Optimal Buffer Management Using Hybrid Systems

Wei Zhang and Jianghai Hu

***Abstract*— This paper studies a class of dynamic buffer management problems with one buffer inserted between two interacting components. The overall system is modeled as a hybrid system and the power minimization problem is formulated as an optimal control problem. Different from many previous studies, the objective function of the proposed problem depends on the switching cost and the size of the continuous state space, making its solution much more challenging. A simplified version of the problem with some extra assumptions is solved in [1]. This paper relaxes all these assumptions and gives a complete solution to the proposed problem using some variational approach. Simulation result shows that the proposed method can save about 60% energies compared with some heuristic schemes.**

I. Introduction

Dynamic buffer management (DBM) is an effective power management scheme that can reduce the power consumptions of electronic devices by inserting buffers among interacting components. The buffer insertion makes it possible to turn off underutilized components at appropriate times without affecting the services for other components, thus reducing the power consumptions. The optimal buffer size resulting in the largest power reduction is derived in [2]. The result is further extended to the case where two buffers are inserted between three streamlined components [3] [4]. A major limitation of these previous studies is that they all assume that the components to be controlled have only two power modes, “on” and “off”. However, in practice, many components can work in more than two power modes, such as variable speed processors [5] and multi-speed disks [6]. For such a component, instead of completely turning it off, we can properly design a switching strategy, namely the scheduling of different power modes of the component, to further reduce the overall power consumption.

This paper studies a more general DBM problem, where the component to be controlled has multiple power modes. Since different power modes correspond to different data changing rates in the buffer, the overall system is perfectly modeled as a piecewise-constant hybrid system, or more accurately, a multi-rate automata [7]. The DBM problem is thus formulated as an optimal control problem of the underlying hybrid system. Despite the richness of the literature in optimal control of hybrid systems [8], [9], [10], [11], previous results cannot be directly applied to our problem as it has the following distinct features: (1) the discrete mode transition depends on the evolution of the continuous state; whereas most previous studies ignore such dependence; (2) the switching (mode) sequence is a decision variable that cannot be assumed fixed as in [10], [9]; (3) the switching cost ignored in most previous papers is an important part of our cost function; (4) The buffer size that determines the range of the continuous states is variable, indicating that both the optimal control and the optimal size of the state space are to be designed at the same time. Few existing results have addressed all of the above issues. A simplified version of our problem with some extra assumptions is solved in [1]. The main contribution of this paper is that we relax all the previous assumptions in [1] and give a complete analytical solution to the proposed problem, despite the challenges mentioned above.

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA. Email:{zhang70, jianghai}@purdue.edu. This work was partially supported by the National Science Foundation under Grant CNS-0643805.

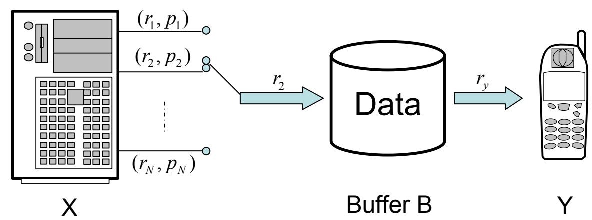


Fig. 1. System Configuration

This paper is organized as follows. In Section II, we formulate the optimal control problem to be studied in this paper. In Section III, some operations on hybrid trajectories are introduced and used in deriving the optimal solutions to our problem. Simulation results are given in Section IV to illustrate the effectiveness of the proposed method. Conclusion remarks and future research directions are discussed in Section V.

II. Problem Formulation

A. System Description

Consider two interacting components X and Y as shown in Fig. 1, where X produces data for Y to consume. Suppose Y is always “on” and consumes data at a constant speed r_y. On the other hand, assume that X has N different operation modes where in mode i, $i = 1, 2, \ldots, N$, it produces data at a constant speed r_i and consumes energy at the rate of p_i. Without loss of generality, assume $r_1 < r_2 < \cdots < r_N$. Usually, a lower data processing rate corresponds to a lower power consumption; thus we require $p_1 < p_2 < \cdots < p_N$. Denote by I and J the sets of indices whose corresponding data rates are greater and smaller than r_y, respectively, i.e.,

$$I = \{i \,|\, r_i > r_y, i = 1, \ldots, N\},$$
$$\text{and} \quad J = \{j \,|\, r_j < r_y, j = 1, \ldots, N\}.$$

Assume that both I and J are nonempty, i.e., $r_N > r_y > r_1$. A mode σ is called an *ascending mode* if $\sigma \in I$ and a *descending mode* otherwise. To ensure smooth operation, a buffer B with capacity Q is inserted between X and Y. See Fig. 1 for the configuration of the overall system.

B. Hybrid System Model

The above problem can be modeled as a hybrid system H. The discrete state space of the hybrid system consists of N modes: $S = \{1, 2, \dots, N\}$, representing all the operation modes of X. The continuous state $q(t)$ is defined as the amount of data stored in the buffer B, and is thus required to take values in the interval $[0, Q]$. The evolution of $q(t)$ is determined by the speed difference between the two components, i.e., $\dot{q}(t) = r_i - r_y$ for mode i. As a physical constraint, there should be no buffer underflow or overflow. Thus we require that whenever $q(t)$ hits the boundary of its domain, namely, $q(t) = 0$ or Q, the system must transit to another mode that can bring $q(t)$ back to the inside of $[0, Q]$. Except for this, there are no other transition rules or guard conditions. The reset map of the system is trivial, i.e., there is no jump in $q(t)$ at the transition instant.

Given a time period $[0, t_f]$, the behavior of the above system can be uniquely determined by the switching strategy $\sigma : [0, t_f] \to S$, which determines the active mode of the system over $t \in [0, t_f]$. The overall trajectory $z(t) = (q(t), \sigma(t))$ of the hybrid system consists of the trajectories of both the continuous state $q(t)$ and the discrete state $\sigma(t)$. For a given initial state $q(0)$, the system is governed by the following differential equation:

$$\frac{dq(t)}{dt} = r_{\sigma(t)} - r_y, \qquad \forall t \in [0, t_f]. \tag{1}$$

In this paper, we study the power consumption of the whole process of transferring a certain amount of data from X to Y. It is thus required that the system must start with an empty buffer at $t = 0$ and end up with an empty buffer at $t = t_f$ when Y have received all the data produced by X. This yields two boundary conditions for the continuous state, namely, $q(0) = 0$ and $q(t_f) = 0$. The hybrid trajectories that satisfy these two conditions are called *feasible trajectories*. (See Fig. 2-(a))

We assume that there is a partition of $[0, t_f]$, $t_0 = 0 \le t_1 \le \ldots \le t_n = t_f$, for some $n \ge 0$, so that $\sigma(t) \equiv \sigma_i \in S$ is constant in each subinterval $[t_{i-1}, t_i)$, $i = 1, \dots, n$. The sequence $(\sigma_1, \dots, \sigma_n)$ is called the *switching sequence* and $(t_0, \dots, t_{n-1})$ is called the *switching instants*[1].

A feasible trajectory is called a Λ-*trajectory* if it consists one ascending mode i and one descending mode j with exactly two switchings as shown in Fig 2-(b). The pair of modes $\{i, j\}$ in a Λ-trajectory is called a Λ-*pair*.

A feasible trajectory $z(t) = (q(t), \sigma(t))$ with switching instants $(t_0, \dots, t_{n-1})$ is called a *boundary-switching trajectory (BST)* if $q(t_i) = Q$ or 0 for any $i = 0, \dots, n$. In

[1]The system is turned on at $t = 0$. Hence, we assume that there is always a switching at $t = 0$ and ignore the switching, if any, at $t = t_f$ for all trajectories.

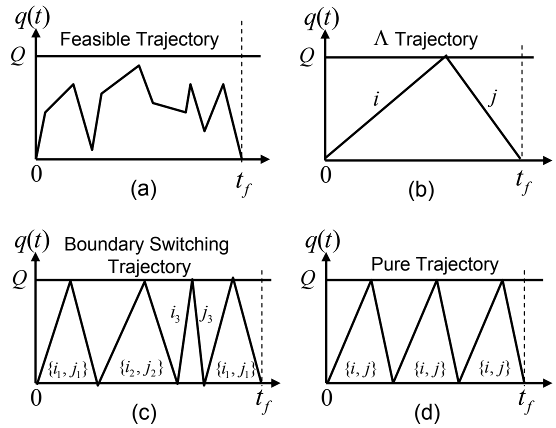


Fig. 2. Illustration of different hybrid trajectories

other words, a BST only switches at the boundary of the range of $q(t)$. Denote by Ω the class of all BSTs. Every BST can be decomposed into a series of Λ-trajectories with the same buffer size. Denote by n_p the number of *distinct* Λ-pairs in a BST. For example, for the BST in Fig 2-(c), we have $n_p = 3$. A BST is called *pure* if $n_p = 1$ and is called *mixed* otherwise. In other words, a pure trajectory must be a BST and is obtained by repeating a Λ-trajectory for a certain number of times. (See Fig. 2-(d)).

The power consumption of a given hybrid trajectory $z(t) = (q(t), \sigma(t))$ consists of three parts: the *running power*, namely the power consumed by component X[2], the *switching power* and the *buffer power*. Since $\sigma(t) \in S$ determines the active power mode of X, $p_{\sigma(t)}$ is the instantaneous power of X at time t. Thus the average running power over $[0, t_f]$ is $\frac{1}{t_f}\int_0^{t_f} p_{\sigma(t)} dt$. Assume that switching among different modes costs the same amount of energy k_s[3]. Then the average switching power over $[0, t_f]$ is nk_s/t_f, where n is the number of switchings in the trajectory z. The buffer power is proportional to the buffer size Q and denoted by $p_b Q$, where p_b is a positive constant. Therefore, the total average power of the system during $[0, t_f]$ can be written as

$$\bar{P}(z; Q, t_f) = \frac{1}{t_f}\int_0^{t_f} p_{\sigma(t)} dt + \frac{nk_s}{t_f} + p_b Q. \tag{2}$$

Similarly, the total energy associated with $z(t)$ during $[0, t_f]$ is

$$E(z; Q, t_f) = \int_0^{t_f} p_{\sigma(t)} dt + nk_s + p_b Q \cdot t_f.$$

The three terms on the right hand side of the above equation represent the *running energy*, the *switching energy*, and the *buffer energy*, respectively.

C. Problem Statements

The main purpose of this paper is to find a feasible trajectory that can finish a given task with the least energy

[2]The power of Y is ignored in this paper since it is a constant independent of the switching strategy.

[3]There may exist other switching penalties, such as the switching delay penalty. To simplify discussion, we assume all the switching penalties are transformed to an equivalent energy cost and incorporated into k_s.

consumption. This problem can be formulated as the following optimal control problem of the hybrid system H.

Problem 1: Find a feasible trajectory $z(t)$ over $[0, t_f]$ and a proper buffer size Q that

$$\text{minimize} \quad \bar{P}(z; Q, t_f) = \frac{1}{t_f}\left(\int_0^{t_f} p_{\sigma(t)} dt + nk_s\right) + p_b Q$$

$$\text{subject to} \quad \frac{dq(t)}{dt} = r_{\sigma(t)} - r_y, \quad \text{with} \quad q(0) = 0, \tag{3}$$

$$q(t_f) = 0, \max_{t\in[0,t_f]} q(t) \le Q, \text{ and } \min_{t\in[0,t_f]} q(t) \ge 0. \tag{4}$$

A simplified version of Problem 1 is solved in [1] under two additional assumptions: (a) t_f approaches infinity, and (b) z is periodic. In next section, we relax both of these assumptions and give a complete analytical solution to Problem 1.

III. OPTIMAL SOLUTIONS

A. Operations on Hybrid Trajectories

We first introduce some useful operations on hybrid trajectories: (i) Denote by $\mathcal{C}_{a,b}[z]$ the ***cropping*** operation that only keeps the part of z within the time interval $[a, b]$, i.e., $\mathcal{C}_{a,b}[z](t) = z(t+a)$, for $t \in [0, b-a]$. (ii) Let $z_1 = (q_1, \sigma_1)$ and $z_2 = (q_2, \sigma_2)$ be two hybrid trajectories with length t_{f1} and t_{f2}, respectively. Denote by $\mathcal{J}[z_1, z_2]$ the ***joining*** operation, which obtains a new trajectory by appending z_2 to the end of z_1. To preserve the continuity of the continuous state, the trajectories to be joined must have consistent boundary conditions, i.e., $q_1(t_{f1}) = q_2(0)$. In particular, suppose z_1 satisfies that $q_1(0) = q_1(t_{f1})$. Denote by $\mathcal{J}_m[z_1]$ a special joining operation that repeats z_1 for m times. (iii) The ***scaling*** operation of z with parameter c is defined as: $\mathcal{S}_c[z] = (cq(t/c), \sigma(t/c))$. The scaling operation does not change the switching sequence; however, the switching instants, the optimal buffer size and the length of z becomes c times the original values.

B. Necessary Conditions

To avoid the unnecessary power consumption by the unused buffer space, Q should be chosen as small as possible so that the buffer is full at least once during $[0, t_f]$. Thus the following lemma follows immediately.

Lemma 1 (Tightness condition): If $z(t) = (q(t), \sigma(t))$ is an optimal solution to Problem 1 (OS1), we must have

$$\min_{t\in[0,t_f]} q(t) = 0, \quad \text{and} \quad \max_{t\in[0,t_f]} q(t) = Q.$$

The following lemma is the key result of this paper and can greatly simplify the derivation of the optimal solution to Problem 1 (OS1).

Lemma 2: If z is an OS1, then $z \in \Omega$.

Proof: Let $z(t)$ be an OS1 with switching sequence $(\sigma_1, \ldots, \sigma_n)$ and switching instants $(t_0, \ldots, t_{n-1})$. Suppose that $z(t)$ has a switching at some interior point of $[0, Q]$, i.e., $0 < q(t_i) < Q$ for some i. Divide $z(t)$ into three parts through the cropping operation as shown in Fig. 3-(a) and define them as three new trajectories.

$$z^{(1)}(t) = \mathcal{C}_{0,t_{i-1}}[z](t), \quad z^{(2)}(t) = \mathcal{C}_{t_{i-1},t_{i+1}}[z](t),$$
$$\text{and} \quad z^{(3)}(t) = \mathcal{C}_{t_{i+1},t_f}[z](t). \tag{5}$$

Assume that $z^{(2)}(t) = (q^{(2)}(t), \sigma^{(2)}(t))$, $q^{(2)}(0) = q_1$ and $q^{(2)}(\tau_i) = q_2$, where $\tau_i = t_i - t_{i-1}$. Thus $z^{(2)}$ has an interior switching from σ_i to σ_{i+1} at instant τ_i. Perturb this switching instant to a neiboring value h and define a perturbed trajectory $z_h^{(2)} = (q_h^{(2)}(t), \sigma_h^{(2)}(t))$ as

$$\sigma_h^{(2)}(t) = \begin{cases} \sigma_i & t \le h \\ \sigma_{i+1} & h < t \le t_h^{(2)} \end{cases},$$
$$\frac{dq_h^{(2)}(t)}{dt} = r_{\sigma_h^{(2)}(t)} - r_y, \quad \text{for } t \in [0, t_h^{(2)}], \tag{6}$$
$$\text{and} \quad t_h^{(2)} = h + \frac{h(r_y - r_{\sigma_i}) + q_2 - q_1}{r_{\sigma_{i+1}} - r_y}.$$

As illustrated in Fig. 3-(b), $z_h^{(2)}$ switches from σ_i to σ_{i+1} at time h instead of τ_i and ends at $t_h^{(2)}$ where the new continuous state evolves to q_2. An important property of $q_h^{(2)}$ is that it has the same initial and final value as $q^{(2)}(t)$, i.e., $q_h^{(2)}(0) = q_1$ and $q_h^{(2)}(t_h^{(2)}) = q_2$. Thus we can join $z_h^{(2)}$ with $z^{(1)}$ and $z^{(3)}$ to obtain $z_h \triangleq \mathcal{J}[z^{(1)}, z_h^{(2)}, z^{(3)}]$. It is obvious that the length of z_h is

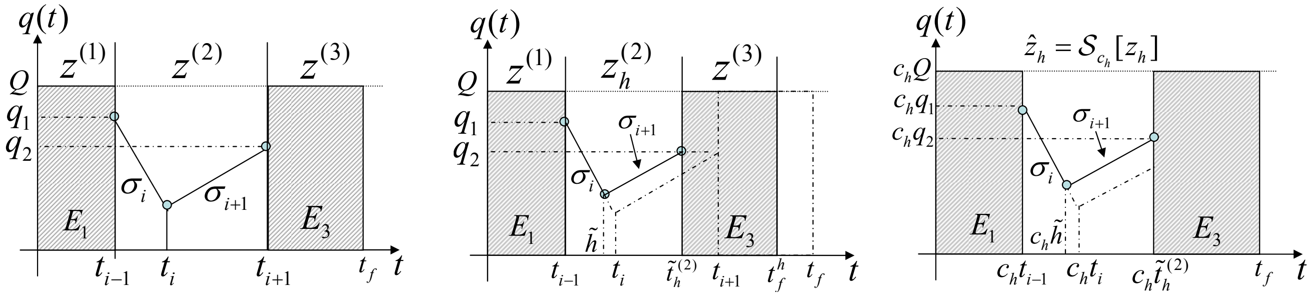
$$t_f^h = t_{i-1} + t_h^{(2)} + (t_f - t_{i+1}). \tag{7}$$

Since $q(t_i) \in (0, Q)$, there exist $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that $\forall h \in D_h \triangleq [\tau_i - \epsilon_1, \tau_i + \epsilon_2]$, $z_h(t)$ stays inside $[0, Q]$ all the time. Thus $\forall h \in D_h$, z_h is a properly defined trajectory for the original buffer size Q and the average power of z_h is

$$\bar{P}(z_h; Q, t_f^h) = \frac{1}{t_f^h}\big[E_1 + E_3 + nk_s + p_{\sigma_i} h$$
$$+ p_{\sigma_{i+1}}(t_h^{(2)} - h)\big] + p_b Q, \tag{8}$$

where E_1 and E_3 are the running energy of $z^{(1)}$ and $z^{(3)}$, respectively. The above perturbation on t_i results in a z_h with length $t_f^h \neq t_f$. To make z_h feasible for Problem 1, define $\hat{z}_h = \mathcal{S}_{c_h}[z_h]$ as shown in Fig. 3-(c), where $c_h = t_f / t_f^h$. According to the properties of the scaling operation, the buffer size of $\hat{z}_h$ becomes cQ and the length of $\hat{z}_h$ is changed back to t_f. Therefore, $\hat{z}_h$ is a feasible trajectory for Problem 1. The average power of $\hat{z}_h$ with buffer $c_h Q$ can be easily computed as:

$$\bar{P}(\hat{z}_h; Q, t_f) = \frac{1}{t_f^h}\big[E_1 + E_3 + 2k_s + p_{\sigma_i} \cdot h$$
$$+ p_{\sigma_{i+1}}(t_h^{(2)} - h)\big] + p_b c_h Q + \frac{nk_s(1 - c_h)}{c_h t_f^h}.$$



(a) $z(t)$ with interior switching (b) Perturbed $z_h(t)$ (c) $\hat{z}_h$ with length t_f

Fig. 3. Illustration of the proof of Lemma 2, where $\tilde{h} = h + t_{i-1}$, $\tilde{t}_h^{(2)} = t_h^{(2)} + t_{i-1}$.

Taking the derivative of $\bar{P}(\hat{z}_h; Q, t_f)$ with respect to h, we have

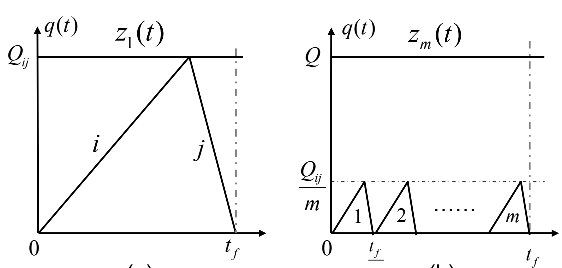
$$\begin{aligned}\frac{d\bar{P}(\hat{z}_h; Q, t_f)}{dh} &= \frac{1}{\left(t_f^h\right)^2}\left[\left(p_{\sigma_i} + p_{\sigma_{i+1}}\frac{r_y - r_{\sigma_i}}{r_{\sigma_{i+1}} - r_y}\right)\right.\\ &\cdot\left(t_f - (t_{i+1} - t_{i-1}) + \frac{q_2 - q_1}{r_{\sigma_{i+1}} - r_y}\right) - \left(1 + \frac{r_y - r_{\sigma_i}}{r_{\sigma_{i+1}} - r_y}\right)\\ &\cdot\left.\left(E_1 + E_3 + k_s + p_{\sigma_{i+1}}\frac{q_2 - q_1}{r_{\sigma_{i+1}} - r_y}\right)\right]\\ &+ \frac{nk_s - p_b Q t_f}{(t_f^h)^2}\left(1 - \frac{r_y - r_{\sigma_i}}{r_{\sigma_{i+1}} - r_y}\right).\end{aligned} \tag{9}$$

Note that the h-related terms in the numerator have been cancelled out. From (9) it is clear that the sign of $\frac{d\bar{P}(z_h; Q, T_h)}{dh}$ does not depend on h, which indicates that $\bar{P}(\hat{z}_h; Q, t_f^h)$ is monotone with respect to h in $D_h = [\tau_i - \epsilon_1, \tau_i + \epsilon_2]$. Thus either $z_{\tau_i - \epsilon_1}$ or $z_{\tau_i - \epsilon_2}$ consumes a less power than z. Therefore, we conclude that $q(t_i) = 0$ or Q for all $i = 1, \ldots, n$, i.e., the OS1 must be a BST. ■

Lemma 2 allows us to consider only the BSTs in deriving the OS1s. Recall that the variable n_p is used to describe the purity of a BST. In the rest of this section, we will first solve a simple case of Problem 1 where only pure BSTs with $n_p = 1$ are considered as candicate solutions. Then we will prove that the optimal solution in this simple case is also an OS1 for an arbitrary n_p.

C. Optimal Pure Trajectory

In this section we derive analytical solutions to Problem 1 with an additional constraint $n_p = 1$, i.e., we only consider pure trajectories as candidate solutions. We call the solution to Problem 1 under this additional constraint the *optimal pure trajectory (OPT)*.



(a) (b)

Fig. 4. Obtaining z_m from z_1

The simplest pure trajectory is the Λ-trajectory. Let $z_1(t)$ be a Λ-trajectory as shown in Fig. 4-(a) with length t_f and Λ-pair $\{i, j\}$. Since t_f is fixed, its optimal buffer size is given by:

$$t_f = \frac{Q_{ij}}{r_i - r_y} + \frac{Q_{ij}}{r_y - r_j} \Rightarrow Q_{ij} \triangleq \frac{tf}{\alpha_{ij}}, \tag{10}$$

where α_{ij} is defined as the ratio of the time horizon T_{ij} to the optimal buffer size Q_{ij} for the given Λ-pair $\{i, j\}$. Denote by β_{ij} the running power of z_1, i.e.,

$$\begin{aligned}\beta_{ij} &= \frac{1}{T_{ij}}\left[\frac{Q_{ij}p_i}{r_i - r_y} + \frac{Q_{ij}p_j}{r_y - r_j}\right]\\ &= \frac{1}{\alpha_{ij}}\left[\frac{p_i}{r_i - r_y} + \frac{p_j}{r_y - r_j}\right].\end{aligned} \tag{11}$$

Note that both α_{ij} and β_{ij} are constants only depending on the given Λ-pair. The general pure trajectory with the Λ-pair $\{i, j\}$ and $2m$ switchings can be obtained from z_1 as $z_m = \mathcal{J}_m[\mathcal{S}_{1/m}[z_1]]$. As shown in Fig. 4-(b), z_m has an optimal buffer size Q_{ij}/m and its average power is

$$\bar{P}_{ij}(m) = \beta_{ij} + \frac{2mk_s}{t_f} + \frac{p_b t_f}{m\alpha_{ij}}. \tag{12}$$

Taking the derivative of $\bar{P}_{ij}(m)$ with respect to m and setting it to zero, we obtain the optimal value of m as

$$\hat{m}_{ij} = t_f\sqrt{\frac{p_b}{2k_s\alpha_{ij}}}. \tag{13}$$

Note that the parameter m must be an integer, and the function $\bar{P}_{ij}(m)$ is convex in m. Therefore, if $\hat{m}_{ij}$ is not an integer, the optimal feasible value of m_{ij}, m_{ij}^*, is whichever of the two closest integers to $\hat{m}_{ij}$ that results in the smallest value of $\bar{P}_{ij}(m)$ as defined in (12). Hence,

$$m_{ij}^* = \underset{m \in \{\lfloor \hat{m}_{ij} \rfloor, \lceil \hat{m}_{ij} \rceil\}}{\arg\min}\ \bar{P}_{ij}(m). \tag{14}$$

The minimal achievable power for any pure trajectory with the Λ-pair $\{i, j\}$ is $\bar{P}_{ij}(m_{ij}^*)$. Then the best Λ-pair $\{\sigma^+, \sigma^-\}$ can be obtained as

$$\{\sigma^+, \sigma^-\} = \underset{\{i \in I, j \in J\}}{\arg\min}\ \bar{P}_{ij}(m_{ij}^*). \tag{15}$$

Denote by Σ_f the set of all minimizers of (15). Thus if $\{i, j\} \in \Sigma_f$, then an OPT can be obtained by repeating the Λ-trajectory with the pair $\{i, j\}$ for m_{ij}^* times. Thus the following theorem follows immediately.

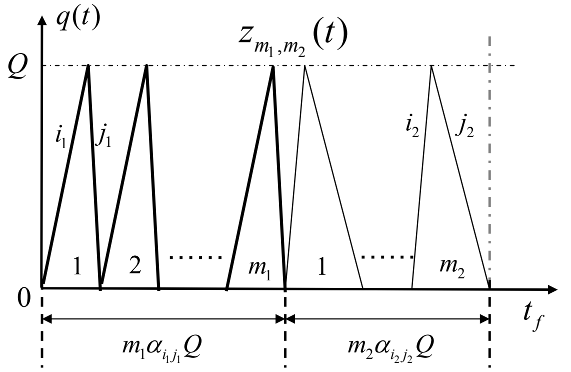


Fig. 5. An example of $z_{m_1,m_2}(t)$

Theorem 1: Let $z_m(t)$ be a pure trajectory as shown in Fig. 4-(b) with $2m$ switchings and Λ-pair $\{i,j\}$. If $\{i,j\} \in \Sigma_f$ and $m = m^*_{i,j}$, then z_m is an OPT with optimal buffer size Q_{ij}/m^*_{ij}.

D. General Optimal Solution

In last subsection, we derive analytically the optimal pure trajectories with $n_p = 1$. A natural question is that whether the power can be further reduced if we relax the constraint on n_p. To answer this question, we start with a simple case where candidate trajectories are allowed to contain at most two distinct Λ-pairs,[4] i.e., $n_p \le 2$. Let z_{m_1,m_2} be a BST consisting of m_1 copies of Λ-pair $\{i_1,j_1\}$ and m_2 copies of Λ-pair $\{i_2,j_2\}$. Without loss of generality, assume that all the same pairs are grouped together as shown in Fig 5. In other words, the switching sequence of z_{m_1,m_2} is assumed to take the following form

$$(\sigma_1,\ldots,\sigma_{2(m_1+m_2)}) = (\underbrace{i_1,j_1,\ldots,i_1,j_1}_{m_1 \text{ pairs}}\underbrace{i_2,j_2,\ldots,i_2,j_2}_{m_2 \text{ pairs}}).$$

For a given pair of m_1 and m_2, the optimal buffer size of z_{m_1,m_2} is uniquely determined by

$$Q = \frac{t_f}{\alpha_{i_1,j_1}m_1 + \alpha_{i_2,j_2}m_2},$$

where $\alpha_{i,j}$ is the ratio of the time duration to the optimal buffer size for the Λ-pair $\{i,j\}$ as defined in (10). Let $\beta_{i,j}$ be the running power of the pair $\{i,j\}$ as defined in (11). Then the total energy consumed by z_{m_1,m_2} is computed as

$$E(m_1,m_2) = 2(m_1+m_2)k_s + \frac{p_b t_f^2 + \beta_{i_1,j_1}\alpha_{i_1,j_1}m_1 t_f + \beta_{i_2,j_2}\alpha_{i_2,j_2}m_2 t_f}{\alpha_{i_1,j_1}m_1 + \alpha_{i_2,j_2}m_2}.$$

Lemma 3: For any $\{i_1,j_1\}$ and $\{i_2,j_2\}$, there exists a pair of nonnegative integers (m_1^*,m_2^*) with either $m_1^* = 0$ or $m_2^* = 0$ such that $E(m_1^*,m_2^*) \le E(m_1,m_2)$ for any other pair of nonnegative integers (m_1,m_2).

[4]Two different Λ-pairs may consist of three or four different modes. For example, $\{\sigma_1,\sigma_2\}$ and $\{\sigma_1,\sigma_3\}$ are also called two different Λ-pairs although they have one mode in common.

Proof: To simplify the notations, let $c = p_b t_f^2$, $a_1 = \beta_{i_1,j_1}\alpha_{i_1,j_1}t_f$, and $a_2 = \beta_{i_2,j_2}\alpha_{i_2,j_2}t_f$. Relax m_1, m_2 to nonnegative real numbers x_1 and x_2. Then

$$E(x_1,x_2) = 2k_s(x_1+x_2) + \frac{a_1x_1 + a_2x_2 + c}{\alpha_{i_1,j_1}x_1 + \alpha_{i_2,j_2}x_2}.$$

Note that all the constants a_1, a_2, c, α_{i_1,j_1}, and α_{i_2,j_2} are positive. To prove the lemma, it suffices to show that there exists a point on the x_1 or x_2 axis that minimizes $E(x_1,x_2)$ in the first quadrant. To find the minimizers of $E(x_1,x_2)$ in the first quadrant, we can first minimize it along each ray in the first quadrant, and then find the ray that gives the best minimum value. Towards this purpose, define $x_2 = \lambda x_1$, where $\lambda \in [0,\infty]$. Then

$$\begin{aligned}
E(x_1,\lambda x_1) =& 2k_s(1+\lambda)x_1 + \frac{(a_1+a_2\lambda)x_1 + c}{(\alpha_{i_1,j_1}+\alpha_{i_2,j_2}\lambda)x_1} \\
\ge& 2\sqrt{\frac{2k_s(1+\lambda)c}{\alpha_{i_1,j_1}+\alpha_{i_2,j_2}\lambda}} + \frac{(a_1+a_2\lambda)}{(\alpha_{i_1,j_1}+\alpha_{i_2,j_2}\lambda)} \\
\triangleq& E(x_1^*,\lambda x_1^*).
\end{aligned}$$

Thus $E(x_1^*,\lambda x_1^*)$ is the minimum value achieved on the ray $x_2 = \lambda x_1$. To prove the lemma, it suffices to show that either $\lambda = 0$ or $\lambda = \infty$ minimizes $E(x_1^*,\lambda x_1^*)$. After some computations, $E(x_1^*,\lambda x_1^*)$ reduces to

$$E(x_1^*,\lambda x_1^*) = d_3\sqrt{d_2 y + \frac{1}{\alpha_{i_2,j_2}}} + d_1 y + \frac{a_2}{\alpha_{i_2,j_2}} \triangleq f(y),$$

where $d_1 = a_1\alpha_{i_2,j_2} - a_2\alpha_{i_1,j_1}$, $d_2 = \alpha_{i_2,j_2} - \alpha_{i_1,j_1}$ and $d_3 = 2\sqrt{2k_s c}$ are all constants and

$$y = \frac{1}{\alpha_{i_2,j_2}^2\lambda + \alpha_{i_1,j_1}\alpha_{i_2,j_2}}.$$

Note that except d_1 and d_2, all the other constants are positive. As λ increases from 0 to ∞, y decreases from $\frac{1}{\alpha_{i_1,j_1}\alpha_{i_2,j_2}}$ to 0. Hence, it suffices to show that either 0 or $\frac{1}{\alpha_{i_1,j_1}\alpha_{i_2,j_2}}$ is a minimizer of $f(y)$ in $[0, \frac{1}{\alpha_{i_1,j_1}\alpha_{i_2,j_2}}]$. Note that the second-order derivative of $f(y)$ is

$$\frac{d^2f}{dy^2}(y) = -\frac{d_2^2 d_3}{4(d_2y + 1/\alpha_{i_2,j_2})^{3/2}} \le 0.$$

Thus $f(y)$ is a concave function of y in $[0, \frac{1}{\alpha_{i_1,j_1}\alpha_{i_2,j_2}}]$. Since the minimizer of a concave function over a bounded set must be on the boundary of the set, we conclude that either 0 or $\frac{1}{\alpha_{i_1,j_1}\alpha_{i_2,j_2}}$ is a minimizer of $f(y)$ in $[0, \frac{1}{\alpha_{i_1,j_1}\alpha_{i_2,j_2}}]$. ∎

According to Lemma 3, for any given two Λ-pairs, we can always use one of them to construct a pure trajectory that performs equally well or better than all the other mixed trajectories involving these two Λ-pairs. Therefore, the following corollary follows immediately.

Corollary 1: The OPT is an optimal solution to Problem 1 under an additional constraint $n_p \le 2$.

The question now becomes that whether we can save more energy by further relaxing the constraint on n_p. It turns out to be not the case. In fact, the OPT is an optimal solution to Problem 1 for an arbitrary n_p. This can be proved by induction. The following lemma is the key of the induction procedure.

Lemma 4: For any BST z with length t_f and $n_p = l+1$, there exists another BST $\hat{z}$ with length t_f and $n_p \leq l$ that consumes equal or less power than z.

Remark 1: Note that Lemma 4 is not a trivial application of Lemma 3. Its proof is much longer and is thus omitted. Interested reader can refer to [12] for a complete proof.

According to Lemma 4, any BST corresponds to a pure trajectory with $n_p = 1$, thus the following theorem follows immediately.

Theorem 2: The OPT defined in Theorem 1 is an OS1 for an arbitrary n_p.

IV. SIMULATION

Our theoretical results can be applied in many real-world applications, such as the power management problem of a multiple-speed disk [6] and the dynamic voltage scheduling (DVS) problem of a variable speed processor [5]. In this section, we use a DVS example to illustrate the effectiveness of our results.

TABLE I

POWER MODES OF INTEL XSCALE PROCESSOR

mode i	1	2	3	4	5
f_i (MHz)	150	400	600	800	1000
r_i (MB/s)	0.45	1.2	1.8	2.4	3
p_i (Watt)	0.08	0.17	0.4	0.9	1.6

Let X be an Intel Xscale processor [5] with five available power modes as defined in Table I. Suppose that Y is a video card that fetches data from X at a constant speed 8 Mbps (1MB/s). The power per megabyte for buffer B is 6.258×10^{-4} W/MB [4]. A typical value of the switching energy is 0.1mJ in a microprocessor [13]. Since the switching cost k_s in our model may also include other switching penalties, such as the switching delay penalty, we will test our method for k_s ranging from 0.1mJ to 100mJ. A heuristic strategy is implemented where X is switched to the highest speed until the buffer is full and then switched to the lowest speed until the buffer is empty. This method is referred to as Scheme2 while Scheme1 refers to the optimal strategy as defined in Theorem 1. Scheme2 is tested for four heuristically selected buffer sizes 0.1MB, 0.3MB, 1MB and 8MB. The power consumptions of Scheme2 in these cases are compared with Scheme1 in Fig. 6. It is clear that our optimal strategy always performs the best for each k_s and can save more than 60% of power consumptions compared with the other strategy.

V. CONCLUSIONS

A dynamic power manage problem is formulated as an optimal control problem of a hybrid system. By exploiting of the particular structure of our system model, the optimal solutions are derived analytically based on some variational approach. Simulation result based on real data shows that the proposed method can save 60% energies compared with a heuristic scheme. Future research will focus on the following two aspects: one is to extend our analysis to the case with more than one buffers inserted among multiple streamlined components. The other one is to study the case where the data rates of components are varying or even random instead of constant.

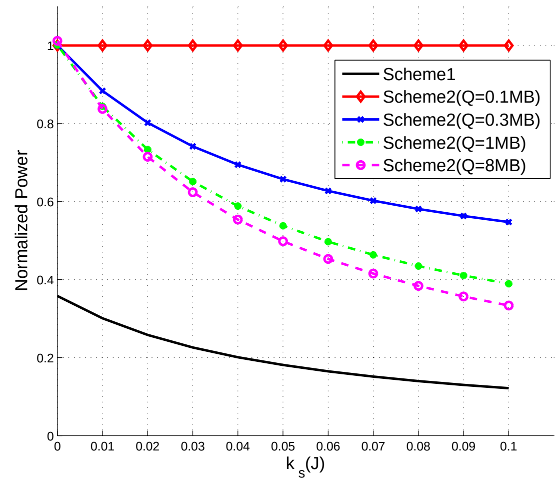


Fig. 6. Power consumptions of various methods under different k_s. For each k_s, all the powers are normalized w.r.t. the largest one.

REFERENCES

[1] W. Zhang, J. Hu, and Y. H. Lu. Optimal power modes scheduling using hybrid systems. In *Proceedings of the American Control Conference*, New York City, NY, July 2007.

[2] L. Cai and Y.-H. Lu. Energy management using buffer memory for streaming data. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 24(2):141–152, 2005.

[3] J. Hu and Y.-H. Lu. Buffer management for power reduction using hybrid control. In *Proceedings of the IEEE Conference on Decision and Control*, pages 6997–7002, Seville, Spain, 2005.

[4] J. Ridenour, J. Hu, N. Pettis, and Y.-H. Lu. Low-power buffer management for streaming data. *IEEE Transactions on Circuits and Systems for Video Technology*, 17(2):143–157, 2007.

[5] R. Xu, C. Xi, R. Melhem, and D. Moss. Practical pace for embedded systems. In *EMSOFT '04: Proceedings of the 4th ACM international conference on Embedded software*, pages 54–63. ACM Press, 2004.

[6] S. Gurumurthi, A. Sivasubramaniam, M. Kandemir, and H. Franke. Drpm: Dynamic speed control for power management in server class disks. In *Proceedings of the International Symposium on Computer Architecture (ISCA)*, pages 169–179, June 2003.

[7] R. Alur, T. Henzinger, G. Lafferriere, and G. J. Pappas. Discrete abstractions of hybrid systems. *Proceedings of the IEEE*, 88(2):971–984, 2000.

[8] S. Hedlund and A. Rantzer. Optimal control of hybrid system. In *Proceedings of the IEEE Conference on Decision and Control*, volume 4, pages 3972–3977, Phoenix, AZ, December 1999.

[9] M. Egerstedt, Y. Wardi, and F. Delmotte. Optimal control of switching times in switched dynamical systems. *IEEE Transactions on Automatic Control*, 51(1):110–115, 2006.

[10] X. Xu and P.J. Antsaklis. Optimal control of switched systems based on parameterization of the switching instants. *IEEE Transactions on Automatic Control*, 49(1):2–16, 2004.

[11] S. C. Bengea and R. A. DeCarlo. Optimal control of switching systems. *Automatica*, 41(1):11–27, 2005.

[12] W. Zhang and J. Hu. Optimal buffer management using hybrid systems. *Automatica*, 2007. submitted.

[13] T. D. Burd and R. W. Brodersen. Design issues for dynamic voltage scaling. In *Proceedings of the international symposium on Low power electronics and design*, pages 9–14. ACM Press, 2000.