Data-Driven Stochastic Models and Policies for Energy Harvesting Sensor Communications

Meng-Lin Ku, *Member, IEEE*, Yan Chen, *Senior Member, IEEE*, and K. J. Ray Liu, *Fellow, IEEE*

***Abstract*—Energy harvesting from the surroundings is a promising solution to perpetually power-up wireless sensor communications. This paper presents a data-driven approach of finding optimal transmission policies for a solar-powered sensor node that attempts to maximize net bit rates by adapting its transmission parameters, power levels and modulation types, to the changes of channel fading and battery recharge. We formulate this problem as a discounted Markov decision process (MDP) framework, whereby the energy harvesting process is stochastically quantized into several representative solar states with distinct energy arrivals and is totally driven by historical data records at a sensor node. With the observed solar irradiance at each time epoch, a mixed strategy is developed to compute the belief information of the underlying solar states for the choice of transmission parameters. In addition, a theoretical analysis is conducted for a simple on-off policy, in which a predetermined transmission parameter is utilized whenever a sensor node is active. We prove that such an optimal policy has a threshold structure with respect to battery states and evaluate the performance of an energy harvesting node by analyzing the expected net bit rate. The design framework is exemplified with real solar data records, and the results are useful in characterizing the interplay that occurs between energy harvesting and expenditure under various system configurations. Computer simulations show that the proposed policies significantly outperform other schemes with or without the knowledge of short-term energy harvesting and channel fading patterns.**

***Index Terms*—Energy harvesting, solar-powered communication, stochastic data-driven model, Markov decision process, transmission policy.**

I. Introduction

In traditional wireless sensor networks, sensor nodes are often powered by non-rechargeable batteries and distributed over a large area for data aggregation. But a major limitation of these untethered sensors is that the network lifetime is often dominated by finite battery capacity. Since the battery charge depletes with time, periodic battery replacement is required for prolonging the sensor node operations, though it becomes infeasible, costly and even impossible in some environments such as a large-scale network. As a result, there has been much research on designing efficient transmission mechanisms/protocols for saving energy in sensor communications [1].

Manuscript received April 1, 2014; revised September 15, 2014; accepted December 16, 2014. Date of publication January 13, 2015; date of current version July 14, 2015.

M.-L. Ku is with the Department of Communication Engineering, National Central University, Taoyuan City 32001, Taiwan (e-mail: mlku@ce.ncu.edu.tw).

Y. Chen and K. J. R. Liu are with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: yan@umd.edu; kjrliu@umd.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2015.2391651

Recently, energy harvesting has become an attractive alternative to circumvent this energy exhaustion problem by scavenging ambient energy sources (e.g., solar, wind, and vibration) to replenish the sensors' power supply [2]. Though an inexhaustible energy supply from the environments enables wireless sensors to function for a potentially infinite lifetime, management of the harvested energy remains a crucial issue due to the uncertainty of battery replenishment. In fact, most ambient sources occur randomly and sporadically in nature. Different sources exhibit different energy renewal processes in terms of predictability, controllability, and magnitude, which results in various design considerations.

In this paper, we focus on solar-powered wireless sensor networks, where each node is equipped with an energy harvesting device and a solar panel to collect surplus energy through the photovoltaic effect. Since the energy generation rate is uncontrollable, the energy is temporarily stored and accumulated up to a certain amount in the capacity-limited rechargeable battery for future data transmissions. But in practice, the amount of energy quanta available to a sensor could fluctuate dramatically even within a short period, and the level depends on many factors, such as the time of the day, the current weather, the seasonal weather patterns, the physical conditions of the environment, and the timescale (from seconds to days) of the energy management, to name but a few. This makes the prediction of energy harvesting very challenging, even though the solar irradiance is partially predictable with the aid of daily irradiance patterns [3]. Hence, there is a need for a stochastic energy harvesting model specific to each node, which is capable of capturing the dynamics of the solar energy associated with real data records. Besides, overly aggressive or conservative use of the harvested energy may either run out of the energy in the battery or fail to utilize the excess energy. Consequently, another essential challenge lies in adaptively tuning the transmission parameters in a smooth way that considers the randomness of energy generation and channel variation, avoids early energy depletion before the next management cycle, and maximizes certain utilities through a finite or infinite horizon of epochs.

Various energy generation models have been adopted in the literature to study the performance of solar-powered sensor networks. They can be categorized into two classes: deterministic models [4], [5] and stochastic models [3], [6]–[18]. Deterministic models, which assume that energy arrival instants and amounts are known in advance by the transmitter, were applied in [4] and [5] for designing transmission schemes. The success of the energy management in this category rests on accurate energy harvesting prediction over a somewhat long time horizon, whereas modeling mismatch occurs when the

prediction interval is enlarged. Recently attention has shifted to stochastic models by accommodating the energy management to the randomness of energy renewal processes. By assuming that energy harvested in each time slot is identically and independently distributed, the energy generation process has been described via Bernoulli models with a fixed harvesting rate [7]–[10]. Other commonly used models that are uncorrelated across time include the uniform process [3], Poisson process [11], and exponential process [12]. In [13]–[17], energy from ambient sources was modeled by a two-state Markov model to mimic the time-correlated harvesting behavior. A generalized Markov model was presented in [18] by introducing a scenario parameter, and discrete harvested energy was assumed for estimating the scenario parameter and the transition probability based on a suboptimal moving average and a Bayesian information criterion. However, there has been little research to validate the assumptions, along with exact physical interpretation, of the aforementioned stochastic models. It is essential to incorporate a data-driven stochastic model, which is capable of linking its underlying parameters to the dynamics of empirical energy harvesting data, into the design of sensor communications to develop more realistic performance characteristics.

Resource management has been studied to optimize the system utility and to harmonize the energy consumption with the battery recharge rate. The optimization of energy usage is subject to a neutral constraint which stipulates that at each time instant, the energy expenditure cannot surpass the total amount of energy harvested so far. With deterministic energy and channel profiles, a utility maximization framework was investigated in [19] to achieve smooth energy spending. The authors of [20] jointly designed power and rate adaption for maximizing data throughput, but the design is solely subject to an average power constraint. Directional water-filling was proposed in [21] for throughput maximization. A major limitation of these works is the requirement for non-causal energy arrival profiles, and they primarily focused on short-term objectives, instead of long-term objectives. Moreover, the optimization problem size grows exponentially with the scheduling interval, thereby increasing the computational burden. With stochastic models, the authors of [12] designed a threshold to decide whether to transmit or drop a message based on its importance. The outage probabilities were analyzed in fading channels by taking into account both the energy harvesting and event arrival processes [7], [22]. A simple power control policy was developed in [23] to attain near optimal throughput in a finite-horizon case. However, joint power control and adaptive modulation that maximize the bit rate have not yet been considered. Some pragmatic issues were neither addressed, e.g., the setting of stochastic models and its relation to system designs and the adaption of transmission to measured solar irradiance.

More recently, Markov decision processes (MDP) have been utilized to deal with the resource management problems for energy harvesting systems. When the battery replenishment, the wireless channel, and the packet arrival are regarded as Markov processes, sleep and wake-up strategies were developed in [6]. Similar investigations were carried out with different reward functions, e.g., buffer delay [14], [24]. Since very simple channel fluctuation and energy harvesting models were adopted, the performance may be considerably degraded in practical scenarios. In addition, the aforementioned works all prearranged stochastic energy generation models for the development of transmission mechanisms without concern for the reality of the assumptions underlying the considered models. Further, none of these works linked the solar irradiance data, gathered by an energy harvesting node, to the constructions of the design frameworks and the optimal transmission policies.

In this paper, we present data-driven transmission policies for an energy harvesting source node that aims to transmit packets to its sink over a wireless fading channel.[1] For this we maximize the long-term bit rates by adapting transmission power and modulation to the source's knowledge of its current battery and channel status. The novelty of this paper is the development of realistic and reliable energy harvesting communication, which enables a sensor node to be aware of the neighborhood environment to adapt its transmission parameters through measurement results. Specifically, the novelty and contribution are summarized as follows:

- We employ a Gaussian mixture hidden Markov model to quantify energy harvesting conditions into several representative solar states, whereby the underlying parameters enable us to effectively describe the statistical properties of the solar irradiance. Our model is different from the generalized Markov model in [18] which is constructed with discrete solar energy as its input regardless of the underlying distribution of solar energy. On the contrary, real solar irradiance is adopted in our model. We justify the validity of Gaussian mixture models for illustrating stochastic solar energy processes and use expectation-maximization (EM) algorithms to extract the underlying parameters.
- Through the discretization, a novel stochastic model that describes the generation of energy quanta is proposed and integrated into our design framework to capture the interplay between the underlying and the system parameters. The adaptive transmission is formulated as a discounted MDP and solved by a value iteration algorithm. Both the energy wastage and the throughput degradation caused by packet retransmission are taken into account. Since the exact solar state is unknown to the sensor, an observation-based mixed strategy is developed to compute the belief state information and to decide the transmission parameters, based on the present measurement of the solar irradiance. To the best of our knowledge, this is the first attempt to develop adaptive transmission schemes which are directly driven by measured data.
- To get more insight, we present a theoretical study on a simple on-off transmission policy. That means packets are transmitted at constant power and modulation levels if the action is "ON," while no transmission occurs if the action is "OFF." In this special case, there exists a threshold structure in the direction along the battery states, and the long-term expected bit rate is increased with the amount of energy quanta in the battery. Our analysis appears to be more general than the previous work [14] that simply

[1]Here, "data" means historical records or present measurement of harvested energy rather than "information-bearing data" in communications.

TABLE I
BRIEF SUMMARY OF MAJOR SYMBOLS

μ_j: Mean of solar irradiance	N_H: Number of solar states	f_D: Doppler frequency, normalized by $1/T_L$
ρ_j: Variance of solar irradiance	N_C: Number of channel states	$P_{e,b}$: Average bit error rate
a_{ij}: Solar state transition probability	N_B: Number of battery states	χ_m: Information bits per data symbol
T_L: Policy management period	N_P: Number of power actions	$P_{f,k}$: Successful packet transmission probability
P_U: Basic transmission power level	N_M: Number of modulation types	D_E: Number of effective data packets
E_U: One unit of energy quantum	$\mathcal{H}$: Set of solar states	γ_U: Average signal-to-noise power ratio
P_H: Harvested solar power	$\mathcal{C}$: Set of channel states	R_a: Reward function
Ω_S: Solar panel area	$\mathcal{B}$: Set of battery states	$\pi(s)$: Transmission policy
ϑ: Energy conversion efficiency	$\mathcal{W}$: Set of power actions	λ: Discount factor
T_P: Packet duration	$\mathcal{M}$: Set of modulation types	$V_i^a(s)$: Long-term expected reward
L_S: Number of symbols during T_P	$\mathcal{A}$: Composite set of $\mathcal{W}$ and $\mathcal{M}$	$\zeta_j^{(t)}$: Belief state information
R_S: Data symbol rate	Γ_i: Channel quantization threshold	$\kappa_{z,x}$: Threshold structure for on-off policy
D: Number of packets during T_L	γ_0: Average channel power	$R_{net,m}$: Expected net bit rate

considers an uncorrelated energy arrival model and a two-state channel model. By exploiting this structure, we provide an energy deficiency condition and an upper bound for the achievable net bit rate to characterize the performance limit. Finally, the existence of structures for a composite policy which contains multi-levels of power and modulation actions is discussed.

- Real data records of the solar irradiance measured by different solar sites in [25] are utilized to exemplify our design framework as well as performance evaluation. The performance of the proposed transmission policies is validated by extensive computer simulations and compared with other radical policies with or without the knowledge of short-range energy harvesting and channel variation patterns.

The rest of this paper is organized as follows. A brief summary of major symbols is listed in Table I. In Section II, we describe the stochastic energy harvesting model, the training of its underlying parameters, and its connection to the real data record. The MDP formulation of the adaptive transmission is presented in Section III, followed by the optimization of the policies and the mixed strategy in Section IV. Section V is devoted to the analysis of a simple on-off transmission policy. Simulation results are presented in Section VI, and concluding remarks are provided in Section VII.

II. STOCHASTIC ENERGY HARVESTING MODELS AND TRAINING

The model for describing the harvested energy depends on various parameters, such as weather conditions (e.g., sunny, cloudy, rainy), sunshine duration (e.g., day and night), and behavior of the rechargeable battery (e.g., storage capacity). We focus on modeling the solar power from the measurements by using a hidden Markov chain, and establish a framework to extract the underlying parameters that can characterize the availability of solar power.

We begin with a toy example to justify the rationality of the proposed energy harvesting models. Consider a real data record of irradiance (i.e., the intensity of the solar radiation in units μW/cm^2) for the month of June from 2008 to 2010, measured by a solar site in Elizabeth City State University (EC), with the measurements taken at five-minute intervals [25]. In Fig. 1(a), the time series of the irradiance is sketched over twenty-four hours for June 15th, 2010, along with the average results for the month of June in 2008 and 2010. We can make the following

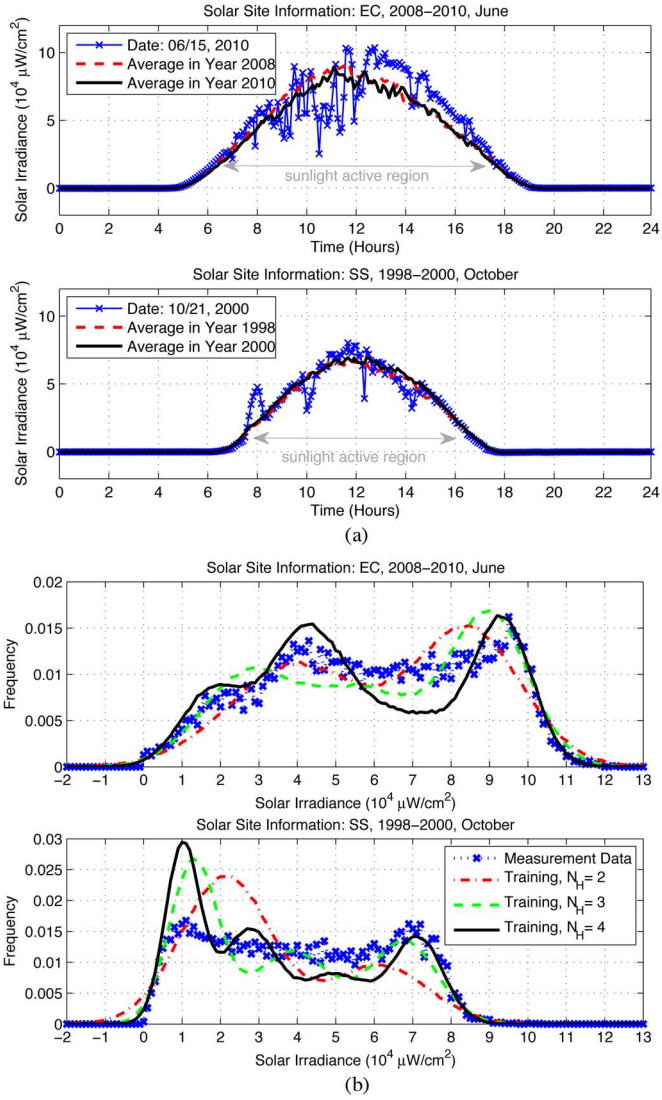

Fig. 1. Toy examples of solar irradiance measured by solar sites in Elizabeth City State University (EC) and Savannah State College (SS). (a) Time series of the daily irradiance. (b) Histogram of the irradiance during a time period of seven o'clock to seventeen o'clock.

observations. First, the daily solar radiation fluctuates slowly within a short time interval, but could suddenly change from the current level to adjacent levels with higher or lower mean values. Second, the average irradiance value is sufficiently high only from the early morning (seven o'clock) to the late afternoon (seventeen o'clock). In fact, the measured irradiance could be positive during daytime hours or negative at night or in

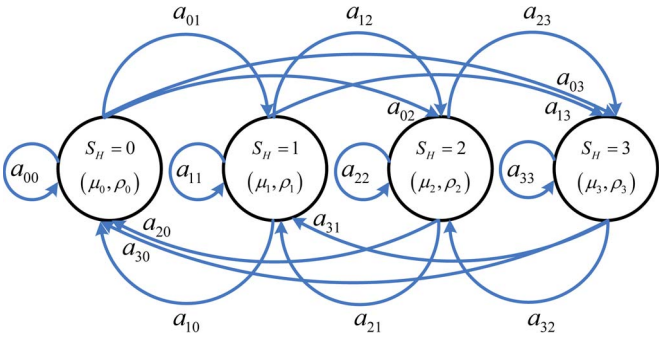

Fig. 2. Gaussian mixture hidden Markov chain of the solar power harvesting model with the underlying parameters (μ_j, ρ_j) $(N_H = 4)$.

the early morning, depending on the total amount of irradiance that comes in and goes out the solar panel. Third, the evolution of the diurnal irradiance follows a very similar time-symmetric mask, whereas the short-term profiles of different days can be very different and unpredictable. The other data record, measured by Savannah State College (SS) in October between 1998 and 2000, also exhibits the same observations, but with a shorter sunlight active duration. By considering the irradiance from seven o'clock to seventeen o'clock, Fig. 1(b) shows the corresponding histogram plotted against the irradiance on the x-axis, which represents the percentage of the occurrences of data samples in each bin of width 10^3 μW/cm^2. We see that the irradiance behaves like a mixture random variable generated by a number of distributions. The prediction of solar irradiance has been an open problem in atmospheric science over the past decades. Some research studies have suggested the use of the Gaussian distribution as the ingredient for describing the irradiance [26], [27]. The assertion stems from the fact that the solar irradiance experiences scattering, diffusion and reflection by molecules, tiny particles in the air, and obstacles (e.g., cloud and terrain) in the surrounding of sensors. Motivated by these discussions, we model the evolution of the irradiance via a hidden Markov chain with a finite number of states, each of which is specified by a normal distribution with unknown mean and variance.

An N_H-state solar power harvesting hidden Markov model is illustrated in Fig. 2, where the underlying normal distribution for the j^{th} state is specified by the parameters of the mean μ_j and the variance ρ_j. The solar irradiance can be classified into several states S_H to represent harvesting conditions such as "Excellent", "Good", "Fair", and "Poor". Without loss of generality, the solar states are numbered in ascending order of the mean values μ_j. Let $S_H^{(t)}$ be the solar state at time instant t. We further assume that the hidden Markov model is time homogeneous and governed by the state transition probability $P(S_H^{(t)} = j | S_H^{(t-1)} = i) = a_{ij}$, for $i, j = 0, \ldots, N_H - 1$. The parameters of the model are thus defined as $\mathbf{\Theta} = \{\boldsymbol{\mu}, \boldsymbol{\rho}, \mathbf{a}\}$, where $\boldsymbol{\mu} = [\mu_0, \ldots, \mu_{(N_H-1)}]^T$, $\boldsymbol{\rho} = [\rho_0, \ldots, \rho_{(N_H-1)}]^T$, and $\mathbf{a} = [a_{00}, a_{01}, \ldots, a_{(N_H-1)(N_H-1)}]^T$. Let $\mathbf{x} = \{X^{(1)} = x_1, \ldots, X^{(T)} = x_T\}$ be a sequence of observed data over a measurement period T, corresponding to a sequence of hidden states $\mathbf{s} = \{S_H^{(1)} = s_1, \ldots, S_H^{(T)} = s_T\}$. The probabilistic model is trained by an EM algorithm, which is a general method of finding the maximum-likelihood (ML) estimate for the state parameters of underlying distributions from incomplete observed data, as follows:

$$\begin{aligned}\mathbf{\Theta}^{(n)} &= \arg\max_{\mathbf{\Theta}} \mathbb{E}_{\mathbf{s}} \left[\log P(\mathbf{x}, \mathbf{s}|\mathbf{\Theta}) | \mathbf{x}, \mathbf{\Theta}^{(n-1)}\right] \\ &= \arg\max_{\mathbf{\Theta}} \sum_{\mathbf{s}} \log P(\mathbf{x}, \mathbf{s}|\mathbf{\Theta}) \cdot P\left(\mathbf{x}, \mathbf{s}|\mathbf{\Theta}^{(n-1)}\right), \quad (1)\end{aligned}$$

where $\mathbf{\Theta}^{(n)}$ is the estimation update at the n^{th} iteration. The problem (1) can be efficiently solved using the well-known iterative forward and backward algorithms, and further details can be referred to [28]. The training procedures are repeated for several iterations until $\mathbf{\Theta}^{(n)}$ gets converged.

The training results with respect to the example above are shown in Fig. 1(b) and Table II, where the measurements are performed every five or fifteen minutes from seven to seventeen o'clock. We can observe that the similarity between the histograms of the training results and the measurement data is improved as N_H is increased from two to four at the expense of the increased complexity. Our experimental experience suggests that a four-state hidden Markov model is good enough to achieve acceptable results. Also in Table II, where the data record of the solar site in EC is used, the transition probabilities from the current solar state to the other adjacent states are very small when the measurements are taken at five-minute intervals, and only a slight increase in the transition probability is observed as the sampling period is increased to fifteen minutes.

The solar power harvesting model is a continuous-time model. In practice, the solar energy is stored in the battery to supply the forthcoming communications, and the transmission strategy is designed on the basis of the required numbers of energy quanta and remains unchanged over a management period of several data packets T_L. Below, we map the continuous-time model into a discrete energy harvesting model, in which the Markov chain states are described by the numbers of harvested energy quanta. Let P_U be the basic transmission power level of sensors, corresponding to one unit of the energy quantum $E_U = P_U T_L$ during the management period. For the harvested solar power P_H, the obtained energy over T_L is given by $E_H = P_H T_L$. At $t = nT_L$, define $E_R^{(n)}$ as the residual energy in the capacitor before harvesting, and $E_C^{(n)}$ as the accumulated energy after harvesting over T_L. Accordingly, the capacitor can provide at most Q energy quanta to recharge the battery, and the remaining part, which is smaller than E_U, is regarded as the residual energy in the capacitor at $t = (n+1)T_L$:

$$E_C^{(n)} = E_R^{(n)} + E_H; \quad (2)$$

$$Q = \left\lfloor \frac{E_C^{(n)}}{E_U} \right\rfloor, \quad E_R^{(n+1)} = E_C^{(n)} - QE_U, \quad (3)$$

where $\lfloor \cdot \rfloor$ is the floor function. By assuming that the fluctuation of the harvested power level is quasi-static over many power management runs, it can be analyzed that if $qE_U \le E_H \le (q+1)E_U$ for some q, then the probability of the number of energy quanta, Q, can be computed as

$$P(Q = i) = \begin{cases} \frac{E_H - qE_U}{E_U}, & i = q+1; \\ 1 - \frac{E_H - qE_U}{E_U}, & i = q; \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

TABLE II
TRAINING RESULTS OF THE HIDDEN MARKOV SOLAR POWER HARVESTING MODEL FOR THE SOLAR SITE IN EC ($N_H = 4$). (a) MEAN, VARIANCE, AND STEADY STATE PROBABILITY. (b) STATE TRANSITION PROBABILITY

(a)

Sampling period	5 minutes				15 minutes			
State ($S_H = i$)	0	1	2	3	0	1	2	3
μ_i (10^4 μW/cm^2)	1.75	4.21	7.02	9.38	1.79	4.56	7.60	9.46
ρ_i	0.65	1.04	2.34	0.54	0.71	1.48	1.55	0.31
$P(S_H = i)$	0.16	0.36	0.21	0.27	0.16	0.39	0.27	0.18

(b)

Sampling period	5 minutes				15 minutes			
a_{ij}	$j=0$	$j=1$	$j=2$	$j=3$	$j=0$	$j=1$	$j=2$	$j=3$
$i=0$	0.979	0.015	0.006	0	0.938	0.057	0.005	0
$i=1$	0.005	0.988	0.007	0	0.023	0.955	0.022	0
$i=2$	0.006	0.009	0.975	0.010	0	0.032	0.950	0.018
$i=3$	0	0	0.007	0.993	0.004	0	0.023	0.973

When a sensor node is operated at the j^{th} solar state with the normal distribution $\mathcal{N}(x;\mu_j,\rho_j)$, the obtained energy E_H is again a normally distributed random variable, which is equal to the solar power per unit area x multiplied by the solar panel area Ω_S, the time duration T_L and the energy conversion efficiency ϑ, i.e., $E_H = x\Omega_S T_L \vartheta$. The conversion efficiency typically ranges between 15% and 20% [2]. Thus, the mean and variance of E_H are respectively given as $\bar{\mu}_j = \mu_j \Omega_S T_L \vartheta$ and $\bar{\rho}_j = \rho_j \Omega_S^2 T_L^2 \vartheta^2$, and the probability of the number of energy quanta is calculated by averaging (4) with respect to the random variable E_H, as follows:

$$P(Q=i|S_H=j) = \begin{cases} \int_{iE_U}^{(i+1)E_U} \frac{(i+1)E_U - E_H}{E_U} \mathcal{N}(E_H;\bar{\mu}_j,\bar{\rho}_j)dE_H, \ i=0; \\ \int_{iE_U}^{(i+1)E_U} \frac{(i+1)E_U - E_H}{E_U} \mathcal{N}(E_H;\bar{\mu}_j,\bar{\rho}_j)dE_H \\ +\int_{(i-1)E_U}^{iE_U} \frac{E_H - (i-1)E_U}{E_U} \mathcal{N}(E_H;\bar{\mu}_j,\bar{\rho}_j)dE_H, \ i\neq 0. \end{cases} \quad (5)$$

Denote the complementary error function as $\mathrm{erfc}(\cdot)$. After some manipulations, we get

$$P(Q=i|S_H=j) = \begin{cases} \left((i+1)-\frac{\bar{\mu}_j}{E_U}\right) g_1(i,\bar{\mu}_j,\bar{\rho}_j) - g_2(i+1,\bar{\mu}_j,\bar{\rho}_j), \ i=0; \\ \left((i+1)-\frac{\bar{\mu}_j}{E_U}\right) g_1(i,\bar{\mu}_j,\bar{\rho}_j) - g_2(i+1,\bar{\mu}_j,\bar{\rho}_j) \\ +\left(\frac{\bar{\mu}_j}{E_U}-(i-1)\right) g_1(i-1,\bar{\mu}_j,\bar{\rho}_j) + g_2(i,\bar{\mu}_j,\bar{\rho}_j), \ i\neq 0, \end{cases} \quad (6)$$

where the relevant terms are defined as

$$g_1(i,\bar{\mu}_j,\bar{\rho}_j) = \frac{1}{2}\left(\mathrm{erfc}\left(\frac{1}{\sqrt{2\bar{\rho}_j}}(iE_U - \bar{\mu}_j)\right) - \mathrm{erfc}\left(\frac{1}{\sqrt{2\bar{\rho}_j}}((i+1)E_U - \bar{\mu}_j)\right)\right); \quad (7)$$

$$g_2(i,\bar{\mu}_j,\bar{\rho}_j) = \sqrt{\frac{\bar{\rho}_j}{2\pi E_U^2}}\left(\exp\left(-\frac{1}{2\bar{\rho}_j}((i-1)E_U - \bar{\mu}_j)^2\right) - \exp\left(-\frac{1}{2\bar{\rho}_j}(iE_U - \bar{\mu}_j)^2\right)\right). \quad (8)$$

III. MARKOV DECISION PROCESS USING STOCHASTIC ENERGY HARVESTING MODELS

We study the adaptive transmissions for sensor communications concerning the channel and battery status, the transmission power, the modulation types, and the stochastic energy harvesting model. Consider a point-to-point communication link with two sensor nodes, where a source node intends to convey data packets to its sink node. Each data packet consists of L_S data symbols at a rate of R_S (symbols/sec), and hence, the packet duration is given by $T_P = L_S/R_S$.

The design framework is formulated as an MDP with the goal of maximizing the long-term net bit rate. As illustrated in Fig. 3, the MDP is mainly composed of the state space, the action set, and the state transition probabilities, and it is operated on the time scale of T_L, covering the time duration of D data packets, i.e., $T_L = DT_P$. Let $\mathcal{S}$ be the state space which is a composite space of the solar state $\mathcal{H} = \{0,\ldots,N_H-1\}$, the channel state $\mathcal{C} = \{0,\ldots,N_C-1\}$ and the battery state $\mathcal{B} = \{0,\ldots,N_B-1\}$, i.e., $\mathcal{S} = \mathcal{H}\times\mathcal{C}\times\mathcal{B}$, where $\times$ denotes the Cartesian product. At the n^{th} battery state, we further denote the action space as $\mathcal{A}$ which consists of two-tuple action spaces: transmission power $\mathcal{W} = \{0,\ldots,\min\{n,N_P-1\}\}$ and modulation types $\mathcal{M} = \{0,\ldots,N_M-1\}$. Since the transition probabilities of the channel and battery states are independent of each other, the transition probability from $(S_H,S_C,S_B) = (j,i,n)$ to $(S_H,S_C,S_B) = (j',i',n')$ with respect to the action $(W,M) = (w,m)$ under the j^{th} solar state can be formulated as

$$\begin{aligned} &P_{w,m}\left((S_H,S_C,S_B) = (j',i',n')|(S_H,S_C,S_B) = (j,i,n)\right) \\ &\quad = P(S_H = j'|S_H = j)P(S_C = i'|S_C = i) \\ &\qquad \cdot P_w\left(S_B = n'|(S_H,S_B) = (j,n)\right), \end{aligned} \quad (9)$$

where the battery state transition is irrespective of the modulation type, and the transition probability of the solar states can be directly obtained by using the training results in Section II. We elaborate on each of the components in Fig. 3 before describing the solution of the Bellman optimality equation.

A. Actions of Transmission Power and Modulation Types

When the action $(w,m) \in \mathcal{W}\times\mathcal{M}$ is chosen by the sensor node, the transmission power and modulation levels are

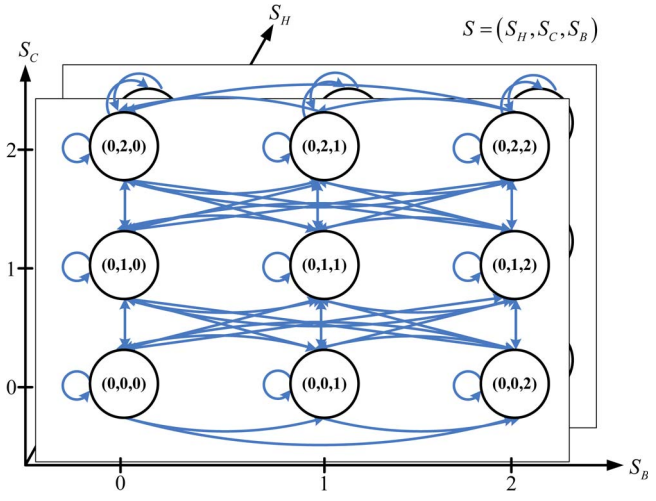

Fig. 3. Markov chain for the Markov decision process ($N_H = 2$, $N_C = 3$ and $N_B = 3$).

respectively set as wP_U and 2^{χ_m}-ary phase shift keying (PSK) or quadrature amplitude modulation (QAM), e.g., QPSK, 8 PSK and 16 QAM, during the policy management period, where χ_m represents the number of information bits in each data symbol. Remember that P_U is the basic transmission power level of the sensor node if data transmission takes place. On the other hand, if $w = 0$, the node remains silent without transmitting data packets.

B. Channel State and State Transition Probability

The wireless channel is quantized using a finite number of thresholds $\mathbf{\Gamma} = \{0 = \Gamma_0, \Gamma_1, \ldots, \Gamma_{N_C} = \infty\}$, where $\Gamma_i < \Gamma_j$ for all $i < j$. The Rayleigh fading channel is said to be in the i^{th} channel state, for $i = 0, \ldots, N_C - 1$, if the instantaneous channel power, γ, belongs to the interval $[\Gamma_i, \Gamma_{i+1})$. We assume that the wireless channel fluctuates slowly and the policy management period is shorter than the channel coherence time. Hence, the channel state transition occurs only from the current state to its neighboring states. The stationary probability of the i^{th} state is

$$P(S_C = i) = \exp\left(-\frac{\Gamma_i}{\gamma_0}\right) - \exp\left(-\frac{\Gamma_{i+1}}{\gamma_0}\right), \tag{10}$$

where $\gamma_0 = \mathbb{E}[\gamma]$ is the average channel power. Define $h(\gamma) = \sqrt{2\pi\gamma/\gamma_0} f_D \exp(-\gamma/\gamma_0)$, where f_D is the maximum Doppler frequency, normalized by $1/T_L$. The state transition probabilities are determined by [29]

$$P(S_C = k|S_C = i) = \begin{cases} \frac{h(\Gamma_{i+1})}{P(S_C=i)}, & k = i+1, \quad i = 0, \ldots, N_C - 2; \\ \frac{h(\Gamma_i)}{P(S_C=i)}, & k = i-1, \quad i = 1, \ldots, N_C - 1; \\ 1 - \frac{h(\Gamma_i)}{P(S_C=i)} - \frac{h(\Gamma_{i+1})}{P(S_C=i)}, & k = i, \quad i = 1, \ldots, N_C - 2, \end{cases} \tag{11}$$

and the transition probabilities of $P(S_C = i|S_C = i)$ for the boundaries are given by

$$\begin{aligned} &P(S_C = 0|S_C = 0) = 1 - P(S_C = 1|S_C = 0); \\ &P(S_C = N_C - 1|S_C = N_C - 1) \\ &\quad = 1 - P(S_C = N_C - 2|S_C = N_C - 1). \end{aligned} \tag{12}$$

C. Battery State and State Transition Probability

Consider a rechargeable battery with finite capacity which is described by N_B states. When the sensor node is run at the n^{th} battery state, the available energy in the battery is stored up to n energy quanta, i.e., nE_U, and the possible action that can be performed is from 0 to $\min\{n, N_P - 1\}$. The w^{th} power action will consume a total of w energy quanta for data transmission. In particular, the sensor is unable to make any transmission when the energy is completely depleted at the 0^{th} state. Once the underlying parameters of the N_H solar states are appropriately estimated through the measurement data, the state transition probabilities for the n^{th} battery state and the w^{th} power action under the j^{th} solar state can be constructed by exploiting (6), as follows:

$$P_w(S_B = k|(S_H, S_B) = (j, n)) = \begin{cases} P(Q{=}k{-}n{+}w|S_H{=}j), & k{=}n{-}w, \ldots, N_B{-}2; \\ 1 - \sum_{i=0}^{N_B-2-n+w} P(Q{=}i|S_H{=}j), & k{=}N_B{-}1, \end{cases} \tag{13}$$

for $n = 0, \ldots, N_B - 1$ and $w = 0, \ldots, \min\{n, N_P - 1\}$.

D. Reward Function

We adopt the average number of good bits per packet transmission as our reward function. It is assumed that the sink node periodically feeds back the channel state information to the source node for planning the next transmission. Let $P_{e,b}((S_C, S_B, W, M) = (i, n, w, m))$ be the average bit error rate (BER) at the i^{th} channel state and the n^{th} battery state when the action $(W, M) = (w, m)$ is taken. By applying the upper bound of the Q-function $Q(x) \le \frac{1}{2}\exp\left(-\frac{x^2}{2}\right)$, it can be computed as

$$\begin{aligned} &P_{e,b}\left((S_C, S_B, W, M) = (i, n, w, m)\right) \\ &\quad = \frac{\int_{\Gamma_i}^{\Gamma_{i+1}} \sum_r \alpha_{m,r} Q\left(\sqrt{\frac{\beta_{m,r} w P_U \gamma}{N_0}}\right) \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right) d\gamma}{\int_{\Gamma_i}^{\Gamma_{i+1}} \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right) d\gamma} \\ &\quad \le \sum_r \frac{\frac{\alpha_{m,r}}{w\beta_{m,r}\gamma_U + 2}}{\exp\left(-\frac{\Gamma_i}{\gamma_0}\right) - \exp\left(-\frac{\Gamma_{i+1}}{\gamma_0}\right)} \\ &\qquad \cdot \left(\exp\left(-\frac{1}{2\gamma_0}(w\beta_{m,r}\gamma_U + 2)\Gamma_i\right)\right. \\ &\qquad \left. - \exp\left(-\frac{1}{2\gamma_0}(w\beta_{m,r}\gamma_U + 2)\Gamma_{i+1}\right)\right) \\ &\quad \triangleq \eta(i, n, w, m), \end{aligned} \tag{14}$$

where N_0 is the noise power, $\gamma_U = P_U\gamma_0/N_0$ is the average signal-to-noise power ratio (SNR) when the basic transmission power level is adopted, and the BER is expressed as a summation of Q-functions with modulation specific constants $\alpha_{m,r}$ and $\beta_{m,r}$ for QPSK, 8 PSK and 16 QAM in Table III [30], [31]. Hence, the probability of successful packet transmission

TABLE III
MODULATION SPECIFIC CONSTANTS

Modulation schemes	Parameters $(\alpha_{m,r},\ \beta_{m,r})$
QPSK	$(\alpha_{m,0},\ \beta_{m,0})= (1, 1)$
8 PSK	$(\alpha_{m,0},\ \beta_{m,0})= (\frac{2}{3}, 2\sin^2(\frac{\pi}{8}))$, $(\alpha_{m,1},\ \beta_{m,1})= (\frac{2}{3}, 2\sin^2(\frac{3\pi}{8}))$
16 QAM	$(\alpha_{m,0},\ \beta_{m,0})= (\frac{3}{4}, \frac{1}{5})$, $(\alpha_{m,1},\ \beta_{m,1})= (\frac{1}{2}, \frac{9}{5})$

(i.e., all $\chi_m L_S$ bits in a packet are successfully detected) is expressed as

$$P_{f,k}\left((S_C, S_B, W, M) = (i, n, w, m)\right) = (1 - P_{e,b}(i, n, w, m))^{\chi_m L_S}. \quad (15)$$

If the sensor fails to decode the received data packet, the retransmission mechanism is employed in the sensor communications. Let Z be the total number of retransmissions required to successfully convey a data packet. By assuming that each transmission is independent, the variable Z can be expressed as a geometric random variable, and the average number of retransmissions for the successful reception of a packet is given by

$$\mathbb{E}[Z] = 1/P_{f,k}(i, n, w, m). \quad (16)$$

Since $T_L = DT_P$, the number of effective data packets due to retransmission during each management period is in average given as

$$D_E = \frac{D}{\mathbb{E}[Z]} = \frac{T_L}{\mathbb{E}[Z]T_P}. \quad (17)$$

From (14)–(17), the net bit rate can therefore be lower bounded by

$$\begin{aligned} G_{w,m}\left((S_C, S_B) = (i, n)\right) &= \frac{1}{T_L} D_E \chi_m L_S \\ &= \frac{1}{T_P} \chi_m L_S \left(1 - P_{e,b}(i, n, w, m)\right)^{\chi_m L_S} \\ &\geq \frac{1}{T_P} \chi_m L_S \left(1 - \eta(i, n, w, m)\right)^{\chi_m L_S}. \end{aligned} \quad (18)$$

Since $Q(x) \leq \frac{1}{2}\exp\left(-\frac{x^2}{2}\right)$ is asymptotically tight as x is large, or equivalently, $P_{e,b}(i, n, w, m) \approx \eta(i, n, w, m)$ for a sufficiently large γ_U, this implies that the lower bound of the net bit rate is tight in high SNR regimes.

Definition 1: The reward function for the action $(W, M) = (w, m)$ at the state $(S_C, S_B) = (i, n)$ is defined as

$$R_{w,m}\left((S_C, S_B) = (i, n)\right) = \begin{cases} 0, & w = 0; \\ \frac{1}{T_P}\chi_m L_S \left(1 - \eta(i, n, w, m)\right)^{\chi_m L_S}, & w \in \mathcal{W}\backslash\{0\}. \end{cases} \quad (19)$$

The reward function has the following properties:

(a) $R_{w,m}((S_C, S_B) = (i, n)) = 0$ for $w = 0$, because no data transmission occurs when the transmission power is zero.

(b) $R_{w,m}((S_C, S_B) = (i, n)) = R_{w',m}((S_C, S_B) = (i, n'))$ for any $w = w'$, because the immediate reward is independent of the battery state.

(c) $R_{w,m}((S_C, S_B) = (i, n)) \geq R_{w,m}((S_C, S_B) = (i', n))$ for any $i \geq i'$, which means a higher immediate reward is obtained as the channel condition improves.

E. Transmission Policies

Two transmission policies are implemented regarding the affordable actions in the action set $\mathcal{A}$.

Definition 2 (Composite Policy): A transmission policy is composite, if $N_P \geq N_B$. The action set at the n^{th} battery state is given by $\mathcal{A} = \{0, \ldots, n\} \times \{0, \ldots, N_M - 1\}$.

Definition 3 (On-Off Policy): A transmission policy is on-off, if $N_P = 2$ and $N_M = 1$. The action set at the n^{th} battery state is given by $\mathcal{A} = \{0, \ldots, \min\{n, 1\}\} \times \{0\}$.

In the composite policy, the power action could be unconditional as long as the resultant energy consumption during the management period is below the battery supply. On the contrary, only a single power and modulation level is accessible in the on-off policy whenever the sensor is active. The composite policy undoubtedly has better performance than the on-off policy, whereas the later one, as its name suggests, operates in a relatively simple on-off switching mode for data transmission.

IV. OPTIMIZATION OF TRANSMISSION POLICIES

A. Optimal Policies and Belief Update

The main goal of the MDP is to find a decision policy $\pi(s) : \mathcal{S} \to \mathcal{A}$ that specifies the optimal action in the state s and maximizes the objective function. Since we are interested in maximizing some cumulative functions of the random rewards in the Markov chain, the expected discounted infinite-horizon reward is formulated by using (19):

$$V_\pi(s_0) = \mathbb{E}_\pi\Big[\sum\nolimits_{k=0}^{\infty} \lambda^k R_{\pi(s_k)}(s_k)\Big], \qquad s_k \in \mathcal{S},\ \pi(s_k) \in \mathcal{A}, \quad (20)$$

where $V_\pi(s_0)$ is the expected reward starting from the initial state s_0 and continuing with the policy π from then on, and $0 \leq \lambda < 1$ is a discount factor. The adjustment of λ provides a wide range of performance characteristics, and the long run average objective can be closely approximated by choosing a discount factor close to one.[2] It is known that the optimal value of the expected reward is unrelated to the initial state if the states of the Markov chain are assumed to be recurrent. From (9) and (20), there exists an optimal stationary policy $\pi^*(s)$ that satisfies the Bellman's equation [32]:

$$V_{\pi^*}(s) = \max_{a \in \mathcal{A}} \left(R_a(s) + \lambda \sum_{s' \in \mathcal{S}} P_a(s'|s) V_{\pi^*}(s') \right), \qquad s \in \mathcal{S}. \quad (21)$$

The well-known value iteration approach is then applied to iteratively find the optimal policy [32]:

$$V_{i+1}^a(s) = R_a(s) + \lambda \sum_{s' \in \mathcal{S}} P_a(s'|s) V_i(s'),\ s \in \mathcal{S},\ a \in \mathcal{A}; \quad (22)$$

$$V_{i+1}(s) = \max_{a \in \mathcal{A}} \left\{ V_{i+1}^a(s) \right\}, \quad s \in \mathcal{S}, \quad (23)$$

[2] A link between average and discounted objective problems is provided in [32]. Define the long run average reward as $\bar{V}_\pi(s_0) = \limsup_{N\to\infty} \frac{1}{N} \cdot \mathbb{E}[\sum_{k=0}^{N-1} R_{\pi(s_k)}(s_k)]$. For any stationary policy π, $\bar{V}_\pi(s_0) = \lim_{\lambda\to 1}(1 - \lambda)V_\pi(s_0)$. Hence, a policy that maximizes $V_\pi(s_0)$ for $\lambda \approx 1$ also approximately maximizes the average cost $\bar{V}_\pi(s_0)$.

TABLE IV
EFFECT OF PARAMETERS ON SYSTEM PERFORMANCE

Parameters	Summary of effect on system performance
N_B	The energy overflow problem is relieved as N_B increases. With a larger energy buffer size, the sensor can get higher rewards by opportunistically spending energy at good channels before energy overflows.
N_C	Let $\bar{\mathbf{\Gamma}}$ be a sub-partition of a channel quantization set $\mathbf{\Gamma}$, i.e. $\mathbf{\Gamma} \subseteq \bar{\mathbf{\Gamma}}$. As compared with $\mathbf{\Gamma}$, the partition $\bar{\mathbf{\Gamma}}$, which has a larger value of N_C, gives better performance due to higher channel resolution.
f_D	When f_D increases, the sensor experiences more channel transitions before energy overflows, and it can take the opportunity to spend energy at good channels. (See Appendix A for details.)
Ω_S, ϑ	The amount of harvested energy at the sensor becomes larger with the increase of Ω_S and ϑ.
N_P, N_M	Enlarged sets of power levels and modulation types can facilitate performance improvement because more diversified actions are acquirable in response to the changes of channel and battery conditions.

where i is the iteration index, and the initial value of $V_0(s)$ is set as zero for all $s \in \mathcal{S}$. The update rule is repeated for several iterations until a stop criterion is satisfied, i.e., $|V_{i+1}(s) - V_i(s)| \leq \varepsilon$. Based on the definition in Section III-E, the optimal solutions for the two policies can be found by alternatively executing (22) and (23). In general, the convergence of the value iteration algorithm is guaranteed, and interested readers are referred to [32] for more details. To get more insight, we also summarize the impact of various parameters on the system performance in Table IV.

In real applications, the channel state of the communication link can be reliably obtained at the transmitter via channel feedback information. The belief of the solar state can be calculated from the observation prior to the action decision. Let $x^{(t)}$ be the average value of the measured solar data during the t^{th} management period, and $\zeta_j^{(t-1)} = P(S_H^{(t-1)} = j|x^{(1)}, \ldots, x^{(t-1)})$ be the belief of the j^{th} solar state according to the historical observation up to the $(t-1)^{th}$ period. With the solar power harvesting model, the belief information at the t^{th} period is updated using Bayes' rule as follows:

$$\zeta_j^{(t)} = \frac{\sum_{i=0}^{N_H-1} \zeta_i^{(t-1)} a_{ij} f_j\left(x^{(t)}\right)}{\sum_{j'=0}^{N_H-1}\sum_{i'=0}^{N_H-1} \zeta_{i'}^{(t-1)} a_{i'j'} f_{j'}\left(x^{(t)}\right)}, \quad (24)$$

where $f_j(x) = \mathcal{N}(x; \mu_j, \rho_j)$ and a_{ij} are the normal distribution and the state transition probability, as obtained in the training results of Section II. The final task is to apply the belief information for deciding the action at each management period. We consider the following mixed strategy. Remember that in the construction of the solar power harvesting model, each observed data sample contributes to the values of all underlying parameters at different states in *a posteriori* probability sense in the EM training procedures [28]. Thus, the optimization of the transmission policy inherently accounts for the probability of the observation that belongs to each solar state. This implies that the mixed strategy, which randomly plays the optimal action corresponding to the j^{th} solar state with probability proportional to $\zeta_j^{(t)}$, is the optimal strategy for the observations up to the present time.

B. Computational Complexity

We now discuss the computational complexity of finding the optimal transmission policies. The main complexity of the value iteration algorithm arises from the multiplication in (22), and the required number of multiplications per iteration is given as

$$\sum_{j=0}^{N_H-1}\sum_{i=0}^{N_C-1}\sum_{n=0}^{N_B-1}\sum_{a=0}^{\min\{n,N_P-1\}} N_H N_C (N_B - n + a)$$
$$= \begin{cases} (N_H N_C)^2\left(N_B^2 + N_B - 1\right), & \text{on-off policy;} \\ \frac{1}{6}(N_H N_C)^2\left(2N_B^3 + 3N_B^2 + N_B\right), & \text{composite policy.} \end{cases} \quad (25)$$

In summary, the on-off policy has the complexity of $\mathcal{O}(N_H^2 N_C^2 N_B^2)$, while the composite policy has $\mathcal{O}(N_H^2 N_C^2 N_B^3)$. In real applications, the optimal policy can be precalculated offline and stored in memory as a look-up table. Thus, the involved online computation for the sensor node is to update the belief information in (24), which has the complexity of $\mathcal{O}(N_H^2)$.

V. OPTIMAL ON-OFF TRANSMISSION POLICIES

A. Threshold Structure of Transmission Policies

To facilitate analysis, we focus on a simple on-off transmission policy and drop the modulation type index m, i.e., $a = w \in \{0, 1\}$. From (6), (9) and (11)–(13), the expected reward function with respect to the action a in (22) can be rewritten as an expected form:

$$V_{i+1}^a(z, x, y) = R_a(x, y) + \lambda \sum_{j=0}^{N_H-1} P(S_H = j|S_H = z)$$
$$\cdot \sum_{l=\max\{0,x-1\}}^{\min\{x+1,N_C-1\}} P(S_C = l|S_C = x)$$
$$\cdot \sum_{q=0}^{\infty} P(Q = q|S_H = z) V_i\left(j, l, \min\{N_B - 1, y - a + q\}\right)$$
$$= R_a(x, y) + \lambda \cdot \mathbb{E}_{z,x,y}\left[V_i\left(j, l, \min\{N_B - 1, y - a + q\}\right)\right], \quad (26)$$

where the subscript in $\mathbb{E}_{z,x,y}[\cdot]$ is used to indicate the associated solar, channel and battery states.

Lemma 1: For any fixed solar state $z \in \mathcal{H}$ and channel state $x \in \mathcal{C}$, $V_i^a(z, x, y-1) \leq V_i^a(z, x, y)$, $\forall\, y \in \mathcal{B} \setminus \{0\}$ and $a = 0, 1$. Moreover, $V_i(z, x, y-1) \leq V_i(z, x, y)$, $\forall\, y \in \mathcal{B} \setminus \{0\}$.

Proof: From (23), if $V_i^a(z, x, y-1) \leq V_i^a(z, x, y)$ is satisfied, it implies

$$V_i(z, x, y-1) = \max_{a\in\{0,1\}} \{V_i^a(z, x, y-1)\}$$
$$\leq \max_{a\in\{0,1\}} \{V_i^a(z, x, y)\} = V_i(z, x, y). \quad (27)$$

We prove the lemma by the induction. From (26) and the initial condition $V_0(s)=0$, the statement is held for $i=1$ because $V_1^a(z,x,y-1)$ and $V_1^a(z,x,y)$ only relate to the same reward, for $a\in\{0,1\}$. Hence, we obtain $V_1(z,x,y-1)=V_1(z,x,y)$. Assume $i=k$ holds, and for any $z\in\mathcal{H}$ and $x\in\mathcal{C}$, it gives $V_k(z,x,y-1)\le V_k(z,x,y)$, $\forall\, y\in\mathcal{B}\setminus\{0\}$. Using (26), we prove that for $i=k+1$:

$$\begin{aligned}
&V_{k+1}^a(z,x,y)-V_{k+1}^a(z,x,y-1)\\
&\quad=\lambda\sum_{j=0}^{N_H-1}P(S_H=j|S_H=z)\\
&\qquad\cdot\sum_{l=\max\{0,x-1\}}^{\min\{x+1,N_C-1\}}P(S_C=l|S_C=x)\sum_{q=0}^{\infty}P(Q=q|S_H=z)\\
&\qquad\cdot\big(V_k\left(j,l,\min\{N_B-1,y-a+q\}\right)\\
&\qquad\underbrace{-V_k\left(j,l,\min\{N_B-1,y-1-a+q\}\right)}_{\ge 0,\text{ since the assumption holds for } i=k.}\big)\ge 0. \quad (28)
\end{aligned}$$

This thereby implies that $V_{k+1}(z,x,y-1)\le V_{k+1}(z,x,y)$, and the statement holds for $i=k+1$. ■

Theorem 1: For the optimal on-off policy, the long-term expected reward is non-decreasing with respect to the battery state. That is, for any $z\in\mathcal{H}$ and $x\in\mathcal{C}$, $V_{\pi^*}(z,x,y-1)\le V_{\pi^*}(z,x,y)$, $\forall\, y\in\mathcal{B}\setminus\{0\}$.

Proof: By applying Lemma 1 and following the value iteration algorithm, the theorem is proved when the algorithm has converged. ■

Now we turn to describing the structure of the on-off transmission policy. Since no transmission (i.e., $a=0$) is the only action when the battery state is zero, we concentrate on the actions for $y\in\mathcal{B}\setminus\{0\}$ in the following.

Lemma 2: For each $z\in\mathcal{H}$, $x\in\mathcal{C}$ and $y\in\mathcal{B}\setminus\{0\}$, define two difference functions:

$$\Theta_i(z,x,y)=V_i^1(z,x,y)-V_i^0(z,x,y); \quad (29)$$

$$\begin{aligned}
\Lambda_i(z,x,y)=&\mathbb{E}_{z,x,y}\left[V_i^1\left(j,l,\min\{N_B-1,y+q\}\right)\right.\\
&\left.-V_i^0\left(j,l,\min\{N_B-1,y-1+q\}\right)\right]. \quad (30)
\end{aligned}$$

The function $\Theta_i(z,x,y)$ is monotonically non-decreasing in $y\in\mathcal{B}\setminus\{0\}$, if the function $\Lambda_t(z,x,y)$ is non-increasing in $y\in\mathcal{B}\setminus\{0\}$, $\forall\, t<i$, $z\in\mathcal{H}$ and $x\in\mathcal{C}$.

Proof: We use induction to prove this lemma. When $i=1$, the statement is true because $\Theta_1(z,x,y)=V_1^1(z,x,y)-V_1^0(z,x,y)=R_1(x,y)$, for $y\neq 0$, and the reward function $R_1(x,y)$ keeps the same value in $y\in\mathcal{B}\setminus\{0\}$ for any given $x\in\mathcal{C}$. Assume $i=k$ holds, the function $\Theta_k(z,x,y)$ is non-decreasing in $y\in\mathcal{B}\setminus\{0\}$, $\forall\, z\in\mathcal{H}$ and $\forall\, x\in\mathcal{C}$. It immediately implies that the following two functions are both non-decreasing in y:

$$\Delta_k^{\max}(z,x,y)=\max\{0,\Theta_k(z,x,y)\}\ge 0; \quad (31)$$

$$\Delta_k^{\min}(z,x,y)=\min\{0,\Theta_k(z,x,y)\}\le 0. \quad (32)$$

For $i=k+1$, the difference function $\Theta_{k+1}(z,x,y)$ can be derived from (23) and (26) as follows:

$$\begin{aligned}
\Theta_{k+1}(z,x,y)&=V_{k+1}^1(z,x,y)-V_{k+1}^0(z,x,y)\\
&=R_1(x,y)-R_0(x,y)\\
&\quad+\lambda\mathbb{E}_{z,x,y}\left[\max\left\{V_k^0\left(j,l,\min\{N_B-1,y-1+q\}\right),\right.\right.\\
&\qquad\left.\left.V_k^1\left(j,l,\min\{N_B-1,y-1+q\}\right)\right\}\right]\\
&\quad-\lambda\mathbb{E}_{z,x,y}\left[\max\left\{V_k^0\left(j,l,\min\{N_B-1,y+q\}\right),\right.\right.\\
&\qquad\left.\left.V_k^1\left(j,l,\min\{N_B-1,y+q\}\right)\right\}\right]. \quad (33)
\end{aligned}$$

Inserting (31) and (32) into (33) yields

$$\begin{aligned}
\Theta_{k+1}(z,x,y)&=R_1(x,y)-\lambda\Lambda_k(z,x,y)\\
&\quad+\lambda\mathbb{E}_{z,x,y}\left[\Delta_k^{\max}\left(j,l,\min\{N_B-1,y-1+q\}\right)\right]\\
&\quad+\lambda\mathbb{E}_{z,x,y}[\Delta_k^{\min}\left(j,l,\min\{N_B-1,y+q\}\right)]. \quad (34)
\end{aligned}$$

According to the non-decreasing property of the functions $\Delta_k^{\max}(z,x,y)$, $\Delta_k^{\min}(z,x,y)$ and $R_1(x,y)$, it can be shown from (34) that $\Theta_{k+1}(z,x,y)$ preserves the non-decreasing property in $y\in\mathcal{B}\setminus\{0\}$, if $\Lambda_k(z,x,y)$ is non-increasing in $y\in\mathcal{B}\setminus\{0\}$, $\forall\, z\in\mathcal{H}$ and $\forall\, x\in\mathcal{C}$. ■

In fact, the validity of the non-decreasing property of $\Theta_i(z,x,y)$ relies on the transition probabilities of the solar states, channel states and battery states, and this property is not necessarily satisfied in $z\in\mathcal{H}$ and $x\in\mathcal{C}$. Below we show that the function $\Lambda_t(z,x,y)$ is indeed non-increasing in the direction along the battery states for a given solar state and channel state, and the following theorem is provided.

Theorem 2: For any $z\in\mathcal{H}$ and $x\in\mathcal{C}$, the difference function $\Theta_i(z,x,y)$ is non-decreasing in $y\in\mathcal{B}\setminus\{0\}$, and the optimal on-off policy has a threshold structure.

Proof: We first show that $\Lambda_t(z,x,y+1)-\Lambda_t(z,x,y)\le 0$, for $y=1,\ldots,N_B-2$, in the following. It can be derived from the definition in Lemma 2 that

$$\begin{aligned}
\Lambda_t(z,x,y+1)-\Lambda_t(z,x,y)&=\sum_{j=0}^{N_H-1}P(S_H=j|S_H=z)\\
\cdot\sum_{l=\max\{0,x-1\}}^{\min\{x+1,N_C-1\}}&P(S_C=l|S_C=x)\cdot\sum_{q=0}^{\infty}P(Q=q|S_H=z)\Phi_y(j,l,q), \quad (35)
\end{aligned}$$

where the term $\Phi_y(j,l,q)$, for $y=1,\ldots,N_B-2$, is defined as

$$\begin{aligned}
\Phi_y(j,l,q)=&V_t^1\left(j,l,\min\{N_B-1,y+1+q\}\right)\\
&-V_t^0\left(j,l,\min\{N_B-1,y+q\}\right)\\
&-V_t^1\left(j,l,\min\{N_B-1,y+q\}\right)\\
&+V_t^0\left(j,l,\min\{N_B-1,y-1+q\}\right). \quad (36)
\end{aligned}$$

The third summation over the variable q in (35) can be further divided into three cases according to the range of q, and after some straightforward manipulations, we obtain

$$\Phi_y(j,l,q)=\begin{cases}0, & q=0,\ldots,(N_B-y-2);\\ -\left(V_t^0(j,l,N_B-1)-V_t^0(j,l,N_B-2)\right)\leq 0, & q=(N_B-y-1);\\ 0, & q=(N_B-y),\ldots,\infty,\end{cases} \tag{37}$$

where the values for $q \neq N_B - y - 1$ are equal to zero due to self-cancellation, and the inequality in the second line comes from Lemma 1. As a result, $\Phi_y(j,l,q)$ is always non-positive. From (35) and (37), it leads to $\Lambda_t(z,x,y+1) \leq \Lambda_t(z,x,y)$, and thus the function $\Lambda_t(z,x,y)$ is non-increasing in y. By applying Lemma 2, it suffices to prove that $\Theta_i(z,x,y)$ is non-decreasing in $y \in \mathcal{B} \setminus \{0\}$. When the value iteration algorithm is converged, a threshold structure $\boldsymbol{\kappa} = \{\boldsymbol{\kappa}_0, \ldots, \boldsymbol{\kappa}_{N_H-1}\}$, where $\boldsymbol{\kappa}_z = \{\kappa_{z,0}, \ldots, \kappa_{z,N_C-1}\}$, is given by using the non-decreasing property of $\Theta_i(z,x,y)$:

$$\pi^*(z,x,y)=\begin{cases}0, & y \leq \kappa_{z,x};\\ 1, & y \geq \kappa_{z,x}+1,\end{cases} \tag{38}$$

for a threshold $\kappa_{z,x}$ that is satisfied with $\Theta_i(z,x,\kappa_{z,x}) < 0$ and $\Theta_i(z,x,\kappa_{z,x}+1) \geq 0$ if $\kappa_{z,x} \in \mathcal{B} \setminus \{0\}$, and $\Theta_i(z,x,\kappa_{z,x}+1) \geq 0$ if $\kappa_{z,x} = 0$. ■

In fact, whether the threshold structure exists or not strongly depends on the state transition probabilities. The proof in Lemma 2 and Theorem 2 also implicitly indicates that the threshold structures for the solar or channel states do not necessarily occur, since for any fixed battery and solar (or channel) state, the function $\Lambda_i(z,x,y)$ after taking the expectation over the transition probabilities may not be guaranteed to be non-increasing with respect to the channel (or solar) states. By taking the training results in Table II as an example, a threshold structure is demonstrated in Fig. 4 for $S_H = 0$. It appears that there exists a threshold $\boldsymbol{\kappa}_0 = \{7,7,0,0,0,0\}$ above which data transmission occurs to gain the maximum long-term expected reward. Furthermore, it can be seen that for a fixed channel state, the long-term expected reward is non-decreasing with respect to the battery state. The simplicity of the threshold structure makes the on-off transmission policy attractive for hardware implementation, and it also helps reduce the computational burden in obtaining the optimal policy.

B. Energy Deficiency Condition

From (4) and (6), the harvested energy is quantized into two consecutive levels, $Q = 0$ and $Q = 1$, if the harvested power is less than P_U (i.e., the mean and variance of each solar state are small). The energy level of $Q = 0$ is referred to as energy deficiency. A necessary energy deficiency condition for the existence of an optimal threshold policy at $\boldsymbol{\kappa} = \{\boldsymbol{\kappa}_0, \ldots, \boldsymbol{\kappa}_{N_H-1}\}$ is provided in the following.

Theorem 3: Let $V_{\pi^*}(z,x,y)$ be the long-term expected reward of the on-off policy π^*. Define $\Xi(z,x,y) = \mathbb{E}_{z,x}[V_{\pi^*}(j,l,\min\{N_B-1,y+1\}) - V_{\pi^*}(j,l,\min\{N_B-1,y\})]$ as a difference function of $V_{\pi^*}(z,x,y)$ at the two battery states $\min\{N_B-1,y+1\}$ and $\min\{N_B-1,y\}$,

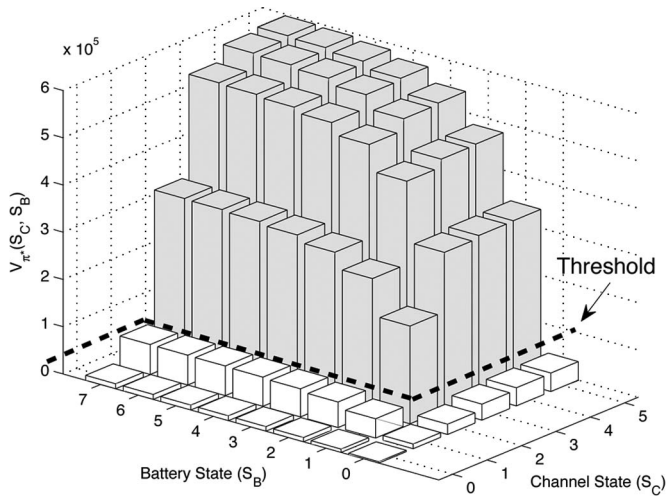

Fig. 4. Threshold structure policy and long-term expected reward for the solar state $S_H = 0$ ($N_C = 6$, $N_B = 8$, $R_S = 10^5$ symbols/sec, $L_S = 10^3$ symbols/packet, $T_L = 300$ sec, $P_U = 1.8 \times 10^4$ μW, $\gamma_U = 18.5$ dB, $\Omega_S = 0.1$ cm^2, $\vartheta = 1$, $f_D = 5 \times 10^{-2}$, $\boldsymbol{\Gamma} = \{0, 0.3, 0.6, 1.0, 2.0, 3.0, \infty\}$, $\lambda = 0.5$ and 8 PSK).

which is averaged over the channel and solar state transition probabilities from the state $(z,x) \in \mathcal{H} \times \mathcal{C}$ to its adjacent states. Consider two possible energy quantum levels $Q = 0$ and $Q = 1$. There exists an optimal policy with the threshold $\boldsymbol{\kappa} = \{\boldsymbol{\kappa}_0, \ldots, \boldsymbol{\kappa}_{N_H-1}\}$, only if the energy deficiency probability belongs to the interval $\mathcal{D}_z = \bigcap_{x=0}^{N_C-1} \mathcal{D}_{z,x}$, where $\mathcal{D}_{z,x}$ is defined as

$$\mathcal{D}_{z,x}=\begin{cases}P(Q=0|S_H=z) \leq \phi(z,x,1), & \kappa_{z,x}=0;\\ P(Q=0|S_H=z) \geq \phi(z,x,0), & \kappa_{z,x}=N_B-1;\\ \phi(z,x,0) \leq P(Q=0|S_H=z) \leq \phi(z,x,1), & \text{otherwise},\end{cases} \tag{39}$$

where $\phi(z,x,n) = \frac{R_1(x)/\lambda - \Xi(z,x,\kappa_{z,x}+n)}{\Xi(z,x,\kappa_{z,x}+n-1) - \Xi(z,x,\kappa_{z,x}+n)}$ and $R_1(x) = R_1(x,\kappa_{z,x}+1) = R_1(x,\kappa_{z,x})$.

Proof: By applying Theorem 2, it is sufficient to show that $\boldsymbol{\kappa}$ is the optimal threshold policy, only if the following conditions are satisfied, $\forall\, z \in \mathcal{H}$ and $\forall\, x \in \mathcal{C}$:

$$\begin{cases}V_{\pi^*}^1(z,x,\kappa_{z,x}+1) \geq V_{\pi^*}^0(z,x,\kappa_{z,x}+1), & \kappa_{z,x}=0;\\ V_{\pi^*}^1(z,x,\kappa_{z,x}) \leq V_{\pi^*}^0(z,x,\kappa_{z,x}), & \kappa_{z,x}=N_B-1;\\ V_{\pi^*}^1(z,x,\kappa_{z,x}) \leq V_{\pi^*}^0(z,x,\kappa_{z,x}) \text{ and} & \\ \quad V_{\pi^*}^1(z,x,\kappa_{z,x}+1) \geq V_{\pi^*}^0(z,x,\kappa_{z,x}+1), & \text{otherwise}.\end{cases} \tag{40}$$

From the definition in (26), the condition of $V_{\pi^*}^1(z,x,\kappa_{z,x}) \leq V_{\pi^*}^0(z,x,\kappa_{z,x})$ in (40) becomes

$$R_1(x) \leq \lambda \sum_{q=0}^{1} P(Q=q|S_H=z)\Xi(z,x,\kappa_{z,x}-1+q), \quad z \in \mathcal{H} \text{ and } x \in \mathcal{C}. \tag{41}$$

On the other hand, the condition of $V_{\pi^*}^1(z,x,\kappa_{z,x}+1) \geq V_{\pi^*}^0(z,x,\kappa_{z,x}+1)$ implies that

$$R_1(x) \geq \lambda \sum_{q=0}^{1} P(Q=q|S_H=z)\Xi(z,x,\kappa_{z,x}+q), \quad z \in \mathcal{H} \text{ and } x \in \mathcal{C}. \tag{42}$$

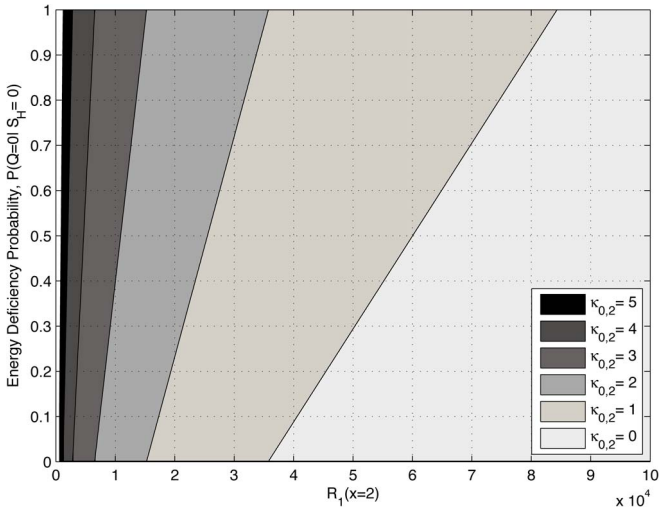

Fig. 5. Energy deficiency regions $P(Q=0|S_H=0)$ versus immediate rewards $R_1(x=2)$ for different thresholds $\kappa_{0,2}$.

In addition, it can be derived that $\Xi(z,x,\kappa_{z,x}-1)-\Xi(z,x,\kappa_{z,x})\geq 0$ as follows:

$$\begin{aligned}
&\Xi(z,x,\kappa_{z,x}-1)-\Xi(z,x,\kappa_{z,x})\\
&\quad=\mathbb{E}_{z,x}\left[V^0_{\pi^*}\left(j,l,\min\{N_B-1,\kappa_{z,x}\}\right)\right.\\
&\qquad\qquad\left.-V^0_{\pi^*}\left(j,l,\min\{N_B-1,\kappa_{z,x}-1\}\right)\right]\\
&\qquad-\mathbb{E}_{z,x}\left[V^1_{\pi^*}\left(j,l,\min\{N_B-1,\kappa_{z,x}+1\}\right)\right.\\
&\qquad\qquad\left.-V^0_{\pi^*}\left(j,l,\min\{N_B-1,\kappa_{z,x}\}\right)\right]\\
&\quad\geq\mathbb{E}_{z,x}\left[V^1_{\pi^*}\left(j,l,\min\{N_B-1,\kappa_{z,x}\}\right)\right.\\
&\qquad\qquad\left.-V^0_{\pi^*}\left(j,l,\min\{N_B-1,\kappa_{z,x}-1\}\right)\right]\\
&\qquad-\mathbb{E}_{z,x}\left[V^1_{\pi^*}\left(j,l,\min\{N_B-1,\kappa_{z,x}+1\}\right)\right.\\
&\qquad\qquad\left.-V^0_{\pi^*}(j,l,\min\{N_B-1,\kappa_{z,x}\})\right]\geq 0, \quad (43)
\end{aligned}$$

where the threshold structure is used in the first equality; for instance, $V_{\pi^*}(z,x,y)=V^0_{\pi^*}(z,x,y)$, for $y\leq\kappa_{z,x}$, and the last inequality holds due to (36) and (37). Similarly, we get $\Xi(z,x,\kappa_{z,x})-\Xi(z,x,\kappa_{z,x}+1)\geq 0$. By applying $P(Q=0|S_H=z)+P(Q=1|S_H=z)=1$ and (41)–(43) into (40), the necessary conditions can be rewritten as in (39). Hence, there exists an optimal threshold at $\boldsymbol{\kappa}=\{\boldsymbol{\kappa}_0,\ldots,\boldsymbol{\kappa}_{N_H-1}\}$, only if the probability $P(Q=0|S_H=z)\in\mathcal{D}_z=\bigcap_{x=0}^{N_C-1}\mathcal{D}_{z,x}$. ■

This necessary condition gives an important insight into how the energy deficiency probability affects the threshold of the policy. Taking the long-term expected reward in Fig. 4 and $S_H=0$ as an example, the energy deficiency regions versus the immediate rewards $R_1(x=2)$ for different thresholds $\kappa_{0,2}$ are plotted in Fig. 5, where the other thresholds are fixed at $\{\kappa_{0,0},\kappa_{0,1},\kappa_{0,3},\kappa_{0,4},\kappa_{0,5}\}=\{7,7,0,0,0\}$. It is observed that for $R_1(x=2)=2\times10^4$ and 6×10^4, the threshold $\kappa_{0,2}=1$ could be the optimal policy, only if $P(Q=0|S_H=0)\leq 0.25$ and $P(Q=0|S_H=0)\geq 0.5$, respectively.

C. Expected Net Bit Rate Analysis

Here we use the expected net bit rate to assess the performance of the optimal threshold policy. Consider a threshold policy $\boldsymbol{\kappa}=\{\boldsymbol{\kappa}_0,\ldots,\boldsymbol{\kappa}_{N_H-1}\}$, and denote $\nu_{j,i\times N_B+n}$ as the stationary probability of the state $(S_H,S_C,S_B)=(j,i,n)$, for $i=0,\ldots,N_C-1$ and $n=0,\ldots,N_B-1$. Define $\boldsymbol{\nu}_j=[\nu_{j,0},\ldots,\nu_{j,i\times N_B+n},\ldots,\nu_{j,N_C\times N_B-1}]^T$, for $j=0,\ldots,N_H-1$, and $\boldsymbol{\nu}=[\boldsymbol{\nu}_0^T,\ldots,\boldsymbol{\nu}_{N_H-1}^T]^T$. Let $\boldsymbol{\Pi}_{j,i}$ be an $N_B\times N_B$ battery state transition probability matrix associated with the threshold policy $\boldsymbol{\kappa}$ at the j^{th} solar state and the i^{th} channel state, given by

$$[\boldsymbol{\Pi}_{j,i}]_{p,q}=\begin{cases}
P\left(Q=(p-q)|S_H=j\right),\\
\quad 0\leq q\leq\kappa_{j,i},\ \ q\leq p\leq N_B-2;\\
0,\ 0\leq q\leq\kappa_{j,i},\ \ 0\leq p\leq q-1;\\
P\left(Q=(p-q+1)|S_H=j\right),\\
\quad \kappa_{j,i}+1\leq q\leq N_B-1,\ \ q-1\leq p\leq N_B-2;\\
0,\ \ \kappa_{j,i}+1\leq q\leq N_B-1,\ \ 0\leq p\leq q-2,
\end{cases} \quad (44)$$

and $[\boldsymbol{\Pi}_{j,i}]_{N_B-1,q}=1-\sum_{p=0}^{N_B-2}[\boldsymbol{\Pi}_{j,i}]_{p,q}$, for $q=0,\ldots,N_B-1$, where the $(p,q)^{th}$ entry of the matrix $[\boldsymbol{\Pi}_{j,i}]$ represents the transition probability from the state $(S_H,S_C,S_B)=(j,i,q)$ to the state $(S_H,S_C,S_B)=(j,i,p)$. Therefore, the stationary probability with respect to the threshold policy $\boldsymbol{\kappa}$ can be computed by solving the balance equation:

$$\begin{bmatrix}\boldsymbol{\Phi}-\mathbf{I}_{(N_B\times N_C\times N_H)}\\ \mathbf{1}^T_{(N_B\times N_C\times N_H)}\end{bmatrix}\boldsymbol{\nu}=\begin{bmatrix}\mathbf{0}_{(N_B\times N_C\times N_H)}\\ 1\end{bmatrix}, \quad (45)$$

where $\boldsymbol{\Phi}$ is the state transition probability matrix of size $(N_B\times N_C\times N_H)\times(N_B\times N_C\times N_H)$, whose $(zN_C+x,jN_C+i)^{th}$ sub-matrix is equal to $P(S_H=z|S_H=j)\cdot P(S_C=x|S_C=i)\cdot\boldsymbol{\Pi}_{j,i}$, for $z,j=0,\ldots,N_H-1$, $i=0,\ldots,N_C-1$, and $x=\max\{0,i-1\},\ldots,\min\{i+1,N_C-1\}$, and the remaining sub-matrices all equate to zero. By taking the expectation of the reward function in (19), the expected net bit rate using the 2^{χ_m}-ary modulation scheme is given by

$$R_{net,m}=\frac{1}{T_P}\sum_{j=0}^{N_H-1}\sum_{i=0}^{N_C-1}\sum_{n\geq\kappa_{j,i}+1}^{N_B-1}\nu_{j,(i\times N_B+n)}\cdot\chi_m L_S\left(1-\eta(i,n,1,m)\right)^{\chi_m L_S}. \quad (46)$$

Theorem 4: Define an energy harvesting rate as $\bar{q}=\lim_{T\to\infty}\bar{q}_T=\lim_{T\to\infty}\mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T}q_t\right]$, where q_t denotes the number of energy quanta obtained by a sensor at the t^{th} policy management period. The expected net bit rate of the on-off policy is upper bounded by

$$R_{net,m}\leq\min\{\bar{q},1\}\cdot\left(\frac{1}{T_P}\chi_m L_S\cdot\left(1-\eta(N_C-1,N_B-1,1,m)\right)^{\chi_m L_S}\right). \quad (47)$$

At asymptotically high SNR, the upper bound value converges to $\min\{\bar{q},1\}\cdot\frac{1}{T_P}\chi_m L_S$.

Proof: Let $a_t\in\{0,1\}$ be the optimal action at the t^{th} policy management period, corresponding to a sequence of channel states x_t and battery states y_t, for $t=1,\ldots,T$. From (19), the immediate reward can be rewritten as $R_m(a_t,x_t,y_t)=a_t\frac{1}{T_P}\chi_m L_S(1-\eta(x_t,y_t,1,m))^{\chi_m L_S}$. Thus, the average net

bit rate, $R_{net,m} = \lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} R_m(a_t, x_t, y_t)\right]$, is calculated as

$$\begin{aligned} R_{net,m} &= \lim_{T\to\infty} \sum_{i_t} P(x_t = i_t, t = 1, \ldots, T) \\ &\quad \cdot \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[a_t \, \frac{1}{T_P}\chi_m L_S \left(1 - \eta(x_t, y_t, 1, m)\right)^{\chi_m L_S} \middle| \right. \\ &\qquad\qquad \left. x_t = i_t, t = 1, \ldots, T\right] \\ &\le \lim_{T\to\infty} \sum_{i_t} P(x_t = i_t, t = 1, \ldots, T) \\ &\quad \cdot \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[a_t | x_t = i_t, t = 1, \ldots, T\right] \\ &\quad \cdot \left(\frac{1}{T_P}\chi_m L_S \left(1 - \eta(N_C - 1, N_B - 1, 1, m)\right)^{\chi_m L_S}\right), \end{aligned} \quad (48)$$

where the marginal probability is performed in the first equality by summing over all channel states i_t, and the BER relationship of $\eta(N_C - 1, N_B - 1, 1, m) \le \eta(x_t, y_t, 1, m)$ is used in the second inequality. For any transmission policy, the accumulated energy consumption cannot exceed the initial energy in the battery plus the total amount of harvested energy, and it yields the constraint:

$$\frac{1}{T}\sum_{t=1}^{T} a_t \le \frac{1}{T}(N_B - 1) + \frac{1}{T}\sum_{t=1}^{T} q_t. \quad (49)$$

Besides, the on-off transmission imposes another energy expenditure constraint of $\frac{1}{T}\sum_{t=1}^{T} a_t \le 1$. Substituting this constraint and (49) into (48), we finally obtain the upper bound of the expected net bit rate in (47). From (14), it is found that the function $\eta(N_C - 1, N_B - 1, 1, m) \to 0$ as $\gamma_U \to \infty$, and the upper bound converges to $\min\{\bar{q}, 1\} \cdot \frac{1}{T_P}\chi_m L_S$ at asymptotically high SNR. ■

D. Some Structure Results for Composite Policies

Fig. 6 depicts the optimal composite policies with the same parameters of Fig. 4, except as otherwise stated. A monotonic policy is observed in the direction along the battery states. To be explicit, for any fixed $z \in \mathcal{H}$ and $x \in \mathcal{C}$, $\pi^*(z, x, y) \preceq \pi^*(z, x, y'), \forall\, y \le y'$, where $\preceq$ is a generalized inequality. From Lemma 4.7.1 in [33], there exists such a monotonic property, if $V_{i+1}^a(z, x, y)$ in (23) is a superadditive function on $\mathcal{A} \times \mathcal{B}$.[3] Since it is tough to directly inspect the superadditivity of $V_{i+1}^a(z, x, y)$, a sufficient condition is provided in the following theorem.

Theorem 5: The optimal composite policy is a monotonic policy, if the energy harvesting condition:

$$\begin{aligned} &\sum_{i=\max\{\alpha - n^+ + w^+, 0\}-1}^{\max\{\alpha - n^+ + w^+, 0\}-1} P(Q = i | S_H = j) \\ &\qquad \le \sum_{i=\max\{\alpha - n^- + w^-, 0\}}^{\max\{\alpha - n^- + w^+, 0\}-1} P(Q = i | S_H = j), \end{aligned} \quad (50)$$

is satisfied at each solar state, $\forall\, \alpha \in \mathcal{B}$, $\forall\, n^+ \ge n^- \in \mathcal{B}$, and $\forall\, w^+ \ge w^- \in \mathcal{W}$.

Proof: From (13), let us first define $\beta_w(\alpha | j, n) = \sum_{k=\alpha}^{N_B - 1} P_w(S_B = k | (S_H, S_B) = (j, n))$, for $\alpha \in \mathcal{B}$. By applying Theorem 6.11.6 in [33], $V_{i+1}^a(z, x, y)$ is superadditive, if the following three conditions hold: (a) $R_a(z, x, y)$ is superadditive on $\mathcal{A} \times \mathcal{B}$; (b) $V_i^a(z, x, y)$ is nondecreasing in $y \in \mathcal{B}$, $\forall\, a \in \mathcal{A}$; (c) $\beta_w(\alpha | j, n)$ is superadditive on $\mathcal{W} \times \mathcal{B}$, $\forall\, \alpha \in \mathcal{B}$. It is straightforward to show that the condition (a) holds because the reward is independent of the battery state. Also, the condition (b) can be assured by extending the proof in Lemma 1 to the case of multiple power and modulation actions. From (13) and the condition (c), the sufficient condition (50) is then obtained after some manipulations. ■

The theorem implicitly indicates that the optimal composite policy tends to be monotonic when the probabilities of harvesting higher numbers of energy quanta are large enough to allow for a quick battery recovery. For example, a monotonic policy is given, if $P(Q = i | S_H = j)$ increases with the number of energy quanta i. Unlike the on-off policy, where the optimal threshold structure is always promised, the existence of such

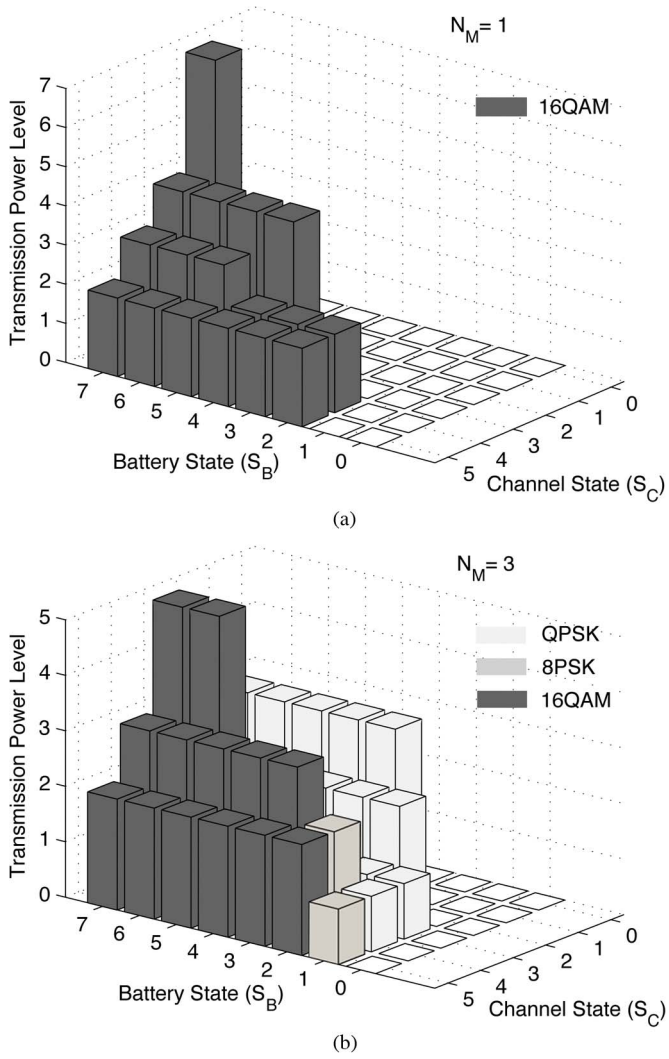

Fig. 6. Monotonic structures of the optimal composite policies with $N_M = 1$ and $N_M = 3$ ($S_H = 3$ and $\gamma_U = 12.5$ dB). (a) $N_M = 1$ and $\Omega_S = 0.1$ cm^2; (b) $N_M = 3$ and $\Omega_S = 0.4$ cm^2.

[3]Let $\mathcal{X}$ and $\mathcal{Y}$ be two partially ordered sets. If a real-valued function $f(x, y)$ is superadditive on $\mathcal{X} \times \mathcal{Y}$, then $f(x^+, y^+) - f(x^-, y^+) \ge f(x^+, y^-) - f(x^-, y^-)$, $\forall\, x^+, x^- \in \mathcal{X}$ and $\forall\, y^+, y^- \in \mathcal{Y}$ such that $x^+ \ge x^-$ and $y^+ \ge y^-$.

a monotonic structure in the composite policy indeed depends on various system parameters, although a sufficient condition regarding the energy harvesting is presented here. Fortunately, this elegant structure often appears according to our experimental observations.

VI. Simulation Results

Simulation results are presented in this section to evaluate the performance of the proposed data-driven transmission policies. In the system model, the numbers of solar states, battery states, channel states are set as four, twelve, and six, respectively. The data record of the irradiance collected by the solar site in EC in June from 2008 to 2012 is adopted throughout the simulation [25]. A four-state solar power harvesting model is trained using the data in the first three years, where the underlying parameters are given in Table II. The irradiance data of the subsequent two years are then applied for performance evaluation. Sensor communications usually require a bandwidth of a few hundreds of kHz to support a data rate of hundreds of kbps. In the system configuration, the symbol rate R_S is operated at 100 kHz, and a medium-sized packet of $L_S = 10^3$ data symbols is used. In other words, the packet duration T_P is given by 0.01 sec. Depending on sensor network applications, the transmission power typically ranges on the order of several tens of mW. Here, we set the basic transmission power level as $P_U = 40 \times 10^3$ μW. The modulation types could be QPSK, 8 PSK and 16 QAM. These three modulation types are considered as potential candidates for the composite policy, while only one modulation type is preselected for the on-off policy. To avoid frequent change of transmission actions, the policy management period is set to five minutes, i.e., $T_L = 300$ sec. In the value iteration algorithm, the discount factor λ and the stopping criterion ε are selected as 0.99 and 10^{-6}, respectively. Since the size of sensor nodes is small, the solar panel area is set as $\Omega_S = 1$ cm^2. From [2], the energy conversion efficiency is assumed to be $\vartheta = 20\%$. We assume that the battery state is randomly initialized. The channel quantization levels are randomly selected as $\mathbf{\Gamma}1 = \{0, 0.3, 0.6, 1.0, 2.0, 3.0, \infty\}$, and the optimization of quantization levels is beyond the scope of this paper. By assuming that sensor nodes are located in a rich-scattering environment, Jakes' model is applied to generate channel gains under a deterministic relative mobility between the transmitter and the receiver [34]. It is assumed that nodes have low mobility, and for a normalized Doppler frequency $f_D = 0.05$, the channel coherence time, $\frac{T_L}{2f_D}$, is around one hour. The above parameters are used as default settings, except as otherwise stated. Finally, since the average transmission power for a sensor is unknown and depends on real solar irradiance, a normalized average SNR γ_C is defined with respect to the transmission power of 10^3 μW throughout the simulation.

As a benchmark, two myopic policies are included for performance comparisons. For these two policies, the actions are performed without concern for the channel state and battery state transition probabilities, and data packets are transmitted as long as the battery storage is non-empty. The first policy (Myopic Policy I) attempts to transmit data packets at the lowest transmission power level, if the energy storage is positive. Regarding with the second one (Myopic Policy II), the largest available battery power is consumed for data transmission, if the battery state is non-zero. In addition, we compare the proposed schemes with a deterministic energy harvesting scheme in [19], called t-time fair rate assignment (t-TFR), which requires perfect knowledge of the channel fading and energy harvesting patterns for determining the optimal transmission power over a short-term period t to maximize the reward function in (19).

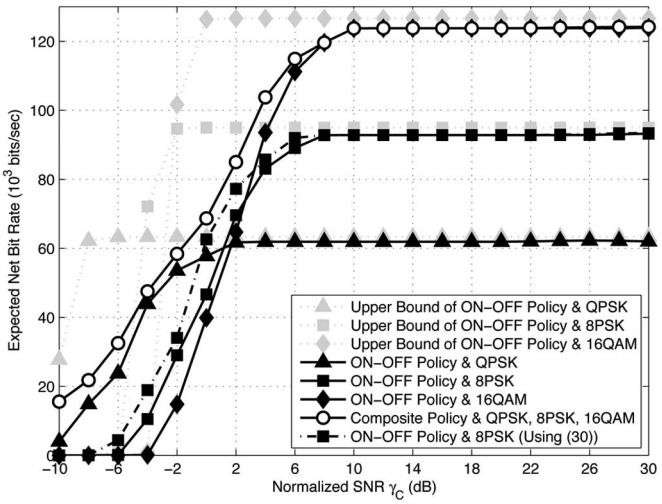

Fig. 7. Expected net bit rate versus normalized SNR γ_C for different transmission policies ($\Omega_S = 1$ cm^2, and $f_D = 0.05$).

Fig. 7 shows the expected net bit rates for the composite and on-off transmission policies. The expected net bit rate of the on-off policy is calculated according to (46), while that for the composite policy can be analyzed in a similar way although the accessible transmission actions appear to be more sophisticated. The performance upper bound of the on-off policy in (47) is also included for calibration purposes. For the on-off policy, it is observed that the expected net bit rate is monotonically increased with the operating SNRs, while the performance finally becomes saturated at 0.6×10^5 bits/sec, 0.9×10^5 bits/sec and 1.2×10^5 bits/sec for QPSK, 8 PSK and 16 QAM, respectively, when γ_C is sufficiently high. A saturation effect is observed because the BER becomes extremely small at high SNR regime and the net bit rate is thus limited by the permissible modulation schemes and the energy arrival rate. It is clear that the policy with QPSK modulation exhibits a better bit rate, as compared to 8 PSK and 16 QAM modulation when $\gamma_C \leq 2$ dB. On the contrary, it is advisable to employ high-level modulation schemes, e.g., 8 PSK and 16 QAM, to achieve better performance. This is because the adoption of high-level modulation schemes generally requires larger SNRs to guarantee a low packet error rate. As expected, the composite policy offers an expected net bit rate better than the on-off policy because it has more diversified actions, and the performance gap between these two policies could be as large as 60×10^3 bits/sec. However, the on-off policy with a mixture of QPSK and 16 QAM modulation can still achieve a large fraction of bit rate regions as available in the composite policy, and its simple implementation makes it attractive for practical applications. Besides, we demonstrate the exact performance for the on-off policy with 8 PSK by applying numerical integration in (18).

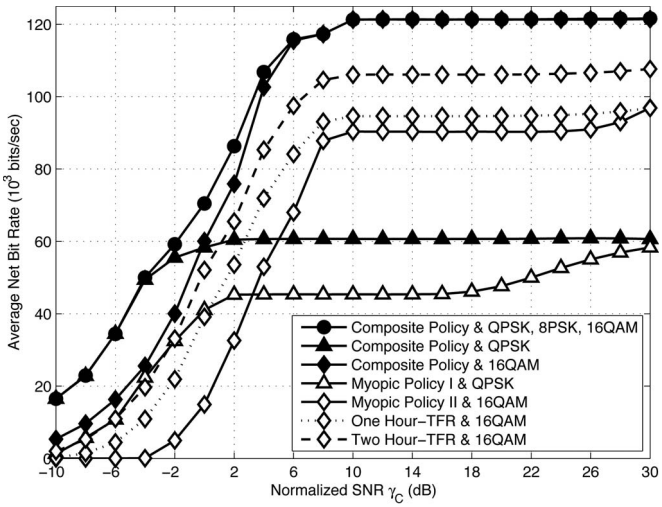

Fig. 8. Average net bit rate performances of the composite policy, Myopic Policy I, Myopic Policy II and t-TFR with the real data record of irradiance in June from 2011 to 2012, measured by a solar site in EC ($\Omega_S = 1$ cm^2, and $f_D = 0.05$).

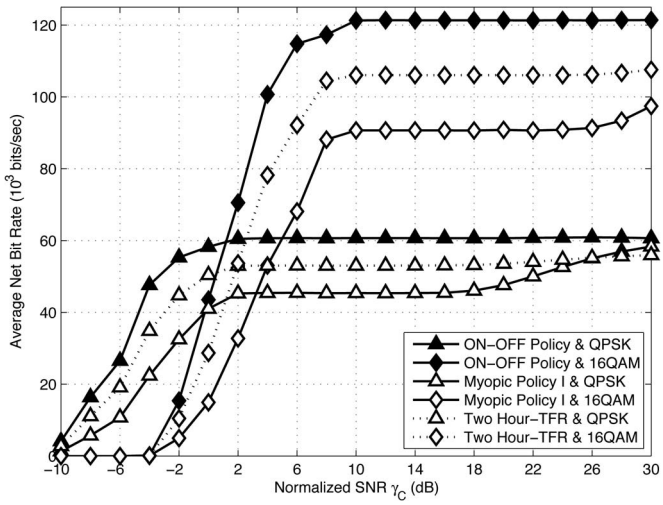

Fig. 9. Average net bit rate performances of the on-off and other benchmark policies ($\Omega_S = 1$ cm^2, and $f_D = 0.05$).

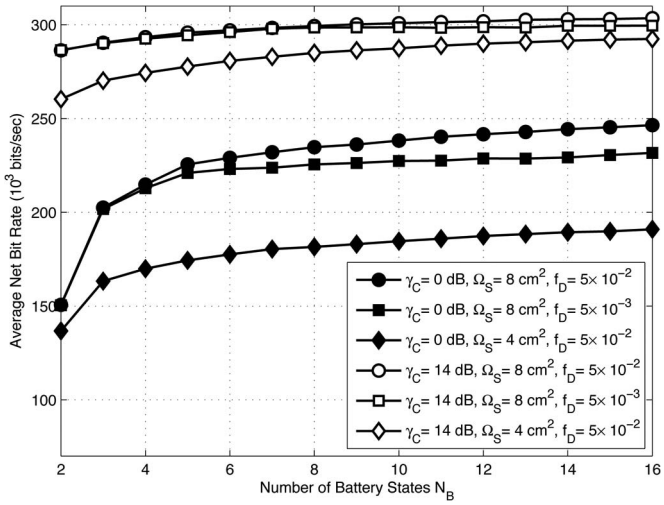

Fig. 10. Average net bit rate of the composite policy versus number of battery states under different Doppler frequencies and solar panel areas.

There is a minor gap between the exact performance without applying the lower bound and the expected net bit rate when γ_C is small, whereas the curves become identical at high SNRs. Similar results can be achieved for the proposed policies with other modulation types, although they are not shown in this figure.

Fig. 8 shows the average net bit rates of the proposed composite policy and other benchmark schemes, in which the real data record in EC from 2011 and 2012 is utilized to assess the performance. We can observe from this figure that Myopic Policy I with QPSK is superior to Myopic Policy II with 16 QAM in terms of the average net bit rates for low SNR regions, whereas the reverse trend is found for high SNR regions. This is because aggressive energy expenditure merits better bit rate performance when the operating SNR is high, and conservative use of energy is more preferable at low SNRs. Actually, the average net bit rate of Myopic Policy II with 16 QAM gets saturated at 1.2×10^5 bits/sec when $\gamma_C \geq 46$ dB, although this effect is not depicted. Moreover, the composite transmission policy is capable of achieving much better average net bit rates than these two myopic policies under the same modulation type. We can also find that the average net bit rate of the composite policy is superior to that of the t-TFR scheme, even if the energy harvesting and channel variation patterns are assumed to be perfectly predicted for one or two hours. Though the t-TFR scheme could attain better performance with an increased prediction interval, it suffers from the problems of larger prediction error and higher computational complexity for a long prediction interval. Finally, the composite policy in conjunction with the three modulation types has much better performance than that with a single modulation type.

The average net bit rate of the on-off transmission policy is shown in Fig. 9 for different modulation types. Moreover, the performances of the Myopic Policy I and the Two Hour-TFR schemes, in conjunction with various modulation types, are included in this figure. To make a fair comparison, the t-TFR scheme also adopts on-off power actions for the short-term scheduling of energy expenditure. It can be seen that the maximum spectrum efficiency provided by our proposed on-off policy is approximately given by 0.6 bits/sec/Hz and 1.2 bits/sec/Hz for QPSK and 16 QAM, respectively. With a fixed modulation scheme, the on-off policy offers significant performance gains over the myopic policy by taking advantage of channel fluctuation gains. A closer look at this figure reveals that the performance gap between these two policies becomes wider as the modulation level increases. When compared with the Two Hour-TFR scheme, the on-off policy can still achieve better average net bit rates, no matter which modulation type is used.

Fig. 10 illustrates the average net bit rate of the composite policy as a function of the number of battery states. To clearly understand the relationship between the Doppler frequency and the battery storage capacity, the normalized Doppler frequency, f_D, is chosen as 0.005 and 0.05. We can observe that the average net bit rate can be dramatically enhanced by enlarging the energy buffer size to store more energy quanta, especially when the operating SNR is low. For instance, the performance

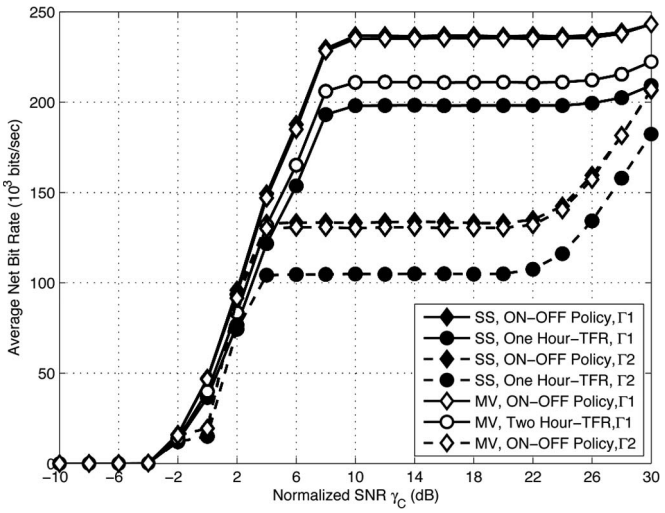

Fig. 11. Average net bit rate of the on-off policy for other data records in different locations/months ($N_B = 12$, $\Omega_S = 4$ cm^2, $f_D = 0.05$, and 16 QAM).

with $N_B = 16$ at $\gamma_C = 0$ dB, $\Omega_S = 8$ cm^2 and $f_D = 0.05$ is about 2.5×10^5 bits/sec, probably 1.6 times that being achieved by the same policy with $N_B = 2$. Also, the bit rate is increased with the increase of the solar panel area due to a higher energy harvesting rate. The sensor node additionally benefits from channel fluctuation gains if the energy spending is carefully governed to respond to the change in channel conditions, and the bit rate becomes better as f_D increases.

Fig. 11 shows the average net bit rate for the data records, measured by two solar sites in SS (in October from 1998 to 2002) and Mississippi Valley State University (MV) (in March from 2000 to 2004) [25]. Another channel quantization $\mathbf{\Gamma}2 = \{0, 1.0, 3.0, \infty\}$ $(\subseteq \mathbf{\Gamma}1)$ with $N_C = 3$ is applied here. With the same quantization set, the proposed scheme can still outperform the One/Two Hour-TFR schemes in different locations and months. When comparing the results from different quantization sets, one can see that the performance can be further improved by partitioning the channel into a larger number of channel states.

VII. Conclusion

In this paper, we have studied the problem of maximizing long-term net bit rates in sensor communication that solely relies on solar energy for data transmission. A node-specific energy harvesting model was developed to classify the harvesting conditions into several solar states with different energy quantum arrivals. Unlike previous works, which were not concerned with the real-world energy harvesting capability, a data-driven MDP framework was formulated to obtain the optimal transmission parameters from a set of power and modulation actions in response to the dynamics of channel fading and battery storage. Since different nodes may possess different energy harvesting capabilities, the parameters of the underlying energy harvesting process were completely determined by the solar irradiance observed at a sensor node. In practice, the exact solar state at each time epoch is unavailable, and a mixed strategy was proposed to associate the adaptive transmission parameters with the beliefs of the solar states. The validity of the proposed data-driven approach was rigorously justified by the real data of solar irradiance. We also analyzed the properties and the net bit rates of the optimal on-off transmission policy, and it was proved that this policy has an inherent threshold structure in the direction along the battery states. Through extensive computer simulations, the proposed data-driven approach was shown to achieve significant gains with respect to other radical approaches, while it did not require non-causal knowledge of energy harvesting and channel fading patterns. As a final remark, this work can be served as an important step for investigating other upper-layer issues in energy harvesting sensor networks, e.g., wake-up and sleep cycles and routing protocols, that involve more sophisticated settings with practical considerations in the future.

Appendix A
Effect of f_D and N_B on the Performance

We explain the idea of how the parameters f_D and N_B affect the performance by considering a simple model. Assume that $N_C = 2$ and $N_H = 1$, and the solar energy periodically arrives at a constant rate f_E, i.e., the sensor can harvest one energy quantum every $T_E (\propto 1/f_E)$ (time unit: T_L). Moreover, the channel alternates between the two states every $T_C (\propto 1/f_D)$ time units. It is undoubted that the more the energy is harvested, the better the performance is; however, a capacity-limited battery will cause an energy overflow problem and reduce the chance to harvest energy. Thus, when the sensor knows the capacity of its battery is going to be saturated, it should at least spend one energy quantum for data transmission, even if the current channel is bad. Actually, there is a tradeoff for the sensor between the risk of energy overflow and the chance of transition to good channel states. Since the sensor will take the opportunity to transmit data in the good channel as long as the battery is nonempty, we focus on the situation in the bad channel as follows. If the current battery state is y and $(N_B - 1 - y)T_E < T_C$, the sensor is impelled to spend the energy in the bad channel, even though the obtained bit rate is low. On the other hand, if $(N_B - 1 - y)T_E \geq T_C$, the sensor is expected to experience several times of good channels during $(N_B - 1 - y)T_E$, and thus, the harvested energy is spent more efficiently to achieve a higher bit rate as N_B and f_D increase.

References

[1] C. Pandana and K. J. R. Liu, "Near optimal reinforcement learning framework for energy-aware wireless sensor communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 4, pp. 788–797, Apr. 2005.

[2] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: Survey and implications," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 443–461, 2011.

[3] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 6, no. 4, pp. 32/1–38/1, Sep. 2007.

[4] M. Tacca, P. Monti, and A. Fumagalli, "Cooperative and reliable ARQ protocols for energy harvesting wireless sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 2519–2529, Jul. 2007.

[5] S. Reddy and C. R. Murthy, "Profile-based load scheduling in wireless energy harvesting sensors for data rate maximization," in *Proc. IEEE Int. Conf. Commun.*, 2010, pp. 1–5.

[6] D. Niyato, E. Hossain, and A. Fallahi, "Sleep and wakeup strategies in solar-powered wireless sensor/mesh networks: Performance analysis and

optimization," *IEEE Trans. Mobile Comput.*, vol. 6, no. 2, pp. 221–236, Feb. 2007.
[7] B. Medepally, N. B. Mehta, and C. R. Murthy, "Implications of energy profile and storage on energy harvesting sensor link performance," in *Proc. IEEE Globe Commun. Conf.*, 2009, pp. 1–6.
[8] N. Michelusi, K. Stamatiou, and M. Zorzi, "On optimal transmission policies for energy harvesting devices," in *Proc. IEEE Inf. Theory Appl. Workshop*, 2012, pp. 249–254.
[9] N. Michelusi and M. Zorzi, "Optimal random multiaccess in energy harvesting wireless sensor networks," in *Proc. IEEE Int. Conf. Commun.*, 2013, pp. 463–468.
[10] A. Aprem, C. R. Murthy, and N. B. Mehta, "Transmit power control policies for energy harvesting sensors with retransmissions," *IEEE J. Sel. Topics Signal Process.*, vol. 7, no. 5, pp. 895–906, Oct. 2013.
[11] K. J. Prabuchandran, S. K. Meena, and S. Bhatnagar, "Q-learning based energy management policies for a single sensor node with finite buffer," *IEEE Wireless Commun. Lett.*, vol. 2, no. 1, pp. 82–85, Feb. 2013.
[12] J. Lei, R. Yates, and L. Greenstein, "A generic model for optimizing single-hop transmission policy of replenishable sensors," *IEEE Trans. Wireless Commun.*, vol. 8, no. 2, pp. 547–551, Feb. 2009.
[13] S. Mao, M. H. Cheung, and V. W. S. Wong, "An optimal energy allocation algorithm for energy harvesting wireless sensor networks," in *Proc. IEEE Int. Conf. Commun.*, 2012, pp. 265–270.
[14] M. Kashef and A. Ephremides, "Optimal packet scheduling for energy harvesting sources on time varying wireless channels," *J. Commun. Netw.*, vol. 14, no. 2, pp. 121–129, Apr. 2012.
[15] Z. Wang, A. Tajer, and X. Wang, "Communication of energy harvesting tags," *IEEE Trans. Commun.*, vol. 60, no. 4, pp. 1159–1166, Apr. 2012.
[16] H. Li, N. Jaggi, and B. Sikdar, "Cooperative relay scheduling under partial state information in energy harvesting sensor networks," in *Proc. IEEE Globe Commun. Conf.*, 2010, pp. 1–5.
[17] N. Michelusi, K. Stamatiou, and M. Zorzi, "Transmission policies for energy harvesting sensors with time-correlated energy supply," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2988–3001, Jul. 2013.
[18] C. K. Ho, P. D. Khoa, and P. C. Ming, "Markovian models for harvested energy in wireless communications," in *Proc. IEEE Int. Conf. Commun. Syst.*, 2010, pp. 311–315.
[19] M. Gorlatova, A. Wallwater, and G. Zussman, "Networking low-power energy harvesting devices: Measurements and algorithms," *IEEE Trans. Mobile Comput.*, vol. 12, no. 9, pp. 1853–1865, Sep. 2013.
[20] P. S. Khairnar and N. B. Mehta, "Power and discrete rate adaptation for energy harvesting wireless nodes," in *Proc. IEEE Int. Conf. Commun.*, 2011, pp. 1–5.
[21] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.
[22] S. Zhang, A. Seyedi, and B. Sikdar, "An analytical approach to the design of energy harvesting wireless sensor nodes," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4010–4024, Aug. 2013.
[23] Q. Wang and M. Liu, "When simplicity meets optimality: Efficient transmission power control with stochastic energy harvesting," in *Proc. IEEE INFOCOM*, 2013, pp. 580–584.
[24] T. Zhang, W. Chen, Z. Han, and Z. Cao, "A cross-layer perspective on energy harvesting aided green communications over fading channels," in *Proc. IEEE INFOCOM*, 2013, pp. 3225–3230.
[25] NREL. (2012, Feb.). Solar radiation resource information, Golden, CO, USA. [Online]. Available: http://www.nrel.gov/rredc/
[26] R. D. Rugescu, *Solar Power*. Rijeka, Croatia: InTech, 2012.
[27] P. L. Zervas, H. Sarimveis, J. A. Palyvos, and N. C. G. Markatos, "Prediction of daily global solar irradiance on horizontal surfaces based on neural-network techniques," *Renew. Energy*, vol. 33, no. 8, pp. 1796–1803, Aug. 2008.
[28] J. A. Bilmes, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," Int. Comput. Sci. Inst., Berkeley, CA, USA, Tech. Rep. ICSI-TR-97-021, Apr. 1998.
[29] H. S. Wang and N. Moayeri, "Finite-state Markov channel—A useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–171, Feb. 1995.
[30] J. Lu, K. B. Letaief, J. C.-I. Chuang, and M. L. Liou, "M-PSK and M-QAM BER computation using signal-space concepts," *IEEE Trans. Commun.*, vol. 47, no. 2, pp. 181–184, Feb. 1999.
[31] K. Cho and D. Yoon, "On the general BER expression of one- and two-dimensional amplitude modulations," *IEEE Trans. Commun.*, vol. 50, no. 7, pp. 1074–1080, Jul. 2002.
[32] D. P. Bertsekas, *Dynamic Programmming and Optimal Control*, 3rd ed. Belmonth, MA, USA: Athena Scientific, 2005.
[33] M. L. Puterman, *Markov Decision Processes-Discrete Stochastic Dynamic Programming*. New York, NY, USA: Wiley, 1994.
[34] W. C. Jakes, *Microwave Mobile Communications*. New York, NY, USA: Wiley, 1974.

Meng-Lin Ku (M'11) received the B.S., M.S., and Ph.D. degrees in communication engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2002, 2003, and 2009, respectively. Between 2009 and 2010, he was a Postdoctoral Research Fellow with Prof. Li-Chun Wang in the Department of Electrical and Computer Engineering, National Chiao Tung University and with Prof. Vahid Tarokh in the School of Engineering and Applied Sciences, Harvard University. In August 2010, he became a Faculty Member of the Department of Communication Engineering, National Central University, Taoyuan City, Taiwan, where he is currently an Assistant Professor. During the summer of 2013, he was a Visiting Scholar in the Signals and Information Group of Prof. K. J. Ray Liu at the University of Maryland. Dr. Ku was a recipient of the Best Counseling Award in 2012 and the Best Teaching Award in 2013 and 2014 at National Central University. He was also the recipient of the Exploration Research Award of the Pan Wen Yuan Foundation, Taiwan, in 2013. His current research interests are in the areas of green communications, cognitive radios, and optimization of radio access.

Yan Chen (SM'14) received the bachelor's degree from University of Science and Technology of China in 2004, the M.Phil. degree from Hong Kong University of Science and Technology (HKUST) in 2007, and the Ph.D. degree from University of Maryland College Park in 2011. His current research interests are in data science, network science, game theory, social learning and networking, as well as signal processing and wireless communications.

He is the recipient of multiple honors and awards including best paper award from IEEE GLOBECOM in 2013, Future Faculty Fellowship and Distinguished Dissertation Fellowship Honorable Mention from Department of Electrical and Computer Engineering in 2010 and 2011, respectively, Finalist of Deans Doctoral Research Award from A. James Clark School of Engineering at the University of Maryland in 2011, and Chinese Government Award for outstanding students abroad in 2011.

K. J. Ray Liu (F'03) was named a Distinguished Scholar-Teacher of University of Maryland, College Park, in 2007, where he is Christine Kim Eminent Professor of Information Technology. He leads the Maryland Signals and Information Group conducting research encompassing broad areas of signal processing and communications with recent focus on cooperative and cognitive communications, social learning and network science, information forensics and security, and green information and communications technology.

He was a Distinguished Lecturer, recipient of IEEE Signal Processing Society 2009 Technical Achievement Award, 2014 Society Award, and various best paper awards. He also received various teaching and research recognitions from University of Maryland including university-level Invention of the Year Award; and Poole and Kent Senior Faculty Teaching Award, Outstanding Faculty Research Award, and Outstanding Faculty Service Award, all from A. James Clark School of Engineering. An ISI Highly Cited Author, Dr. Liu is a Fellow of AAAS.

Dr. Liu was President of IEEE Signal Processing Society (2012–13) where he has served as Vice President—Publications and Board of Governor. He was the Editor-in-Chief of IEEE Signal Processing Magazine and the founding Editor-in-Chief of EURASIP Journal on Advances in Signal Processing.