all agents to set  $p = 0$  and  $q = 0$ , so that all transitions cease.

This policy will guarantee that the agents get as close to the desired distribution as possible, subject to the variance of  $N_{ss}^{on}$  determined by (13).

#### IV. A MINIMAL KNOWLEDGE FEEDBACK POLICY

Often, it may be useful to consider the case in which the central controller has limited knowledge of the number of  $ON$  agents. Let  $y_t$  be a Boolean measurement which lets the controller know if the current distribution has reached the desired distribution, or is close enough. One definition of this could be when exactly the desired number of agents are in the  $ON$  state,

$$y_t = \begin{cases} true, & N_t^{on} = N^{ref} \\ false, & N_t^{on} \neq N^{ref} \end{cases} \quad (22)$$

The policy space to be searched is all policies for which a constant command  $(p, q)$  is broadcast, until the minimal feedback measurements determine that the desired state has been reached. At this point, all state transitions are commanded to cease, by setting  $p = q = 0$ . The broadcast command is assumed to be constant because, in the absence of additional information, there is no good reason for changing the command. This problem will be posed as a stochastic shortest path problem. In this framework, the cost function to be minimized

in formulating a policy is the expected time that the system takes to converge to the desired state or output. We will formulate the cost  $J$  to be minimized as a function of the initial probability that agents are  $ON$ ,  $x_0$ , given the additional knowledge that the system has not converged at time  $t = 0$ .

$$J = 1 + E \left[ \sum_{t=1}^{\infty} g(y_t) \middle| x_0 \right] \quad (23)$$

Here  $g(y_t)$  is the cost per stage of the system, equal to 1 when the system has not yet reached the target, and 0 when the system is already there:

$$g(y_t) = \begin{cases} 1, & y_t = false \\ 0, & y_t = true \end{cases} \quad (24)$$

At each point in time, Bellman's equation can be used to express the truncated cost  $J_t$  recursively forward in time,

$$J_t = E[g(y_t)|x_t] + E[J_{t+1}|x_t] \quad (25)$$

Keep in mind that the sequence  $\{x_0, x_1, x_2, \dots, x_t, \dots\}$  is determined by (5), as long as the broadcast command  $(p, q)$  is constant. The stochastic nature of the cost function arises from the uncertainty about when the exact number of desired  $ON$  agents has been reached. After this point, the cost function becomes zero. The practical result of this is that the expectation of the cost-to-go will be equal to the probability of not converging at time  $t$  multiplied by the expected cost-to-go assuming that the desired state has not yet been reached. The likelihood of reaching the desired state can be described by thinking of the number of  $ON$  agents as a Bernoulli variable (the sum of many binary random variables),

$$H_t = P(y_t|x_t) = \binom{N}{N^{ref}} x_t^{N^{ref}} (1 - x_t)^{N - N^{ref}} \quad (26)$$

Similarly, the likelihood of not reaching the desired state will be defined as  $\bar{H}_t$ ,

$$\bar{H}_t = P(\bar{y}_t|x_t) = 1 - H_t \quad (27)$$

Using this, the expectation in (25) can be evaluated,

$$J_t = \bar{H}_t + \bar{H}_t J_{t+1} = \bar{H}_t [1 + J_{t+1}] \quad (28)$$

The benefit of rewriting the series representation of  $J_t$  recursively is that each term in the series carries a multiplicative term from the previous time step,

$$J_t = \bar{H}_t + \bar{H}_t \bar{H}_{t+1} + \bar{H}_t \bar{H}_{t+1} \bar{H}_{t+2} + \dots + \prod_{k=t}^{m-1} \bar{H}_k J_m$$

The recursive form of this expression can be used to derive the optimal policy. In order to demonstrate that a policy locally minimizes this cost function, we must first find a policy for which the gradient of  $J$  with respect to the policy parameters is zero, indicating that the policy lies at an extremum of  $J$  in parameter space. As in the no-knowledge case, nothing is known about the number of  $ON$  agents until the target

distribution is achieved, so a policy with constant  $p$  and  $q$  must be pursued until  $y_t$  is true. The optimal policy to pursue is not hard to imagine; Essentially, it is to re-run the one-shot policy again and again until  $y_t$  is true. This can be proved analytically.

$$(p, q) = \left( \frac{N^{ref}}{N}, 1 - \frac{N^{ref}}{N} \right) \quad (29)$$

The first condition is that the gradient of  $J$  with respect to  $p$  and  $q$  is zero, for any initial probability  $x_0$ .  $J$  can be expanded recursively forward as a recursive series from any point in time using (28),

$$\frac{\partial J_t}{\partial p} = \frac{\partial \bar{H}_t}{\partial p} [1 + J_{t+1}] + \bar{H}_t \frac{\partial J_{t+1}}{\partial p} \quad (30)$$

The sign of each term in this series is determined by the sign of the partial derivative of  $\bar{H}_t$  at each point in time. The other terms in the expression are probabilities, which are positive, or truncated cost functions, which must also be positive. The derivative of  $H$  with respect to  $x_t$  reduces to an expression in terms of  $H_t$ ,

$$\frac{\partial H_t}{\partial x_t} = H_t \frac{N^{ref} - Nx_t}{x_t(1-x_t)} \quad (31)$$

The partial derivative of  $\bar{H}_t$  with respect to  $p$  is:

$$\frac{\partial \bar{H}_t}{\partial p} = H_t \frac{Nx_t - N^{ref}}{x_t(1-x_t)} \frac{\partial x_t}{\partial p} \quad (32)$$

One multiplicative term within this expression,  $Nx_t - N^{ref}$ , is particularly interesting. Using the allegedly optimal policy from (29), the value of  $x_t$  can be found for all  $t$  using (5):

$$x_t = \begin{cases} x_0, & t = 0 \\ N^{ref}/N, & t > 0 \end{cases} \quad (33)$$

For all  $t > 0$ ,  $Nx_t - N^{ref}$  is equal to  $N \cdot N^{ref}/N - N^{ref} = 0$ . Consequently  $\partial \bar{H}_t / \partial p = 0$  for  $t > 0$ , and by extension every term in the series defining  $\partial J_t / \partial p$ . Furthermore, this line of reasoning also works to show that  $\partial J_t / \partial q = 0$ , because each term of the series defining it will also contain a factor of  $N^{ref} - Nx_t$ . The task remaining is to show that  $J$  is increasing as the policy parameters deviate from the optimal policy. It will not be shown here, but it is straightforward to demonstrate that  $J$  increases with even infinitesimal perturbations in  $p_0$ . As Figure 5, shows, however, the increase in  $J$  as  $\beta$  varies is very gradual, and second derivative tests do not suffice. One approach is to show that the gradient of  $J$  at some finite distance away from the critical point is always pointing away from the optimal policy. This is shown by making sure that the inner product between the vector distance to the optimal policy and the gradient at that point is positive:

$$(\Delta p, \Delta q) \cdot \nabla J_t = Z_t = \frac{H_t[1 + J_{t+1}]}{x_t(1-x_t)} G_t + \bar{H}_t z_{t+1} \quad (34)$$

The sub-expression  $G_t$  determines the sign of each term in the series,

$$G_t = (Nx_t - N^{ref}) \left( \Delta p \frac{\partial x_t}{\partial p} + \Delta q \frac{\partial x_t}{\partial q} \right) \quad (35)$$

As above, the only terms in the series defining the inner product which can be negative are within  $G_t$ .  $G_t$  can be evaluated for  $p = \beta N^{ref}/N$  and  $q = \beta(1 - N^{ref}/N)$ , to obtain the expression:

$$G_t = Nt(x_0 - p_0)^2(1 - \beta)^{2t} \quad (36)$$

This is greater than zero for  $\beta \neq 1$ , so the cost function is also increasing in that direction.  $G_t$  can be shown to be positive for any small perturbation in  $p$  and  $q$ , but the expressions involved are lengthy and will not be repeated here.

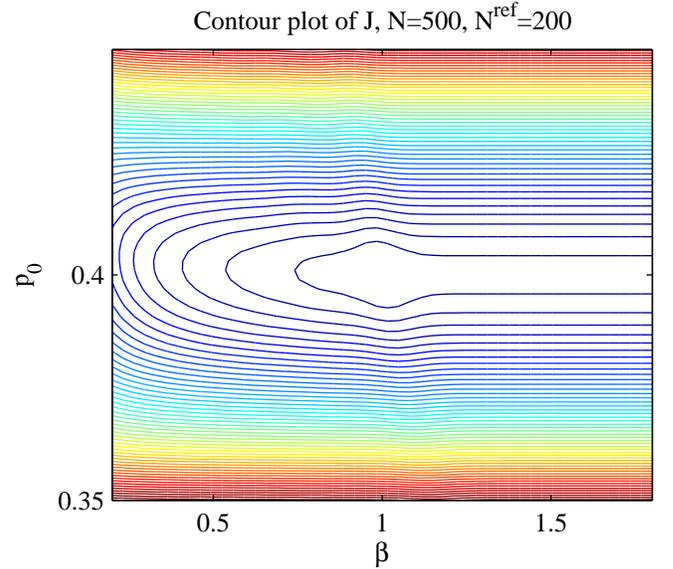


Fig. 5. The contour lines of the cost function for  $N = 500$ ,  $N^{ref} = 200$  show the shallowness of the minimum with respect to the scaling factor  $\beta$ . The cost steeply increases as  $p_0$  is varied for a fixed  $N^{ref}$ . In contrast, changes in  $\beta$  result in only modest increases in  $J$ .

## V. COMPUTATIONAL RESULTS

A swarm of 500 two-state agents was simulated with the goal of recruiting 200 agents into the  $ON$  state. The results illustrate the behaviors described above. Simple feed-forward control policies were computed, one setting  $\beta = 0.1$  and the other  $\beta = 1$ . As predicted, the number of  $ON$  agents approaches 200 in both cases. The recruitment behavior of the  $\beta = 0.1$  policy seems less random; however, this is the result of the auto-covariance of the system, not a difference in the actual variance of  $N_t^{on}$ . The cumulative distribution of values of  $N_t^{on}$  in the steady-state regime of both simulations shows that the overall distribution of both processes is nearly identical, as shown in Fig. 7 using the cumulative distribution function of  $N_t^{on}$  for each policy. However, the number of state transitions per unit time, shown in red on Figure 6, is much lower for the  $\beta = 0.1$  policy, as predicted by (15).

The one-shot policy was simulated using the same reference as the constant policies,  $N^{ref} = 200$ . Figure 8 shows the result of 20 simulations. This control policy produces constant

steady-state errors, unlike the constant policies. However, the distribution of these errors is exactly equal to those plotted in Fig. 7. Figure 9 demonstrates this using the result of 1000 simulations.

The minimal feedback policy introduced in Section IV was simulated for  $N^{ref} = 200$ , showing the typical convergence behavior. Figure 10 shows a simulated result, as well as the expected convergence time to the desired distribution, 27.5 time intervals. The authors' previous work found that the expected convergence time under these same conditions, but with full knowledge of  $N_t^{on}$  was about 4.5 time intervals [5]. This gives a sense scale when considering the value of state information for feedback in recruitment problems.

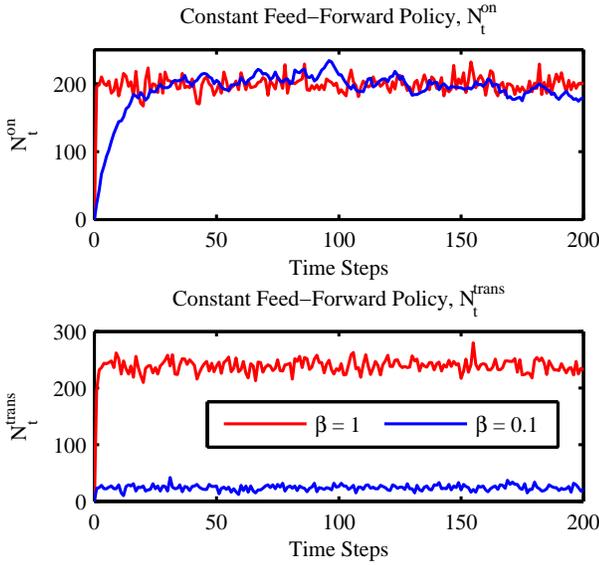


Fig. 6. A simulation of the constant policy recruitment algorithm. The number of ON cells for  $N^{ref} = 200$  is plotted (blue) for  $\beta = 0.1$  and  $\beta = 1$ , with the number of transitions per unit time (red). As predicted, reducing  $\beta$  by a factor of 10 reduces the number of transitions per unit time by a factor of 10.

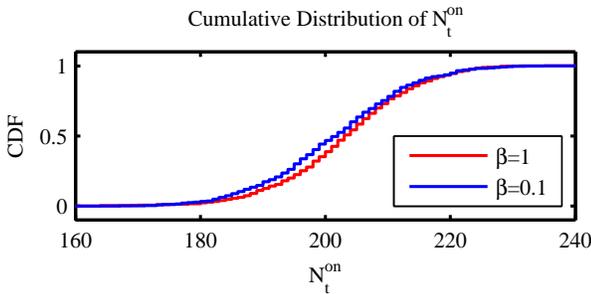


Fig. 7. The cumulative distribution of  $N_t^{on}$  in steady state, for the two simulations shown in Fig. 6. Despite the difference in convergence rate and smoothness of the time evolution, any value of  $\beta$  produces the same cumulative distribution.

## VI. CONCLUSION

This paper has demonstrated that “recruiting” finite state agents into specified states in specified numbers is possible

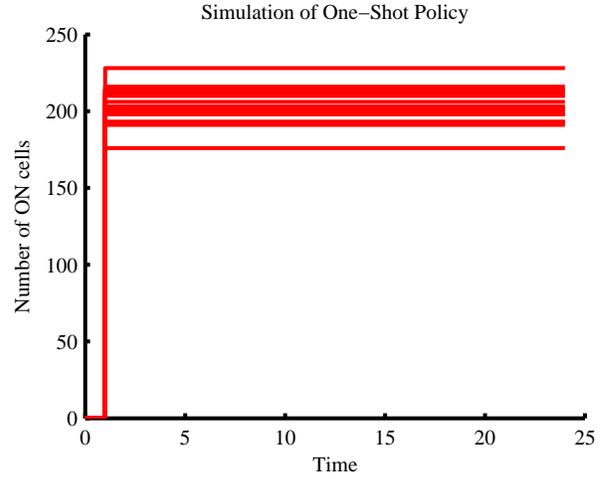


Fig. 8. Twenty simulation runs of the one-shot policy. This plot illustrates the trade-off made in this controller. In exchange for a steady-state error having the same variance as the constant policy, the agents make no state transitions at steady state.

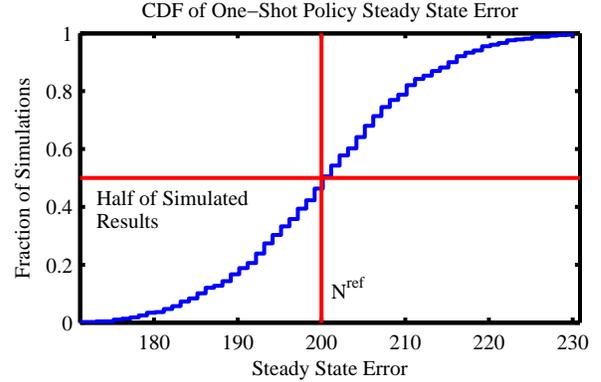


Fig. 9. A cumulative distribution function for the steady-state errors in the one-shot distribution. This distribution will approach exactly the same binomial limit distribution as the constant control policies.

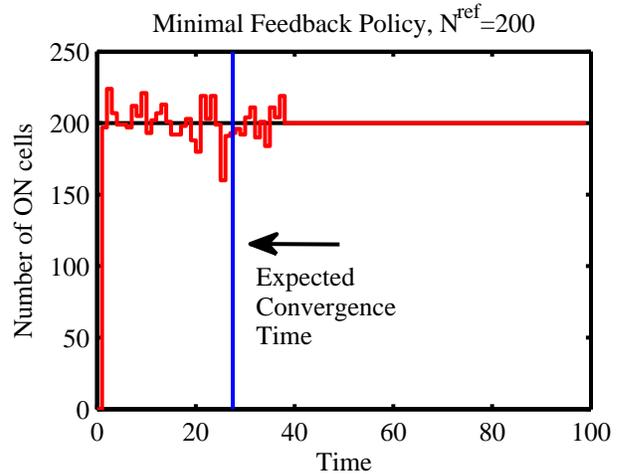


Fig. 10. The minimal feedback policy consists of using the  $\beta = 1$  constant policy until some measurement confirms that the desired number of agents are recruited; The central controller then commands all agents to cease any state transitions.

using stable, feed-forward control policies as long as the controller can find inputs that satisfy a few eigenvector and eigenvalue constraints on the state evolution equation. Unlike the previously described closed-loop policies, these feed-forward control laws cannot cause the system to converge. However, the random distribution of agents among states will be tightly grouped about the desired distribution with a well-characterized variance. Because the variance is well-characterized and relatively independent of the other performance measures for the system, it is straightforward to assess whether a closed-loop law or a feed-forward law provides greater accuracy simply by comparing the variance in the response of the closed-loop policy with uncertain information to the variance of the feed-forward law. This analysis is restricted to very limited knowledge of the state of agents in the swarm; Past work on recruitment policies having full knowledge of  $N_t^{on}$  has yielded good results in finding numerically optimal control policies. Finding policies for systems in which partial knowledge is obtainable in the form of an estimating filter, or for systems with a limited range of control inputs both remain interesting future directions.

#### REFERENCES

- [1] N. Mantzaris, S. Webb, and H. Othmer, "Mathematical modeling of tumor-induced angiogenesis", *Mathematical Biology*, Vol. 49, pp.111-187, 2004
- [2] S. McDougall, R.A. Anderson, and M. Chaplain, "Mathematical modeling of dynamic adaptive tumor-induced angiogenesis: Clinical implications and therapeutic targeting strategies", *Journal of Theoretical Biology*, Elsevier, 2006.
- [3] J. Ueda, L. Odhner, S. Kim, H. Asada, "Distributed Stochastic Control of MEMS-PZT Cellular Actuators with Broadcast Feedback," in *2006 IEEE-RAS International Conference on Biomedical Robotics and Biomechatronics*, 20-22 February 2006, p. 272-277.
- [4] L.Odhner, J. Ueda, H. Asada, "Feedback Control of Stochastic Cellular Actuators," in *The 10th International Symposium on Experimental Robotics*, 2006, p. 481-490.
- [5] L. Odhner, J. Ueda, H. Asada, "Stochastic Optimal Control Laws for Cellular Artificial Muscles," in *2007 International Conference on Robotics and Automation*, 1-14 April 2007, p. 1554-1559.
- [6] J. Ueda, L. Odhner, H. Asada, "Broadcast Feedback of Stochastic Cellular Actuators Inspired by Biological Muscle Control", *International J. of Robotics Research*, v.26, n.11-12, pp.1251-1265
- [7] Zajac, Felix, "Muscle and Tendon: Properties, Models, Scaling and Application to Biomechanics and Motor Control", *Crit. Reviews in Biomedical Engineering*, v.17, n.4, 1989, pp. 359-411
- [8] E. Henneman, G. Somjen, and D.O. Carpenter, "Functional significance of cell size in spinal motoneurons," *J Neurophysiol*, 1965, v. 28, p. 560-580.
- [9] A. Julius, A. Halasz, V. Kumar, G. Pappas, "Controlling biological systems: the lactose regulation system of *Escherichia coli*," in *2007 American Control Conference*, 9-13 July 2007 p. 1305-1310.
- [10] D. Gillespie, "Exact Stochastic Simulation of Coupled Chemical Reactions," *The Journal of Physical Chemistry*, Vol, 8 1, No. 25, 1977
- [11] S. Berman, A. Halasz, V. Kumar, S. Pratt, "Bio-Inspired Group Behaviors for the Deployment of a Swarm of Robots to Multiple Destinations," in *2007 International Conference on Robotics and Automation*, 1-14 April 2007, p. 2318-2323.