Sensing-Constrained Power Control in Digital Health

Daniel Miller*, Zhengyuan Zhou*, Nicholas Bambos* and Irad Ben-Gal†
* Department of Electrical Engineering
Stanford University, Stanford, CA 94305
Email: {danielrm, zyzhou, bambos}@stanford.edu
†Department of Industrial Engineering
Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel
Email: bengal@eng.tau.ac.il

***Abstract*—This paper considers the problem of transmitting data packets from sensors recording health information in a wireless body area network (WBAN). We model a resource-constrained paradigm in which measuring the channel interference incurs a non-trivial power cost. Each data packet has a deadline before the health information it contains is no longer relevant and the packet is dropped from the queue. In each time time slot the sensor must arbitrate whether to measure the channel interference and incur the associated cost, or rely on a previous measurement that may be inaccurate. Once the interference estimate is determined, the sensor may attempt to transmit a health packet to the central controller. The transmission power is chosen to minimize the costs associated with the power consumption, the cost of packet delays in the backlog buffer, and the cost of information loss from dropping packets. We formulate the dual channel-measurement/transmission-power control problem, and prove a performance bound on the overall expected cost incurred from using aging channel measurements.**

I. INTRODUCTION

In this paper, we consider a power control model for streaming data packets from personal health sensors in a wireless body area network (WBAN), accounting for the cost to measure the channel interference. Each sensor's transmission buffer begins with an initial backlog of B packets to be transmitted sequentially. Each head-of-line (HOL) packet is given a relevant lifespan D, before it violates the deadline and is dropped. The cost of losing potentially vital health data is dictated by the specific sensor type, frequency of operation, information type, and redundancy with other sensors. Once the HOL packet is either successfully transmitted or is dropped, the subsequent packet is moved to the head-of-line and attempts to transmit before its deadline is violated.

The transmitter must make two decisions to balance the varying costs of the system, based on the time-sensitivity and importance of its buffered data. First, the transmitter controls each packet's transmission power to regulate the probability of a successful packet communication. Increasing the transmission power improves the chances of the packet being successfully received, but incurs a corresponding power cost. This is also confounded by the channel interference, which increasingly reduces the chance of a successful packet reception. The effects of a known interference may be directly counteracted by increasing transmission power. Many wireless power control models assume that the channel interference may be measured freely, and is a known quantity at any given time. However, in general, determining this interference requires the dedication of both hardware resources and computation time, both of which are non-trivial in our resource-constrained paradigm. We therefore consider a non-zero *cost-to-sense*, and must decide whether it is advantageous to measure the current interference, or avoid this cost by recycling an older measurement of decaying accuracy. The optimal control scheme for this system balances the total power consumed by packet transmission, the cost of dropping any expired packets or delaying packets containing time-sensitive health information, and the cost to measure channel interference.

Optimal control of WBANs is fundamentally a resource-constrained problem, rising from the battery limitations in very small, distributed sensor nodes [1], [2]. Frequent battery charge requirements have been identified as an issue limiting patient compliance [3], and strategies for maximizing battery life in a WBAN have been well-examined [4]. This constraint is shared by other well-studied applications like wireless media streaming to mobile devices, and an extremely wide variety of other applications [5]. One common goal is to simply maximize the communication rate of the system [6]. Other approaches involve formulations to minimize power consumption under a direct constraint to overcome interference effects [7]. However, care must be taken to ensure that the communication requirements of any centralized power control algorithm do not exceed its benefits [8]. This issue motivates the study of distributed control schema, where each node determines its optimal control using local information.

Previous work has examined optimal power control for wireless packet transmission in a distributed setting [9], [10], [11]. An alternative hybrid approach is to consider a single uniform global stress signal [12]. Prior papers examine the optimal power control problem without channel interference, and establish properties of monotonicity and approximate control which may be extended to the fixed-rate interference-sensing model in this paper [13]. Alternate approaches involving the receiver's control have also been examined, as in adaptive media playback over a noisy channel unobserved by the controller [14]. More general work has covered the optimal sensing problem in a general noisy Markov-chain [15] and partially-observable Markov decision process (POMDP) frameworks [16]. Dynamic Programming is the dominant method in solving such general problems when restrictive assumptions of linearity or convexity are not required. As such, these models are often limited by complexity, motivating the study of near-optimal, and more practical "sufficient" methods of approximate control or those based on heuristics rather than explicitly optimal formulations [17], [18], [19], [20]. The advantages of and motivations for wireless sensors in digital health monitoring have been well-established [21]. The majority of work in the field focuses primarily on practical implementation [22], [23].

The primary contribution of this paper is the combined power and channel sensing control model. This allows for simultaneous optimization between the potential health risks resulting from undelivered health data from each individual sensor, and the power costs incurred by optimal packet transmission in a dynamic channel shared by a distributed network of similar health sensors. In this paper, we use the wireless packet streaming model to build an optimal power control scheme, with particular attention to the non-trivial costs associated with measuring channel interference. We prove that for an optimal control scheme which assumes the channel interference is known, the performance of that control scheme in a system with an aging interference measurement is bounded by the volatility of the interference.

II. TRANSMISSION MODEL

We consider a WBAN with sensors in a distributed control paradigm, such that each sensor must determine its optimal transmission and interference sensing independently of the others, with no central controller or global control signal. The sensors transmit to a central coordinating device like a smartphone, or directly to the internet. We assume that the sensors' performances are independent, conditioned on the specific patient health state. Due to the self-contained control paradigm, it is sufficient to derive the optimal control of a single health sensor in an arbitrary environment.

Our model is constructed as a discrete-time, no-arrival system in which the sensor's buffer is initially pre-loaded with a backlog of data packets $b = 1, 2, ..., B$. In each time step, the transmitter experiences a channel interference $i \in \mathcal{I}$, and must choose a transmission power $p \in \mathcal{P}$. When a packet is transmitted with power p and subject to interference i, it is successfully received by the central receiver with probability $s(p, i)$, which is increasing in p, and decreasing in i.

Each packet is assigned an initial deadline $\bar{D} \geq 1$ before the health information it contains expires. We define $D(b, t)$ as the deadline remaining when packet b reaches the HOL at time t. In other words, $D(\cdot)$ is the number of transmission attempts allowed before the packet violates the deadline. If any packet deadline expires before it has been successfully transferred, it is dropped from the queue. Otherwise, if the deadline has not yet passed, the transmitter attempts to send the packet. If the packet is not successfully received, the receiver requests re-transmission via a low-cost backchannel without interference, and the deadline is decremented. Upon a successful transmission, the next-in-line packet is shifted to the head and its remaining deadline is evaluated.

The optimality of the system is defined in terms of the costs incurred in each time slot. Each packet transmission draws a power cost of $C_p(p)$, increasing in p. A non-empty buffer draws a backlog cost of $C_b(b)$, increasing in $b > 0$. Dropping a data packet due to an exceeded deadline draws a fixed cost of C_d per dropped packet. Finally, we consider an interference measurement to draw a cost of C_m. This measurement cost may be motivated directly by the required power and computation resources, or indirectly by any transmission delays forced by prioritizing the measurement over packet transmission. The exact form will depend on the physical implementation of a sensor, for instance whether the radio multiplexes sensing and transmission on a single antenna. An optimal power control policy μ_p returns the transmission power p to be chosen in any state in order to minimize the expected cumulative cost which will be incurred going forward from the current state.

III. FIXED-SENSING OPTIMAL POWER CONTROL

We next consider the evolution of the system over discrete time slots t. The state X_t is defined by the backlog b_t, deadline d_t, and interference measurement $\hat{i}_t$. In each time slot, we control the system by choosing the transmission power p_t. A decision to postpone packet transmission would be modeled as $p_t = 0$, rather than an independent decision. We first consider the case where the interference level is measured in each time slot, such that $\hat{i}_{t+1} = i_t \; \forall \, t$. Let S_t be a binary random variable denoting a successful transmission. S_t is primarily affected by the transmission power and the channel interference. We accordingly define $s(p, i)$ increasing in p and decreasing in i, so that $\mathbb{P}\{S_t = 1\} = s(p_t, \hat{i}_t)$, and $\mathbb{P}\{S_t = 0\} = 1 - s(p_t, \hat{i}_t)$.

The following state transitions define the system evolution:

1) $\underline{b_t = 0}$: For an empty buffer, we take no action and incur no cost. For the no-arrival buffer drain model, this denotes the terminal state. If we consider packet arrivals, this state could allow interference measurement.
2) $\underline{b_t > 0,\ S_t = 1}$: If the HOL packet is successfully transmitted, we shift the queue to the next packet and assign it a new deadline. For any non-terminal state, we measure the new interference level.

$$X_{t+1} = (b_{t+1}, d_{t+1}, \hat{i}_{t+1}) = (b_t - 1, D(b_t - 1), i_t)$$

3) $\underline{b_t > 0,\ S_t = 0,\ d_t > 1}$: If the HOL packet transmission fails, but has not yet exceeded its deadline, then we retry in the following time step with a decremented deadline.

$$X_{t+1} = (b_{t+1}, d_{t+1}, \hat{i}_{t+1}) = (b_t, d_t - 1, i_t)$$

4) $\underline{b_t > 0,\ S_t = 0,\ d_t = 1}$: Finally, if the HOL packet transmission fails, and violates its deadline, then we drop the

packet and shift the queue to the next packet as in the successful transmission case.

$$X_{t+1} = (b_{t+1}, d_{t+1}, \hat{i}_{t+1}) = (b_t - 1, D(b_t - 1), i_t)$$

To solve for the optimal power control, we formulate our problem as a Dynamic Program (DP). With the individual costs as defined above, our total cost per time slot is given by

$$g(X_t) = g(b_t, d_t, i_t, p_t) = \mathbb{1}_{\{b_t>0\}} \left(C_p(p_t) + C_b(b_t) + C_m + \mathbb{1}_{\{S_t=0 \wedge d_t=1\}} C_d\right) \quad (1)$$

where $\mathbb{1}$ denotes the standard indicator function. Let $\mathcal{J}(X)$ denote the optimal expected cost-to-go for state $X = (b, d, i)$. The recursive Bellman equation is given by

$$\begin{aligned} \mathcal{J}(X) = \min_{p \in P} \Big\{ & C_p(p) + C_b(b) + C_m \\ & + s(p,i) \mathcal{J}(b-1, D(b-1), i) \\ & + (1 - s(p,i)) \big[\mathbb{1}_{\{d>1\}} \mathcal{J}(b, d-1, i) \\ & \qquad + \mathbb{1}_{\{d=1\}} \left(C_d + \mathcal{J}(b-1, D(b-1), i) \right) \big] \Big\} \end{aligned} \quad (2)$$

for positive backlog $b \geq 1$, and for terminal state $\mathcal{J}(b = 0, d, i) = 0$. Let $\mu^*(X)$ be the optimal power control scheme that satisfies the Bellman equation.

IV. PROPERTIES OF THE OPTIMAL FIXED-SENSING POWER CONTROL

If we restrict to no interference, $i = 0$, and let the interference measurement cost $C_i = 0$, then our model reduces to the more simple version discussed in detail in [13]. In this case, it has been shown that the optimal control μ^* is monotone in d, and is the unique solution to the Bellman equation. This property allows for the application of low-complexity approximate power control methods which perform near-optimally. For the fixed-sensing model, in which the interference is measured every time step for a constant cost, it is trivial to extend the proof in [13] to demonstrate a similar monotonicity. The only distinction is that the approximate control methods must be derived independently for each interference level.

We next provide a brief overview of the decision state space with the following parameters:

- Initial buffer length: $B = 25$ packets,
- Minimum packet deadline: $D(b, t) = 10$,
- Power cost: $C_p(p) = p$,
- Backlog pressure cost: $C_b(b) = b$,
- Packet drop cost: $C_d = 1$,
- Successful transmission prob.: $s(p, i) = (1 - e^{-p})e^{-i}$

Figures 1 and 2 provide simple examples of the monotonic structure of the decision curve for the optimal power control μ. Figure 3 shows the full decision boundary surface across interference levels for μ , where the optimal control in states above the surface is to idle ($\mu^*(X) = p = 0$), and for states below the surface to transmit ($\mu^*(X) = p = 10$).

One natural extension of this model would be to allow for periodic interference measurement rather than every time step. The caveat to this method is that when the true interference drifts from an aging measurement, the probability of a successful transmission is no longer known. This may be relaxed by considering instead the conditional probability given the measurement age, which depends on the channel dynamics. This allows the measurement interval to become a control parameter, and trades of the cost of frequent measurement against the risk of making sub-optimal power decisions due to inaccurate interference estimates. In Section V, we take this idea one step further by allowing the health sensor to choose whether to measure in each time step, rather than restricting to a periodic measurement.

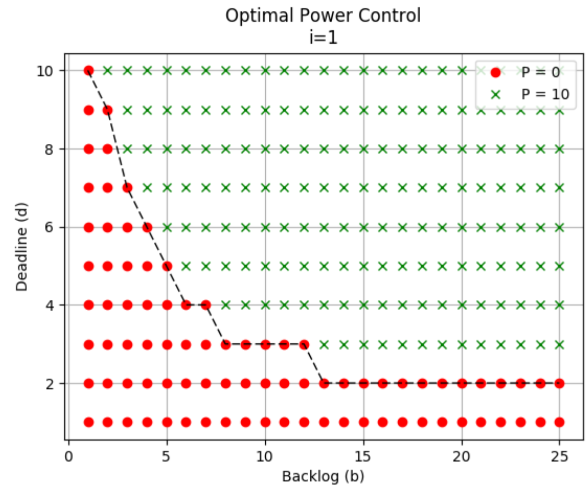

Fig. 1: Binary power control decision curve for a fixed interference level. The interference level is sensed in every time step, and the optimal power control is derived subject to the known interference.

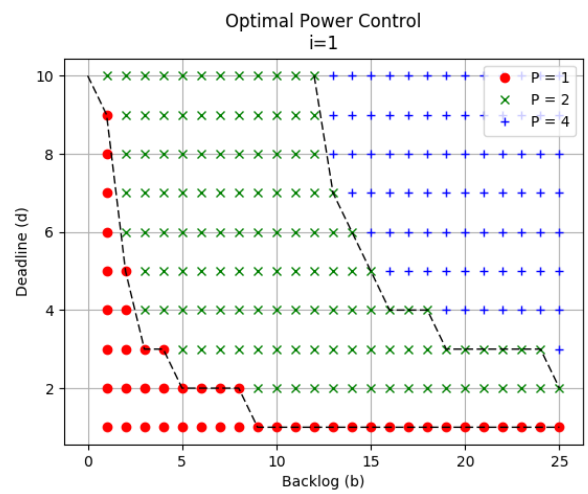

Fig. 2: Ternary power control decision curves. Interference is measured in every time step, and a wider range of allowable transmission powers is considered.

V. CONTROLLED INTERFERENCE SENSING

Let $m_t = \mathbb{1}\{\text{measure interference}\}$ denote the binary decision whether to measure the interference and incur the

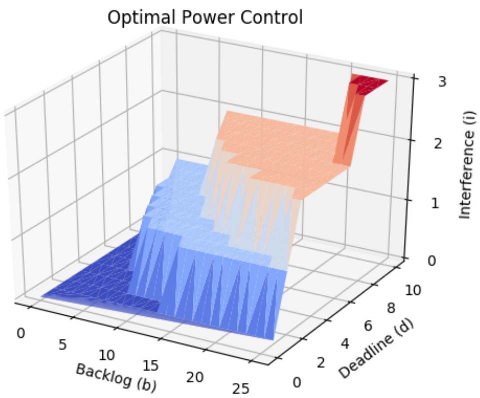

Fig. 3: Binary power control decision surface. States above the control surface represent a low power or transmitter idle decision, and states below represent a high power or active transmission decision.

corresponding cost in the current time step t if $m_t = 1$, or to retain the previous measurement if $m_t = 0$. We therefore have the evolution:

$$\hat{i}(t-1) = \begin{cases} i(t) & \text{if} \quad m_t = 1 \\ \hat{i}(t-1) & \text{if} \quad m_t = 0 \end{cases} \tag{3}$$

Let τ_t be the number of time steps since the most recent measurement, i.e. the age of the current interference estimate, such that $\tau_t = 0$ if $m_t = 1$. As described above, we extend our transmission probability function s to incorporate the age of the measurement, so that

$$s(p, i, \tau) = \mathbb{P}\{S_t = 1 \mid p_t = p, i_{t-\tau} = i\} \tag{4}$$

The extended system evolution is defined by the following state transitions from $X_t = (b_t, d_t, \hat{i}_t, \tau_t)$ to $X_{t+1} = (b_{t+1}, d_{t+1}, \hat{i}_{t+1}, \tau_{t+1})$:

1) $\underline{b_t = 0}$: As previously, for an empty buffer, we take no action and incur no cost. This denotes the terminal state.
2) $\underline{b_t > 0,\ m_t = 0,\ S_t = 1}$: In any case where we decide not to measure the current interference, the previous measurement is carried forward, and the measurement age τ is incremented. As before, if the HOL packet is successfully transmitted, we shift the queue to the next packet and assign it a new deadline.
$$X_{t+1} = (b_t - 1, D(b_t - 1), \hat{i}(t), \tau_t + 1)$$
3) $\underline{b_t > 0,\ m_t = 0,\ S_t = 0,\ d_t > 1}$: As before:
$$X_{t+1} = (b_t, d_t - 1, \hat{i}(t), \tau_t + 1)$$
4) $\underline{b_t > 0,\ m_t = 0,\ S_t = 0,\ d_t = 1}$: As before:
$$X_{t+1} = (b_t - 1, D(b_t - 1), \hat{i}(t), \tau_t + 1)$$
5) $\underline{b_t > 0,\ m_t = 1,\ S_t = 1}$: In any case where we decide to measure the current interference, we update the measurement with the current true interference, and the age of the measurement is reset to $\tau_t = 0$, so that $\tau_{t+1} = 1$. All other transitions are as before.
$$X_{t+1} = (b_t - 1, D(b_t - 1), i(t), 1)$$
6) $\underline{b_t > 0,\ m_t = 1,\ S_t = 0,\ d_t > 1}$: As before:
$$X_{t+1} = (b_t, d_t - 1, i(t), 1)$$
7) $\underline{b_t > 0,\ m_t = 1,\ S_t = 0,\ d_t = 1}$: As before:
$$X_{t+1} = (b_t - 1, D(b_t - 1), i(t), 1)$$

For a known interference level, our previous cost-per-time-slot may be simply extended with the decision of whether to measure the interference in the given time slot.

$$\begin{aligned} g(b_t, d_t, i_t, \tau_t, p_t, m_t) = \mathbb{1}_{\{b_t>0\}} \big[& \\ C_p(p_t) + C_b(b_t) + \mathbb{1}_{\{m_t=1\}} C_m + \mathbb{1}_{\{S_t=0 \wedge d=1\}} C_d \big]. & \end{aligned} \tag{5}$$

However, for a strictly positive measurement cost $C_m > 0$, such a formulation will never actually choose to measure the interference, and will instead continuously re-use the estimate from the previous time slot. To address this issue, we introduce an artificial measurement aging pressure $C_\tau(\tau)$, increasing in the age τ. This pressure is motivated by the assumption that there exists an upper bound on the rate by which the interference changes over time. In this way, a recent interference estimate is more likely to be accurate than an older estimate. The exact cost function should be determined by the statistics of the channel interference. Measurements from a volatile channel should age more quickly than those from one which remains relatively stable.

This leads to the following cost-per-time-slot:

$$\begin{aligned} g(b_t, d_t, i_t, \tau_t, p_t, m_t) = \mathbb{1}_{\{b_t>0\}} \big[C_p(p_t) + C_b(b_t) \\ + \mathbb{1}_{\{m_t=1\}} C_m + \mathbb{1}_{\{m_t=0\}} C_\tau(\tau) + \mathbb{1}_{\{S_t=0 \wedge d=1\}} C_d \big]. \end{aligned} \tag{6}$$

The optimal expected cost-to-go is given by the extended Bellman equation 7. We next explore the state space for the optimal combined power and measurement control, with parameters common to both models chosen as in Section IV, and new parameters $C_\tau(\tau) = \tau$, and $s(p, i, \tau) = (1 - e^{-p})e^{-i}$. The solution to the Bellman equation is given by the optimal power control shown in Figure 4, and the optimal channel measurement shown in Figure 5.

Although any system implementing the optimal control would not know the true interference except in time slots where it takes measurements, we wish to know how the performance varies when using old and potentially inaccurate interference measurements. We therefore establish the error metric $\hat{\epsilon}$ as the increase in the overall expected cost-to-go under the optimal control when using the estimated interference.

$$\hat{\epsilon} = |\mathcal{J}(X_0; \hat{i}) - \mathcal{J}(X_0; i)| \tag{8}$$

Theorem 1 (Bounding Theorem). *Let $h(\tau)$ be a function which is monotone increasing from $h(0) = 0$, and bounds the error in channel interference by:*

$$\left| \mathbb{P}\{i_{t+1} = i \mid \hat{i}_t, \tau_t\} - \mathbb{P}\{i_{t+1} = \hat{i}_t \mid \hat{i}_t, \tau_t\} \right| \leq h(\tau_t) \ \forall\ t \tag{9}$$

$$\begin{aligned}\mathcal{J}(b,d,\hat{i},\tau) = \min_{\substack{p\in P\\ m\in\{0,1\}}} \Big\{ & C_p(p) + C_b(b) + \mathbb{1}_{\{m=0\}}\Big[C_\tau(\tau) + s(p,\hat{i},\tau)\mathcal{J}(b-1,D(b-1),\hat{i},\tau+1) \\ & \qquad + (1-s(p,\hat{i},\tau))\big\{\mathbb{1}_{\{d>1\}}\mathcal{J}(b,d-1,\hat{i},\tau+1) \\ & \qquad\qquad + \mathbb{1}_{\{d=1\}}\big(C_d + \mathcal{J}(b-1,D(b-1),\hat{i},\tau+1)\big\}\Big] \\ & + \mathbb{1}_{\{m=1\}}\Big[C_m + s(p,\hat{i},1)\mathcal{J}(b-1,D(b-1),\hat{i},1) \\ & \qquad + (1-s(p,\hat{i},1))\big\{\mathbb{1}_{\{d>1\}}\mathcal{J}(b,d-1,\hat{i},1) \\ & \qquad\qquad + \mathbb{1}_{\{d=1\}}\big(C_d + \mathcal{J}(b-1,D(b-1),\hat{i},1)\big\}\Big]\Big\}. \end{aligned} \tag{7}$$

Then the overall error in expected cost-to-go under the optimal control strategy $\epsilon^* = |\mathcal{J}(X_0; i) - \mathcal{J}(X_0; \hat{i})|$ is bounded by

$$\epsilon^* \leq \boldsymbol{K} \cdot h(\tau^*) \tag{10}$$

for a constant $\boldsymbol{K} \in \mathbb{R}$ which depends only on the control policy, the initial state, and the cost function, and is independent of the channel interference dynamics.

Proof. *(refer to Appendix A)* □

More specifically, the bounding constant for our model is given by $\boldsymbol{K} = \bar{g}_{\mu^*}(I - \tilde{A}_{\mu^*})^{-1}e_0$, where $\bar{A}_{\mu^*}$ is the controllable transition matrix as defined in Appendix A, e_0 is unit vector of the initial state, $\bar{g}_{\mu^*}$ is the cost vector over each state subject to the optimal control strategy, and $\tau^* = \max_{t\in[0,T^*]} \tau_t$ is the maximum measurement age.

VI. EXTENSIONS AND FUTURE WORK

As with any model reliant on Dynamic Programming, the derivation of the concurrent optimal power and measurement control strategy may become intractable in large state spaces. Prior work with related models has demonstrated the viability of near-optimal approximate power control methods, which may alleviate this issue. Similar methods could be applied to the optimal channel measurement

Additionally, it may be useful to expand our initial exploration of the channel interference time-dynamics, and more directly examine its effect on the optimal control, as well as the sensitivity due to channel drift from aged measurements.

Furthermore, in small-scale systems where the wireless body sensor network is working on a very local scale, and transmitting to a central device either on-patient or nearby, the channel interference may be dominated by the other sensors in the network. In this situation, it may be possible to either control or predict the channel dynamics, and allow for more optimal performance via device cooperation, or ensure robust operation by examining competitive or adversarial situations.

VII. CONCLUSIONS

In conclusion, we have developed a theoretical framework for optimal distributed control of channel measurement and transmission power in a wireless body area network. This framework takes into account the effects of channel interference on packet transmission, and the non-trivial costs inherent in analyzing the channel. Extending prior study in wireless communication allows us to take advantage of the clinical potential of new technologies, while accounting for the inherent energy limitations of small, distributed devices. One future line of work would be considering more sophisticated queueing dynamics for packet arrival and enforce Another worthwhile line of pursuit would be to make minimal assumptions about the channel model and to learn everything only from demonstration data using an inverse reinforcement learning approach [24]. This could make the power control scheme model free.

APPENDIX A
ERROR BOUND FOR INTERFERENCE ESTIMATE

We have seen that using an aging interference estimate leads to a control strategy distinct from one which assumes a known interference. However, we can establish an upper bound on the error incurred from the inaccuracy in this measurement.

We begin by defining $\mathcal{J}_\mu^K(X_0;\theta)$ as the expected total cost incurred through all time steps through K, as follows.

$$\mathcal{J}_\mu^K(X_0;\theta) = \mathbb{E}\left\{\sum_{t=0}^{T} g(X_t;\theta)\right\} \tag{11}$$

where X_t denotes the system state (b,d,i,τ) at time t, and θ denotes the model parameters, e.g. the cost functions, and the channel interference model. As before, g is the final formulation cost-per-time-slot established in equation 6.

We next vectorize over the state space $\mathcal{X} = \{x_n\}$, and define the following vectorized variables:

- π_t: the state distribution at time t, where $[\pi_t]_n = \mathbb{P}\{X_t = x_n\}$
- $\bar{g}_\mu$: the cost of each state, where $[\bar{g}_\mu]_n = g(X_n, \mu(X_n))$
- T_μ: the stochastic transition matrix of the Markov chain resulting from applying control policy μ to the MDP system, where $[T_\mu]_{m,n} = \mathbb{P}\{X_{t+1} = x_m \mid X_t = x_n\}$, as defined by the state transitions in Section V.

We next hold a fixed control policy μ, let θ' denote a second, distinct system model with associated transition matrix T'_μ, and establish Proposition 1.

Proposition 1. *At any time step K, the difference in cost incurred by the distinct models θ and θ' is bounded by:*

$$|\mathcal{J}_\mu^K(X_0;\theta) - \mathcal{J}_\mu^K(X_0;\theta')| \leq \bar{g}_\mu \sum_{t=0}^{K} \Big|(T_\mu)^K - (T'_\mu)^K\Big| e_0 \tag{12}$$

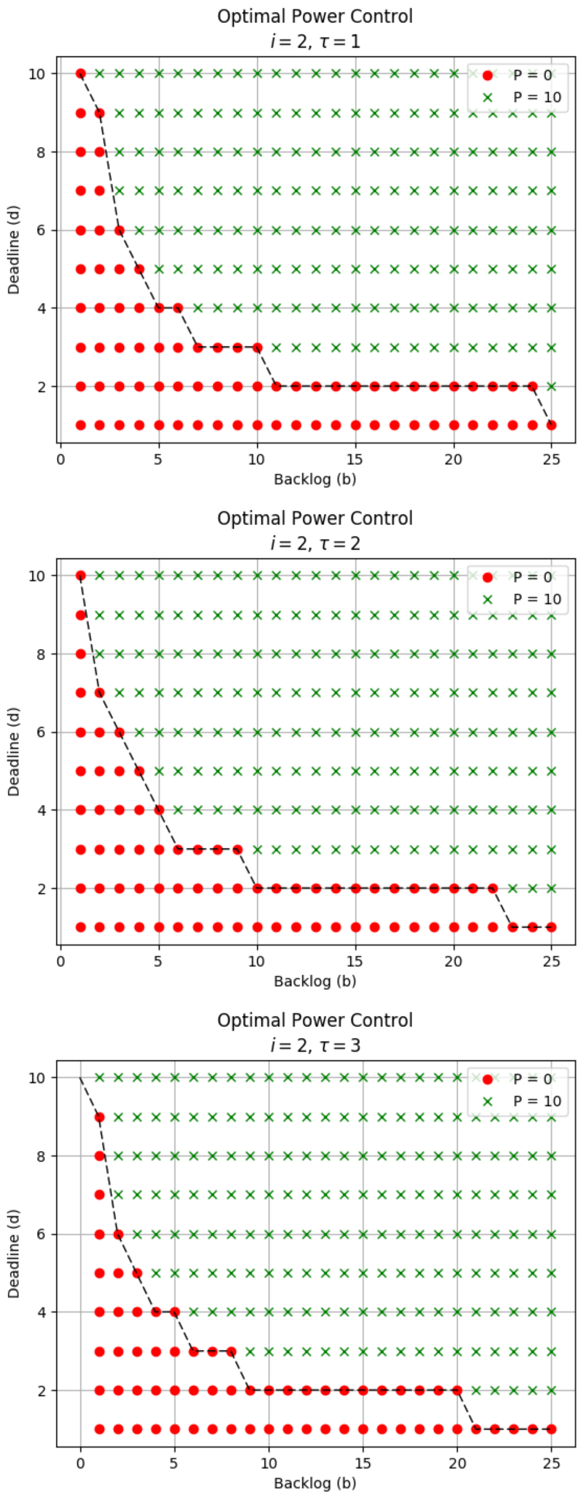

Fig. 4: Optimal power control decision curves for aging interference measurements. Given a fixed channel measurement strategy, it can be shown that the optimal power control remains monotone in d, as with the more simple models in section III and [13]. Proof omitted here for space, but similar to that of [13].

Proof. We first note that the expected cost for time slot t is given by the inner product between the cost vector and state distribution.

$$\mathbb{E}\{g(X_t, \mu)\} = \bar{g}_\mu \cdot \pi_t \tag{13}$$

Therefore, the expected cost can be rewritten as a vectorized recursion, using the state evolution $\pi_t = T_\mu \cdot \pi_{t-1} = (T_\mu)^t \cdot \pi_0$.

$$\begin{aligned} \mathcal{J}_\mu^K(X_0; \theta) &= \sum_{t=0}^{K} \bar{g}_\mu \cdot \pi_t = \bar{g}_\mu \cdot \sum_{t=0}^{K} (T_\mu)^t \cdot \pi_0 \\ &= \bar{g}_\mu (T_\mu)^K \cdot \pi_0 + \mathcal{J}_\mu^{K-1}(X_0; \theta) \end{aligned} \tag{14}$$

For a fixed initial state $e_0 = \pi_0$, where $[e_0]_n = \mathbb{1}_{\{x_n = X_0\}}$, this splits to $\mathcal{J}_\mu^K(X_0; \theta) = \bar{g}_\mu \cdot \left(\sum_{t=0}^{K} (T_\mu)^t\right) \cdot e_0$. We now consider the difference between the cost of a fixed control strategy applied to two separate models. Letting ϵ_μ^K be the accumulated error through time step K, and noting that $\epsilon_\mu^0 = 0$, we have via the triangle inequality that

$$\begin{aligned} \epsilon_\mu^K &= |\mathcal{J}_\mu^K(X_0; \theta) - \mathcal{J}_\mu^K(X_0; \theta')| \\ &= \Big| \bar{g}_\mu \left((T_\mu)^K - (T'_\mu)^K\right) e_0 \\ &\qquad + \mathcal{J}_\mu^{K-1}(X_0; \theta) - \mathcal{J}_\mu^{K-1}(X_0; \theta') \Big| \\ &\leq \left| \bar{g}_\mu \left((T_\mu)^K - (T'_\mu)^K\right) e_0 \right| \\ &\qquad + \left| \mathcal{J}_\mu^{K-1}(X_0; \theta) - \mathcal{J}_\mu^{K-1}(X_0; \theta') \right| \\ &= \bar{g}_\mu \left| (T_\mu)^K - (T'_\mu)^K \right| e_0 + \epsilon_\mu^{K-1} \end{aligned} \tag{15}$$

Unravelling the recursion yields the following:

$$\begin{aligned} \epsilon_\mu^K &\leq \sum_{t=0}^{K} \bar{g}_\mu \left| (T_\mu)^K - (T'_\mu)^K \right| e_0 \\ &= \bar{g}_\mu \left(\sum_{t=0}^{K} \left| (T_\mu)^K - (T'_\mu)^K \right| \right) e_0 \end{aligned} \tag{16}$$

□

It is worth noting that the bound in equation 16 holds with equality for the Markov chain resulting from applying a fixed control policy μ to the system, but the given formulation may be generalized to a broader setting. We now consider a useful factorization of the transition matrix, as given in Proposition 2.

Proposition 2. *Any transition matrix of our system,* $T_\mu : X_{t-1} \longrightarrow X_t$*, may be factorized as follows*

$$T_\mu = A_\mu \cdot B \tag{17}$$

where A_μ *depends only on the control policy, and is independent of the current interference, and where* B *models the interference dynamics.*

Proof. First let N denote the size of the state space $\mathcal{X} = \{x_n\}_{n=1}^N$, and let L denote the size of the interference space $\mathcal{I} = \{i_l\}_{l=1}^L$. Both N and L are assumed here to be finite for simplicity, but this restriction is not necessary in general.

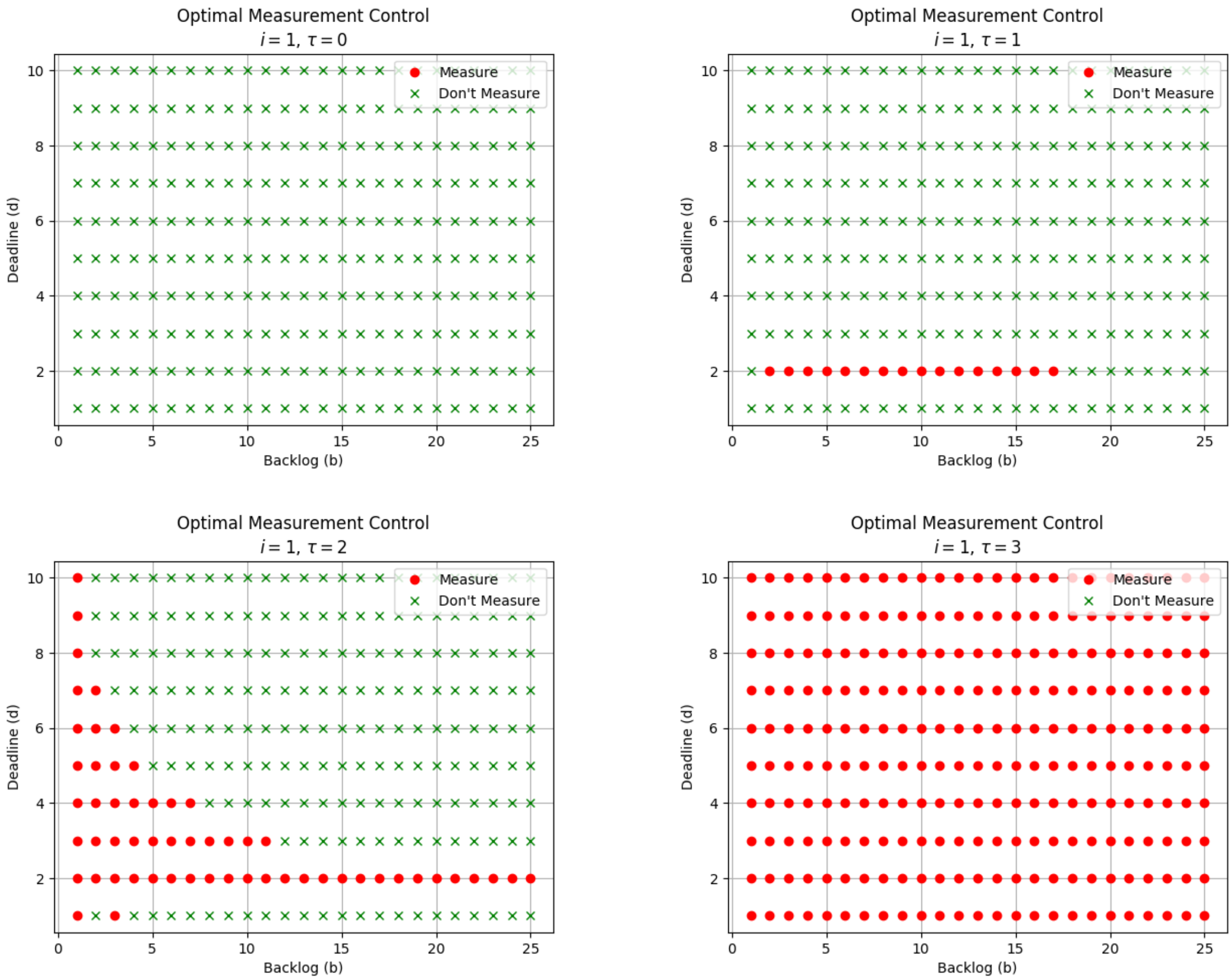

Fig. 5: Optimal measurement control decision curves for aging interference measurements. As expected, in the trivial case of $\tau = 0$, we already know the current channel interference and measurement is unnecessary. At the other extreme, once the measurement has aged beyond a certain point, the optimal decision is to take a new measurement regardless of the buffer backlog and remaining deadline for the HOL packet. In between these extremes, the decision surface takes a shape heavily dependent on the exact choice of cost functions. Unlike the optimal power control, the optimal measurement control is not monotonic with respect to d.

Then T_μ is an $N \times N$ matrix, where the (m, n)'th element is given by

$$[T_\mu]_{m,n} = \mathbb{P}\{X_t = x_m \mid X_{t-1} = x_n\} \tag{18}$$

This transition may be factored by conditioning on the true interference at time t and distributing over total probability.

$$[T_\mu]_{m,n} = \sum_{l=1}^{L} \big(\mathbb{P}\{X_T = x_m \mid X_{T-1} = x_n, i_t = i_l\} \cdot \mathbb{P}\{i_t = i_l \mid X_{T-1} = x_n\}\big) \tag{19}$$

We now define $\bar{A}_\mu$ as an $N \times N \times L$ matrix such that

$$[\bar{A}_\mu]_{m,n,l} = \mathbb{P}\{X_t = x_m \mid X_{t-1} = x_n, i_t = i_l\} \tag{20}$$

and perform a standard reshaping to an $N \times (NL)$ matrix, which we denote by A. Similarly, we define $\bar{B}$ as an $L \times N$ matrix such that

$$[\bar{B}]_{l,n} = \mathbb{P}\{i_t = i_l \mid X_{t-1} = x_n\} \tag{21}$$

and tile $\bar{B}$ vertically N times to $B = [\bar{B}; \bar{B}; ...; \bar{B}]$, so that B is an $(NL) \times N$ matrix.

Now, we have by construction that $T_\mu = A_\mu \cdot B$. □

Proof of Theorem 1. We now fix the control policy as the optimal policy using the measured interference model, $\mu = \hat{\mu}^*$. This allows a substitution with the overall expected-cost-to-go function $\mathcal{J}(X)$ as defined in equation 7.

We consider an alternate model $\bar{\theta}$ which is idential to our original model θ, except where the new channel interference

$\bar{i} = \hat{i}$. The transition matrices T_μ, T'_μ may now be similarly factored. We define A_μ as the controllable transition matrix defined in proposition 2, and similarly B as the portion of the transition matrix corresponding to the random channel, such that $T_\mu = A_\mu \cdot B$. Therefore, we consider $A_\mu = A'_\mu$, but $B \neq B'$, so that $T_\mu - T'_\mu = A_\mu(B - B')$. Since A_μ and B are non-square, T_μ^K does not factor cleanly as in the single-step case. We therefore define $A_\mu^{[k]}$ and $B^{[k]}$ as the transition matrices across k time-steps, so that

$$\begin{aligned} \left[\bar{A}_\mu^{[k]}\right]_{m,n,l} &= \mathbb{P}\{X_t = x_m \mid X_{t-k} = x_n, i_t = i_l\} \\ [\bar{B}^{[k]}]_{l,n} &= \mathbb{P}\{i_t = i_l \mid X_{t-k} = x_n\} \end{aligned} \tag{22}$$

Furthermore, we consider the bound on the change in channel interference $h(\tau)$ as defined in the Theorem 1. If we identify a maximum measurement age τ^*, either defined over the state evolution from an initial state, or as a direct constraint on the control policy, we may bound every state transition by $|B - B'|_{n,m} \leq h(\tau^*)$. Furthermore, $|B^{[k]} - B'^{[k]}|_{n,m} \leq h(\tau^*)$. Finally, we define $\tilde{A}$ as the element-wise upper limit, $\tilde{A}_\mu^{[k]} = \sup_{i \in \mathcal{I}} A_\mu^{[k]} \in \mathbb{R}^{\{N \times N\}}$. Extending Proposition 1 accordingly, and applying the factorization from Proposition 2 yields

$$\begin{aligned} \epsilon^* = \epsilon_{\mu^*}^{K^*} &= |\mathcal{J}_{\mu^*}^{K^*}(X_0;\theta) - \mathcal{J}_{\mu^*}^{K^*}(X_0;\theta')| \\ &\leq \bar{g}_{\mu^*}\left(\sum_{k=0}^{K^*}\left|(A_{\mu^*} \cdot B)^k - (A_{\mu^*} \cdot B')^k\right|\right) e_0 \\ &\leq \bar{g}_{\mu^*}\left(\sum_{k=0}^{K^*}\left|A_{\mu^*}^{[k]}\right| \cdot \left|B^{[k]} - B'^{[k]}\right|\right) e_0 \\ &\leq \bar{g}_{\mu^*}\left(\sum_{k=0}^{K^*}\tilde{A}_{\mu^*}^{[k]} \cdot h(\tau_k)\right) e_0 \\ &\leq \bar{g}_{\mu^*}\left(\sum_{k=0}^{\infty}\tilde{A}_{\mu^*}^{k}\right) e_0 \cdot h(\tau^*) \\ &= \bar{g}_{\mu^*}(I - \tilde{A}_{\mu^*})^{-1} e_0 \cdot h(\tau^*) \end{aligned} \tag{23}$$

If we bound the series above by the infinite sum, the bound holds for any time step K, and most notably for the final time step K^*, where $b_{K^*} = 0$, proving Theorem 1 as follows.

$$\epsilon^* \leq \bar{g}_{\mu^*}(I - A_{\mu^*})^{-1} e_0 \cdot h(\tau^*) \tag{24}$$

ACKNOWLEDGMENTS

Daniel Miller and Zhengyuan Zhou are supported by Stanford University through the Stanford Graduate Fellowship (SGF) in Science & Engineering.

REFERENCES

[1] C. H. Vallejos de Schatz, H. P. Medeiros, F. K. Schneider, and P. J. Abatti, "Wireless medical sensor networks: Design requirements and enabling technologies," *Telemedicine and e-Health*, vol. 18, no. 5, pp. 394–399, 2012.

[2] Z. Zhou, P. Glynn, and N. Bambos, "Repeated games for power control in wireless communications: Equilibrium and regret," in *Decision and Control (CDC), 2016 IEEE 55th Conference on*. IEEE, 2016, pp. 3603–3610.

[3] E. Jovanov, A. Milenkovic, C. Otto, P. De Groen, B. Johnson, S. Warren, and G. Taibi, "A wban system for ambulatory monitoring of physical activity and health status: applications and challenges," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*. IEEE, 2006, pp. 3810–3813.

[4] J. Y. Khan, M. R. Yuce, G. Bulger, and B. Harding, "Wireless body area network (wban) design techniques and performance evaluation," *Journal of medical systems*, vol. 36, no. 3, pp. 1441–1457, 2012.

[5] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[6] G. J. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell labs technical journal*, vol. 1, no. 2, pp. 41–59, 1996.

[7] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE Journal on selected areas in communications*, vol. 13, no. 7, pp. 1341–1347, 1995.

[8] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 4. IEEE, 2001, pp. 2033–2036.

[9] Z. Zhou and N. Bambos, "Wireless communications games in fixed and random environments," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 1637–1642.

[10] Z. Zhou, D. Miller, N. Bambos, and P. Glynn, "A stochastic stability characterization of the foschini-miljanic algorithm in random wireless networks," in *Global Communications Conference (GLOBECOM), 2016 IEEE*. IEEE, 2016, pp. 1–6.

[11] Z. Zhou, P. Mertikopoulos, A. L. Moustakas, S. Mehdian, N. Bambos, and P. W. Glynn, "Stable power control in wireless networks via dual averaging," in *GLOBECOM'17: Proceedings of the 2017 IEEE Global Communications Conference*, 2017.

[12] D. Miller, N. Master, Z. Zhou, and N. Bambos, "Scalable data center power management via a global stress signal," in *Global Communications Conference (GLOBECOM), 2015 IEEE*. IEEE, 2015, pp. 1–7.

[13] N. Master and N. Bambos, "Power control for wireless streaming with hol packet deadlines," in *Communications (ICC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 2263–2269.

[14] M. Kalman, E. Steinbach, and B. Girod, "Adaptive media playout for low-delay video streaming over error-prone channels," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 6, pp. 841–851, 2004.

[15] V. Krishnamurthy, "When to look at a noisy markov chain in sequential decision making if measurements are costly?" *arXiv preprint arXiv:1208.3291*, 2012.

[16] ——, "Structural results for partially observed markov decision processes," *arXiv preprint arXiv:1512.03873*, 2015.

[17] D. P. Bertsekas, D. P. Bertsekas, D. P. Bertsekas, and D. P. Bertsekas, *Dynamic programming and optimal control*. Athena scientific Belmont, MA, 1995, vol. 1, no. 2.

[18] W. B. Powell, *Approximate Dynamic Programming: Solving the curses of dimensionality*. John Wiley & Sons, 2007, vol. 703.

[19] L. Xia, D. Miller, Z. Zhou, and N. Bambos, "Service rate control of tandem queues with power constraints," *IEEE Transactions on Automatic Control*, vol. 62, no. 10, pp. 5111–5123, 2017.

[20] M. P. Vitus, Z. Zhou, and C. J. Tomlin, "Stochastic control with uncertain parameters via chance constrained control," *IEEE Transactions on Automatic Control*, vol. 61, no. 10, pp. 2892–2905, 2016.

[21] J. P. Lynch and K. J. Loh, "A summary review of wireless sensors and sensor networks for structural health monitoring," *Shock and Vibration Digest*, vol. 38, no. 2, pp. 91–130, 2006.

[22] A. Milenković, C. Otto, and E. Jovanov, "Wireless sensor networks for personal health monitoring: Issues and an implementation," *Computer communications*, vol. 29, no. 13, pp. 2521–2533, 2006.

[23] E. Jovanov, A. Milenkovic, C. Otto, and P. C. De Groen, "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation," *Journal of NeuroEngineering and rehabilitation*, vol. 2, no. 1, p. 6, 2005.

[24] Z. Zhou, M. Bloem, and N. Bambos, "Infinite time horizon maximum causal entropy inverse reinforcement learning," *IEEE Transactions on Automatic Control*, 2017.