OPTIMALLY CONTROLLED CCD SHIFT REGISTERS
(OPTIMAL INTERCEPTION ON A RECURRENT TRAJECTORY)
(Revised Version)

by

Micha Hofri, Zvi Rosberg

Technical Report #381
November 1985
OPTIMALLY CONTROLLED CCD SHIFT REGISTERS
(OPTIMAL INTERCEPTION ON A RECURRENT TRAJECTORY)

Micha Hofri, Zvi Rosberg
Department of Computer Science,
Technion, Haifa Israel.

ABSTRACT

The question of optimally controlling the shift rate of a CCD-type shift register, when used as a secondary memory device has been treated a few times in the literature. The usual procedure has been to postulate the functional form of the optimal control action, and then evaluate the optimizing parameter under some assumption about the request arrival process. We present a control-theoretic treatment of the problem, derive the form of the optimal policy (which indeed turns out to be nearly identical to the one postulated before) and compute the optimizing parameter for a variety of cases.

Given the optimal policy we compute the distributions of the queue length and the device response time.

Key Words: Optimal control, Bang-bang, CCD shift register

1. INTRODUCTION

Consider a shift register containing $L$ bits which can only be read or written through a port $P$ - c.f. Figure 1. The register is realized in a technology like that of Charge-Coupled Devices which implies that the contents of the register must rotate continuously. The speed of rotation need not be constant but is bounded below by the charge 'refresh' requirements and from above by the switching rate of the devices or driving clocks. Let these bounds be $a$ and $b$ bits per second, respectively.

We view the register as connected (possibly with many replicas of itself) to a computer that generates requests to read or rewrite the contents of the register. These operations are assumed to require the entire contents of the register to be conducted opposite the port $P$, and to commence only when a certain position (signature) reaches the port. We shall denote this signature by bit 0.

Consider the operation of the device controller. Apparently, whenever there is a pending I/O request, the register contents will be
shifted at the maximum speed \( b \) till the request is satisfied. If another request is pending at that instant it can immediately be processed with no latency. Using standard queueing-theoretical terminology we could say that only a request that starts a busy-period would experience rotational latency, whereas others are serviced as soon as their turn arrives. The non-obvious optimality of this rule is demonstrated in the Appendix. Thus, from the instant of arrival of a request finding an 'idle' register (i.e. a request starting a busy-period) till the termination of that period, the controller will maximize the benefit to the users, which we interpret as minimizing the delays of their requests, by maintaining a maximal rotational speed.

In this paper we address the problem of determining the optimal operation of the controller which will minimize the latency of the requests that start busy-periods. In the sequel we shall assume the arrival process of new requests to be a stationary Poisson process, and then the idle periods are independent and identically distributed. Thus, the optimal operation during a single idle period in conjunction with the trivial optimal operation during the busy-period will determine the global optimal policy. At first glance, it seems that continuously maintaining the maximal speed \( b \) is optimal. It will turn out that this is not the case. The reason is not hard to find - envision the position of bit 0 as it circulates. If a request arrives when bit 0 is in position \( s \) (the distance from the port measured in clockwise sense - see Figure 1) the latency will be at least \( s/b \). Assuming instantaneous speed changes, we take this as its actual
value. Hence, we wish to minimize the position of bit 0 at interception time, figuratively speaking. Therefore, a more promising policy seems to be the policy which applies as high a speed as possible - \( b \) - when bit 0 is still far away from the port and a lower rate, possibly even the lowest - \( a \) - when bit 0 is close to the port.

It turns out that this immediate observation covers most of the ground, with some allowance for the nature of the control policies under consideration.

In the next sections we shall formulate the problem, quantify and prove the optimality of the above approach, even in stronger form, using nearly exclusively extreme versions of that policy, where only the rates \( a \) and \( b \) are used. This type of policy is commonly called a 'bang-bang' policy.

The problem - restricted to the minimization of latency - has been treated before in the literature. Sites [4], and Fuller and McGehearty [1] consider the problem under the assumption that the instant of arrival is uniformly distributed over the duration that the shift register completes a revolution. This however is incompatible with the assumption that the arrival process is a renewal process, though in fact it is the limit of a Poisson arrival process when the rate of arrival approaches zero! In Gelenbe and Mitrani [2, Chap. 2] the problem is considered under the assumption of the same arrival process as we do, but their treatment seems to have been derailed by a computational error. All these treatments assume that a 'bang-bang' policy with a single switch-over point is the best policy to adopt. Then one has only to find the optimal change-over point, which the first two references identify correctly, under their assumptions.

In contradistinction we place the problem in the context of control theory, and derive the explicit form of the optimal control. We show that when the rotational rate can be changed at any instant the optimal control is indeed a 'bang-bang' one, with the instant of change-over depending on the rate of arrival. When the rate of rotation can only be changed at a discrete set of points this policy is not necessarily optimal, though an intermediate rate need only be used once per revolution (at most) between two adjacent decision points.

In Section 2 we formulate the problem. In Section 3 we prove that the 'bang-bang' policy is \( \varepsilon \)-optimal for the discrete-time control problem and in Section 4 we show that it is optimal for the continuous-time control problem. The explicit form of the optimal policy is also given.
CCD shift registers ...

Given the optimal control policy the queueing analysis for the distributions of the number of enqueued I/O requests and their overall delay is straightforward. It is a variant of the so-called "a single-server queueing system with walking time". The analysis is done in Section 5. A discussion in Section 6 concludes the paper.

2. A CONTROLLED MARKOV PROCESS MODEL

We shall formulate a model for the shift register as a controlled Markov process in discrete and continuous time.

Assume the geometry of the register as given in Figure 1. The register contents rotate with speed \( u \) which can be controlled by the device controller. The speed is constrained to be in the interval \([a, b]\). We assume that the rotational speed can be changed instantaneously. No distinction is made between read and write requests, which arrive according to a Poisson counting process with rate \( \lambda \). Hence, at every instant during an idle period, the time until the next request arrival that will initiate a busy-period is exponentially distributed with parameter \( \lambda \), and in particular, it is independent of the position of bit 0.

Theorem A, in the Appendix, lets us concentrate entirely on the idle periods to determine the optimal policy. Thus, the position of bit 0, measured by the distance from the port in clockwise sense, and denoted by \( s \), may be taken as the state of the system.

Let \( L \) be the circular length of the register, then \( s \) varies in the interval \([0, L]\). We shall add to the state space a fictitious absorbing state \( * \), and allow the system to cease operations by entering state \( * \). Once the system enters this state it "terminates".

At every decision epoch, when the system is at state \( s \neq * \), the controller has to determine a rotational speed \( u, a \leq u \leq b \). The objective of the controller is to minimize the expected latency during an idle period; this is the same as minimizing the expected value of \( s \) from which the system will enter the state \( * \). Accordingly, we shall define the immediate cost \( C(s, u) \), when the system is in state \( s \neq * \) and takes action \( u \) as follows:

If there is an arrival before the next decision epoch then \( C(s, u) \), \( s \neq * \), is the state of the system at the moment of that first arrival, and zero otherwise. Obviously, \( C(*, u) = 0 \).
CCD shift registers ...

As long as there is no arrival (the system is still in an idle period) the law of motion (transition probabilities) from any state $s$ is determined by the rotational speed $u$ and the standard motion equation - distance covered equals duration times rate. At the moment of the first arrival the system enters the state $*$ and stays there for ever. A control policy is called **admissible** if it assigns to each state $s$ in the set of controllable states a rate of rotation in the physically realizable range $[a, b]$. For every admissible policy $\pi$, let $V(\pi, s)$ be the expected cost of operating the system given that it is currently at state $s$ and utilizes policy $\pi$.

Let $V(s) = \inf \pi V(\pi, s)$.

$V$ is the **value function**. A control policy $\pi^*$ is optimal if

$$V(\pi^*, s) = V(s) \quad \text{for every controllable } s.$$  

The problem can be treated as a discrete-time problem as well as a continuous-time one:

In the discrete problem we permit the controller to make decisions only at a set of discrete moments of time. These will then determine the set of controllable states. Once a rotational speed is selected at a certain state, the same speed is maintained till the next decision point or till the system enters $*$, whichever comes first. A natural set is the one marked out by bit positions, yielding $L$ decision instants. We shall consider a more comprehensive family of sets though.

In the continuous problem the controller may vary the rotational speed at every moment of time, and then the set of controllable states is the interval $(0, L]$.

**Remark 2.1** Operationally we are only interested in minimizing $V(L)$, since a busy-period is always terminated with $s = L$. Still we have to embed this in the problem of optimizing the value function for all $s$ in the set of controllable states.

3. **THE DISCRETE-TIME PROBLEM**

Let $N$ be any positive integer; $\Delta = L2^{-N}$ and divide the state space interval $(0, L]$ to $2^N$ equally spaced subintervals, each of size $\Delta$. This particular choice will be useful in showing the $\varepsilon$-optimality of the chosen policy.
The controller is permitted to change the rotational speed only when the system is at one of the states \( s = k \Delta, \ 0 < k \leq 2^N \). During the intervals \((k-1)\Delta, k\Delta]\) the speed is maintained constant.

Every value of \( \Delta \) defines a discrete-time control problem which will be referred as the \( \Delta \)-problem. The immediate cost, the expected cost under policy \( \pi \) and the value function will be denoted by \( C(s,u), V(\pi, s) \) and \( V_\Delta(s) \) respectively.

Let \( \overline{C}_{\Delta}(s,u) \) be the expectation of \( C(s,u) \) in the \( \Delta \)-problem.

From the definition of \( \overline{C}_{\Delta}(s,u) \) in Section 2 and the nature of the Poisson arrival process,

\[
\overline{C}_{\Delta}(s,u) = \int_0^{\Delta} \lambda e^{-\lambda t} (s - tu) dt = \Delta e^{-\lambda \Delta u} + \left(1 - e^{-\lambda \Delta u}\right) \left(s - \frac{u}{\lambda}\right),
\]

for every \( s \neq * \).

Note that by our choice of coordinates \( s \) decreases with time over the range \((0,L]\). Now for every \( \Delta \) and state \( s \neq * \), the probability that the system will "terminate" (enter state *) is

\[
\gamma = 1 - e^{-\lambda \Delta/\beta} > 0.
\]

Thus, using terminology from [3, p. 56], we have a discounted dynamic programming problem.

Bellman's optimality equations for this problem are:

\[
V_\Delta(s + \Delta) = \inf_{u} \{ \overline{C}_{\Delta}(s + \Delta, u) + e^{-\lambda \Delta u} V_\Delta(s) \},
\]

for \( s = k \Delta, \ 0 < k \leq 2^N \) and \( s + \Delta \) taken modulo \( L \). Since \( u \) varies on a compact set, \( \overline{C}(s,u) \) is absolutely bounded and uniformly continuous in \( u \). The next theorem follows then from Theorems 6.2 and 6.3 in [3, p. 56].

**Theorem 3.1** For every \( \Delta \)-problem, there exists an optimal stationary control policy which satisfies the optimality equations (3.3). Moreover, if the actions taken by a policy \( \pi^* \) satisfy (3.3) then \( \pi^* \) is optimal.

In the rest of this section we shall use the simplicity of (3.3) to find the optimal policy and to prove that the 'bang-bang' policy is \( \epsilon \)-optimal for small \( \Delta \)'s. Lemma 3.2 provides us with a computational tool.

**Lemma 3.2** For every state \( s \) and control action \( u \)
CCD shift registers ...

\[ V_\Delta(s) - s > \Delta + e^{-\lambda \Delta} (V_\Delta(s+\Delta) - (s+\Delta)). \quad s \neq L \]

**Proof:** Substituting from (3.1) and (3.3)

\[ V_\Delta(s+\Delta) \leq s + \Delta + (V_\Delta(s) - s)e^{-\lambda \Delta} - \frac{u}{\lambda} \left( 1 - e^{-\lambda \Delta} \right) \]

and the lemma follows by algebraic manipulations. \( \square \)

**Note:** Since for \( s = L \) the value of \( s + \Delta \) should be taken as \( \Delta \), the proof does not hold there. This will not impede the following analysis.

From Lemma 3.2 we can deduce the following monotonicity properties of the difference \( V_\Delta(s) - s \).

**Lemma 3.3** For the states \( s = k\Delta, 0 < k \leq 2^N \):

(i) If \( V_\Delta(s) \leq s \) then \( V_\Delta(s+\Delta) < s + \Delta, \ s \neq L \).

(ii) If \( V_\Delta(s) \geq s \) then \( V_\Delta(s-\Delta) > s - \Delta, \ s \neq \Delta \).

**Proof:** part (i) is immediate a fortiori from Lemma 3.2. Part (ii) is the counterpart of (i). \( \square \)

**Corollary 3.4:** Lemma 3.3 implies that the curves\(^1\) \( y_1(s) = V_\Delta(s) \) and \( y_2(s) = s \) intersect at most once.

Clearly, from the interpretation of the cost function \( V_\Delta \) as the expected interception coordinate, \( V_\Delta(L) < L \) and \( V_\Delta(0^+) > 0 \), hence \( V_\Delta(s) \) and \( y_2(s) = s \) intersect exactly once. The relation between \( y_2(s) = s \) and \( V_\Delta(s) \) is given in Figure 2.

Let \( I = (s_0, s_0 + \Delta] \), where \( s \) is of the form \( k\Delta, \) be the interval containing the intersection, i.e.,

\begin{align*}
V_\Delta(s) > s, & \quad \text{for } s < s_0, \quad V_\Delta(s_0) = s_0, \quad (3.4) \\
V_\Delta(s) < s, & \quad \text{for } s = s_0 + \Delta
\end{align*}

Note that Lemma 3.2 implies the monotonicity of \( V_\Delta(s) \) in the interval \((0, s_0 - \Delta]\). However, the monotonicity of \( V_\Delta(s) \) is not used in the derivation of our results.

For every \( \Delta \)-problem define the following bang-bang policy \( \pi_{bb}(\Delta) \).

---

\(^1\) We extend \( V(k\Delta) \) \( 0 < k \leq 2^N \) to the entire interval \((0, L]\) in the natural way which conforms to the definition of \( V(s) \), by adding the expected cost from \( s \) to the next-decision epoch and the value of \( V(\cdot) \) there.
Definition 3.5 The policy \( \pi_{bb}(\Delta) \) is the stationary policy which takes the control action \( u(s) \), where

\[
\begin{align*}
  u(s) = \begin{cases} 
    b & \text{if } s \in \{ s \mid V^\Delta(s) < s \} \\
    a & \text{if } s \in \{ s \mid V^\Delta(s) \geq s \}
  \end{cases}
\end{align*}
\]

(3.5)

From (3.4) we observe that \( \pi_{bb}(\Delta) \) assigns

\[
\begin{align*}
  u(s) = \begin{cases} 
    b & \text{if } s > s_0 \\
    a & \text{if } s \leq s_0
  \end{cases}
\end{align*}
\]

(3.6)

We now give the form of the optimal policy \( \pi^*(\Delta) \) and show that \( \pi_{bb}(\Delta) \) takes the same action as \( \pi^*(\Delta) \) except possibly at \( s_0 + \Delta \).

Let

\[
  g_u(s) = \tilde{C}_\Delta(s,u) + e^{-\lambda\Delta u} V^\Delta(s-\Delta),
\]

(3.7)

where \( \tilde{C}_\Delta(s,u) \) is given in (3.1), and \( V^\Delta(s-\Delta) = V^\Delta(L) \), when \( s = \Delta \).

Lemma 3.6 For every \( u, a \leq u \leq b \),

(i) If \( s > s_0 + \Delta \) then \( g_u'(u) < 0 \),

(ii) If \( s \leq s_0 \) then \( g_u'(u) > 0 \), for \( \Delta \) small enough.
9.

Case 2: The number of passes is bounded.

If the number of distinct messages is finite then $A$ uses only $O(n)$ bits and by Theorem 3 is a regular language; otherwise, by Corollary 3 we conclude that $BIT_A(n) = \Omega(n \log n)$.

5. MISCELLANEOUS

We conclude with few notes about the bit complexity hierarchy, and the trade-off between the number of passes vs. the bit complexity for regular languages.

1. There is a linear language $L$, such that $BIT_L(n) = n^2$.

Let $L = \{ x \mid x = w \cdot w^r, w \in \{a,b\}^* \}$. It is easy to show that at least $\Omega(n^2)$ bits must be sent by every algorithm that recognizes $L$.

2. There is a context-sensitive language that is not context-free and that requires only $O(n \log n)$ bits.

Let $L = \{ 0^n \cdot 1^n \mid n = 1, 2, \ldots, n \}$. $L$ can be recognized in $O(n \log n)$ bits (by sending three counters).

3. For every function $g(n)$, $\Omega(n \log n) < g(n) < O(n^2)$ there exists a language $L$ such that $BIT_L(n) = \Theta(g(n))$.

Let $\Sigma$ be any alphabet, and $L = \{ w \mid \exists x, y \in \Sigma^*, i > 0, \text{ such that } w = x^i y, |x| > |y| \}$ and $\lfloor g(|w|) \rfloor = |x|$. For every $i > 0$, $c^i \in L$, so for every $n$ there is at least one word $w \in L$, such that $n = |w|$.

Every algorithm that recognizes $L$ requires at least $\Omega(|x| \cdot |w|)$ bits. Even if we assume that $n$ is known to every processor, and even more, every processor knows which bit of $x$ it holds, still at least $n - |x| - |y|$ of the processors will send $|x|$ bits for $w \in L$ (otherwise, there will be two distinct words of size $|x|$ for which one of these processors will send the same sequence of bits, and clearly one of these subwords yields a word not in $L$ that will be accepted, a contradiction). Therefore we have the lower bound of $\Omega(g(n))$ bits. The following algorithm recognizes $L$ in $O(g(n))$ bits: The leader determines $n$ (using $O(n \log n)$ bits), and then
CCD shift registers ...

below), and the optimal \( s_0 \) and \( u_0 \) can be directly found. Computational results have indeed shown cases where the optimal \( u_0 \) is not necessarily \( a \) or \( b \). We also could show that when \( \lambda \) is large enough (so that an idle period rarely extends beyond the initial portions of the trajectory of bit 0, after it leaves \( s=L \), \( s_0 \rightarrow 0 \), so that action \( a \) is not used at all. This could be shown, from \((3.3)\) to happen also if \( A \) is not larger than a certain threshold.

**Corollary 3.9** For small enough \( \Delta \)'s, the bang-bang policy \( \pi_{bb}(\Delta) \) takes the same control actions as the optimal policy except possibly at the single node \( s_0 + \Delta \).

We now proceed to show that \( \pi_{bb}(\Delta) \) is \( \varepsilon \)-optimal; or in other words, that

\[
V_{\Delta}(\pi_{bb}(\Delta),s) - V_{\Delta}(s) \rightarrow 0.
\]

This is done implicitly, by computing expressions for the value function under policies that have the form of \( \pi_{bb}(\Delta) \) and \( \pi^*(\Delta) \), with arbitrary values for \( s_0 \) and \( u_0 \) (i.e. without actually requiring that these values satisfy the optimality equations).

We shall also use the fact that by increasing \( N \) (and decreasing \( \Delta \)) we add decision points, since \( \Delta \) is of the form \( 2^{-N} \). Therefore, \( V_{\Delta}(s) \) is non-increasing in \( \Delta \). Thus, the following limits exist:

\[
\lim_{\Delta \to 0} V_{\Delta}(s) = V(s)
\]

and

\[
\lim_{\Delta \to 0} s_0 = \lim_{\Delta \to 0} s_0(\Delta) = s^*.
\]

Clearly, \( s^* \) is the intersection between the curves of \( y(s) = s \) and \( V(s) \).

Using the notation \( t(s) \) and \( s(t) \) to denote the time bit 0 reaches position \( s \), and the position it reaches at time \( t \) respectively, we have for any policy \( \pi \) (that depends on \( \Delta \)):

\[
(1) \quad V_{\Delta}(\pi(\Delta),x) = \int_{t=0}^{t(0^+)} \lambda s e^{-\lambda t}s(t)dt + \lambda s e^{-\lambda t}V_{\Delta}(\pi(\Delta),L), \quad s(0) = x
\]

and in particular,

\[
(2) \quad V_{\Delta}(\pi(\Delta),L) = \int_{t=0}^{T(\pi(\Delta))} \lambda s e^{-\lambda t}s(t)dt + \lambda e^{-\lambda T(\pi(\Delta))}V_{\Delta}(\pi(\Delta),L), \quad s(0) = L
\]

where \( T(\pi) \) is the duration of an uninterrupted full revolution of the register under policy \( \pi \).
We now compute $V_\Delta(\pi_{bb}(\Delta), s)$, using (3.8) and (3.12) to find

$$V_\Delta(\pi_{bb}(\Delta), s) = \begin{cases} 
  s - \frac{\alpha}{\lambda}(1 - e^{-\lambda s/\alpha}) + e^{-\lambda s/\alpha} V_\Delta(\pi_{bb}(\Delta), L) & 0 < s \leq s_0 \\
  s - s_0 e^{-\lambda(s-s_0)/\lambda} + \frac{b}{1 - e^{-\lambda}} V_\Delta(\pi_{bb}(\Delta), s_0) & s_0 < s \leq L
  \end{cases}$$

(3.13a)

and denoting $e^{-\lambda(L-s_0)/b}$ by $\alpha$ and $e^{-\lambda s_0/\alpha}$ by $\beta$ we further find

$$V_\Delta(\pi_{bb}(\Delta), L) = \frac{L - b (1 - \alpha) - \frac{a}{\lambda} \alpha (1 - \beta)}{1 - \alpha \beta}$$

(3.13b)

$$V_\Delta(\pi_{bb}(\Delta), s_0) = s_0 + \frac{L \beta - \frac{a}{\lambda} (1 - \beta) - \frac{b}{\lambda} \beta (1 - \alpha)}{1 - \alpha \beta}.$$

The same computation for $V_\Delta(\pi^*(\Delta), s)$, using also (3.10) is somewhat more cumbersome, yielding

$$V_\Delta(\pi^*(\Delta), s) = \begin{cases} 
  s - \frac{\alpha}{\lambda}(1 - e^{-\lambda s/\alpha}) + e^{-\lambda s/\alpha} V_\Delta(\pi^*(\Delta), L) & 0 < s \leq s^0 \\
  s - s_0 e^{-\lambda(s-s_0)/\alpha} - \frac{u_0}{\lambda} (1 - e^{-\lambda(s-s_0)/\alpha}) + e^{-\lambda(s-s_0)/\alpha} V_\Delta(\pi^*(\Delta), s_0) & s_0 < s \leq s_0 + \Delta \\
  s - (s_0 + \Delta) e^{-\lambda(s-s_0-\Delta)/b} - \frac{b}{\lambda} (1 - e^{-\lambda(s-s_0-\Delta)/b}) + e^{-\lambda(s-s_0-\Delta)/b} V_\Delta(\pi^*(\Delta), s_0 + \Delta) & s_0 + \Delta < s \leq L
  \end{cases}$$

(3.14a)

Defining again $\eta = e^{-\lambda s_0/\alpha}$, $\Psi = e^{-\lambda u_0}$, $\zeta = e^{-\lambda(L-s_0-\Delta)/b}$, one finds
CCD shift registers ...

\[ V_\Delta(\pi^*(\Delta), s_0 + \Delta) = s_0 + \Delta + \frac{L \eta \phi - \frac{u_0}{\lambda} (1 - \phi) - \frac{a}{\lambda} \phi (1 - \eta) - \frac{b}{\lambda} \eta \phi (1 - \xi)}{1 - \eta \phi} \] (3.14b)

Comparing (3.13) and (3.14), noting that as \( \Delta \to 0 \), \( \phi \) approaches 1, we obtain

**Theorem 3.10** \( V_\Delta(\pi_{\text{bb}}(\Delta), s) - V_\Delta(s) \to 0 \), uniformly in \( s \).

**Proof** The limit follows from the continuity of the terms that take part in \( V_\Delta(\pi_{\text{bb}}(\Delta), s) \) and \( V_\Delta(\pi^*(\Delta), s) \) as can be seen from (3.13-14). The uniformity follows from the terms being also bounded, and the limit holds for every \( s \) in the interval \( (0, L) \).

We have thus reached our goal. True, the statements of the policies \( \pi_{\text{bb}}(\Delta) \) (or \( \pi^*(\Delta) \)) still contain a parameter (or two) we have not yet determined. When one solves (3.3) numerically they are naturally obtained as well as the functional values of \( V(s) \), as given generically in Figure 2. The parameters could in principle be also extracted from equation (3.14) (or (3.13)); this will be however far simpler to do in the context of the continuous-time version of the problem we investigate in the next Section. Theorem 3.10 assures us of the quality of the values thus derived even for discrete-time control.

**4. THE CONTINUOUS-TIME PROBLEM**

In this section we permit a continuous-time control of the rotational speed, i.e., the controller may change the speed at every state \( s \in (0, L) \).

Let \( \pi \) be a continuous-time control and \( V(\pi, s) \) its expected cost given that the system is currently at state \( s \).

Let \( V(s) = \inf_\pi V(\pi, s) \).

Since the control variable \( u \) is bounded it follows that \( V(\pi, s) \) is continuous in \( s \).
Let $V(\pi, s)$ be the left derivative of $V(\pi, s)$. Dividing (4.2) by $\Delta$ and letting $\Delta \to 0$

For every $\pi$, $\Delta > 0$ and state $s > \Delta$ let $\tau(s, \pi)$ be the travel time from state $s$ to state $s - \Delta$ under policy $\pi$. Then, the expected policy cost satisfies

$$V(\pi, s) = \int_0^{\Delta} \bar{\sigma}(s - x, u(s - x))dx + e^{-\lambda\tau(s, \pi)}V(\pi, s - \Delta).$$

(4.2)

Let $\bar{\sigma} V(\pi, s)$ be the left derivative of $V(\pi, s)$. Dividing (4.2) by $\Delta$ and letting $\Delta \to 0$

$$V(\pi, s) = s - \frac{u(s)}{\lambda} \bar{\sigma} V(\pi, s).$$

(4.3)

Similarly, we obtain the optimality equation

$$V(s) = \inf_{a \in \mathcal{A}} \left\{ s - \frac{u(s)}{\lambda} \bar{\sigma} V\right\}. $$

(4.4)

The optimal control, if it exists, satisfies the optimality equation (4.4), whence the optimal stationary action function $u(s)$ is defined.

We show in the next theorem that the 'bang-bang' control policy is the unique solution to (4.4), up to the control action at a single point.

A direct consequence of (4.4) is

**Lemma 4.1** For every state $s$, $0 < s \leq L$

- $V(s) > s$ implies $\bar{\sigma} V(s) < 0$,
- $V(s) < s$ implies $\bar{\sigma} V(s) > 0$,
- $V(s) = s$ implies $\bar{\sigma} V(s) = 0$.

Let $s^*$ be a state satisfying $V(s^*) = s^*$. Its existence follows from the fact that $V(L) < L$, $V(0^+) > 0$ and the continuity of $V(s)$. The last statement of Lemma 4.1 provides its uniqueness.

**Definition 4.2** Let $\pi_{bb}$ be the stationary bang-bang policy which takes the control actions $u^*(s)$, where
Theorem 4.3 The bang-bang policy \( \pi_{bb} \) is optimal for the continuous time problem.

Proof: Since the control variable \( u \in [a, b] \) it is easy to show that the set of all control policies \( \pi \) is a compact metric space. Moreover, \( V_\pi(s) \) is continuous in \( \pi \). Thus, the \( \inf \limits_v V_\pi(s) \) is obtained and there exists an optimal policy.

From Lemma 4.1 and (4.4) it follows that \( V(s) > s \) for \( s < s^* \) and \( V(s) < s \) for \( s > s^* \). Thus, from (4.4) \( \pi_{bb} \) is the only solution (up to the decision at state \( s^* \)) to the optimality equations.

The existence and the fact that the optimal policy satisfies the optimality equation complete the proof. \( \square \)

Note that \( \pi_{bb} \) is optimal among all possible policies, and we did not restrict ourselves to policies under which the rotation rate and \( V_\pi(s) \) are smooth.

It is straightforward now to parametrize explicitly the optimal policy. Under \( \pi_{bb} \), \( V_{\pi_{bb}}(s) \) is differentiable at any \( s \), except 0 and \( s^* \). Thus, solving (4.3) as a first order differential equation we have for some constants \( C \) and \( D \)

\[
V(s) = \begin{cases} 
C e^{-\lambda s/a} + s - \frac{a}{\lambda} & 0 < s \leq s^* \\
D e^{-\lambda s/b} + s - \frac{b}{\lambda} & s^* < s \leq L 
\end{cases} \tag{4.5}
\]

The requirement that both branches agree at \( s = s^* \) and the relation \( V(0) = V(L) \) determine \( C \) and \( D \). The latter follows from the continuity of \( V(s) \), and this also implies that \( C \) is a linear function of \( D \) with a positive slope \( e^{-\lambda L/b} \); thus they are minimized together. The value \( s^* \) is the one that minimizes \( V(s) \). From (4.5) we find

\[
D = \frac{\frac{b-a}{\lambda} (e^{\lambda s^*/a} - 1) + L}{e^{-\lambda s^*(1/a - 1/b)} - e^{-\lambda L/b}} \tag{4.6}
\]

Differentiating \( D \) with respect to \( s^* \) and equating the derivative to zero yields
CCD shift registers ...

\[ b \left( e^{-\lambda L s^* \sqrt{b}} - e^{\lambda s^* / a} \right) + (b - a) \left( e^{\lambda s^* / a} - 1 \right) + \lambda L = 0. \]  
\[ (4.7) \]

Consider first an approximate solution, valid for low arrival rates. By dropping from (4.7) terms of order \( O \left( \frac{\lambda}{a L} \right)^3 \) we obtain

\[ \left( \frac{L - s^*}{s^*} \right)^2 = \frac{b}{a}. \]
\[ (4.8) \]

This is precisely the value found in [1] upon assuming only bang-bang policies and uniform arrival density over \((0, L]\).

This is consistent with the observation that for small \( \lambda \), the arrival process becomes close to uniform. Also note that \( \lambda < \frac{b}{L} \) is the stability condition for the system when considered as queueing model.

The solution of (4.8) provides

\[ s^* \approx \frac{L}{1 + \sqrt{r}}, \quad r = \frac{b}{a}, \]  
\[ (4.9) \]

and \( V(L) \), as given by (3.13b) with \( s_c \) replaced by \( s^* \) results in

\[ V(L) \approx \frac{r^2}{a} + \frac{L^2 - s^*}{b} + O(\lambda) \]

and using (4.9), we get after some manipulations

\[ V(L) \approx \frac{L}{1 + \sqrt{r}} + O(\lambda). \]
\[ (4.10) \]

Note the accidental(?) identity of the right-hand-sides of (4.9) and (4.10). The limiting value of (4.10), as \( \lambda \to 0 \) is precisely the value found by [1].

Brief reflection will convince the reader that for non-zero values of \( \lambda \) the probability of interception across equal sections of the trajectory of bit 0 increases \(^3\) with \( s \), so one should expect then higher values for \( V(L) \).

Another instructive quantity is the ratio of \( V(L) \) to its value under the naive policy that maintains the rate constantly at its highest value \( b \). For low \( \lambda \) it provides \( V(L) \approx \frac{b}{2} L + O(\lambda) \), so the relative gain is \( (1 + \sqrt{r})/2 \). Consider that realistic values of \( \sqrt{r} \) are in the range of 10 to 100+!

\[^2\] Theorem 3.3 and the compactness of the policy space allow this borrowing.

\[^3\] More precisely, the probability of interception on \((s, s + z], z < s \leq L \) is \( \exp(-2\lambda z)\exp(-\lambda s) \).
CCD shift registers ...

In the enclosed table we bring some numerical examples obtained by solving (4.7) and substituting in (3.13b), for a range of $\lambda$ and $r$. The values in the row $\lambda = 0$ are from (4.9) and (4.10). The two values for each entry are $s^{*}/L$ above, and $V(L)/L$ below. The values of $\lambda$ are also given in dimensionless form, as $\lambda L$, since $bL$ is the maximal input rate consistent with the stability of the system as we show in the next Section.

5. A QUEUEING-THEORETICAL ANALYSIS

In this Section we compute the distribution of the response time of the device under the optimal control, as shown in the Appendix and in Section 4 (the continuous control case). The behavior of this system is similar to a well-known queueing model, first analyzed in [5]. There is a difference though: there, when the server terminates a busy-period it embarks on a sequence of i.i.d. "vacations"; at the end of each the queue is inspected, and once it is found non-empty, service is resumed. Here the situation is nearly the same, with the vacations having a constant value, except that once there is an arrival, the server "hurries" to finish the vacation, giving it thereby a different duration than the others, a duration moreover, that depends on the instant of arrival.

Thus one cannot use the results of [5] to estimate the performance in this case, but the difference is easy to handle. (Our approach below is rather different from Skinner's).

Let $X$ denote the queue length at the device, and $x_n$ represent the steady-state probability of $n$ requests being queued at service

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$r$</th>
<th>1.25</th>
<th>12.5</th>
<th>125</th>
<th>1250</th>
<th>12500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$=0$</td>
<td></td>
<td>.4721</td>
<td>.2204</td>
<td>.0820</td>
<td>.0273</td>
<td>.0087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4721</td>
<td>.2204</td>
<td>.0820</td>
<td>.0273</td>
<td>.0087</td>
</tr>
<tr>
<td>.0010</td>
<td></td>
<td>.4721</td>
<td>.2204</td>
<td>.0820</td>
<td>.0273</td>
<td>.0087</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4722</td>
<td>.2207</td>
<td>.0824</td>
<td>.0278</td>
<td>.0092</td>
</tr>
<tr>
<td>.0050</td>
<td></td>
<td>.4719</td>
<td>.2200</td>
<td>.0814</td>
<td>.0267</td>
<td>.0081</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4728</td>
<td>.2215</td>
<td>.0835</td>
<td>.0291</td>
<td>.0106</td>
</tr>
<tr>
<td>.0250</td>
<td></td>
<td>.4710</td>
<td>.2180</td>
<td>.0787</td>
<td>.0241</td>
<td>.0061</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4745</td>
<td>.2256</td>
<td>.0893</td>
<td>.0359</td>
<td>.0183</td>
</tr>
<tr>
<td>.1250</td>
<td></td>
<td>.4883</td>
<td>.2083</td>
<td>.0675</td>
<td>.0182</td>
<td>.0029</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.4837</td>
<td>.2462</td>
<td>.1198</td>
<td>.0743</td>
<td>.0626</td>
</tr>
<tr>
<td>.5250</td>
<td></td>
<td>.4434</td>
<td>.1692</td>
<td>.0400</td>
<td>.0088</td>
<td>.0010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.5299</td>
<td>.3519</td>
<td>.2781</td>
<td>.2601</td>
<td>.2569</td>
</tr>
</tbody>
</table>
CCD shift registers ...

completion. These probabilities, if they exist (cf. below), have to satisfy the equations

\[ n = x_0 \sum_{r=1}^{n+1} P(\bar{S} = r)P(\bar{S} = n-r+1) + \sum_{r=1}^{n+1} x_r P(\bar{S} = n-r+1) \quad n \geq 0 \quad (5.1) \]

where \( \bar{S} \) is the number of arrivals during a service and has the probability mass function (pmf)

\[ P(\bar{S} = i) = e^{-\lambda L/b} \frac{(\lambda L/b)^i}{i!} \]

and \( I \) is the number of requests enqueued at the end of an "idle" rotation of the register during which there was at least one arrival.

Denoting by \( X(z) \) the probability generating function (pgf) of \( \{x_n\} \) we immediately get from equation (5.1)

\[ X(z) = \sum_{n=0}^{\infty} x_n z^n = x_0 \frac{S(z)}{z} I(z) + \frac{1}{z} S(z)(X(z) - x_0) \]

\[ = x_0 S(z) \frac{I(z) - 1}{z - S(z)} \quad (5.2) \]

where \( S(z) \), the pgf of \( S \) is given by \( e^{-\lambda L(1-z)/b} \), and \( I(z) \) is the pgf of \( I \) to be computed below.

The empty queue probability is obtained by setting \( z = 1 \) in equation (5.2).

\[ x_0 = (1 - \rho)/E(I), \quad \rho = \lambda L/b \quad (5.3) \]

whence the obvious stability criterion \( \rho < 1 \) follows.

The evaluation of \( I(z) \) is simple once we note that if there is an arrival when bit 0 is at \( s \), the time to the end of that rotation is \( s/b \), and during that period additional arrivals may come, at rate \( \lambda \). Let \( A(\xi) \) denote the Laplace-Stieljes transform (LST) of such an \( s \), then \( I(z) = zA(\lambda(1-z)/b) \).

To determine \( A(\xi) \) we start with the time from the beginning of an "idle" rotation till the first arrival, under the condition that such an arrival does transpire. This time has the density

\[ f(t) = \frac{\lambda e^{-\lambda t}}{1 - e^{-\lambda U}} \quad (5.4) \]

where \( U = (L - s^*)/b + s^*/a \) is the duration of a full (uninterrupted) idle rotation. We shall denote \( (L - s^*)/b \) by \( t^* \); this is the time from the beginning of an idle rotation until \( s^* \) is reached. Given that a first arrival occurred at time \( t \) the time till the end of the rotation, \( T(t) \), is given by
CCD shift registers ...

\[ T(t) = \frac{s(t)}{b} = \begin{cases} \frac{L - bt}{s} & 0 \leq t < t^* \\ \frac{-a(t - t^*)}{s} & t^* \leq t < U \end{cases} \quad (5.5) \]

Using equation (5.4) the LST of this time is immediate if cumbersome:

\[
A(\xi/b) = \frac{1}{1 - e^{-\lambda \xi / b}} \left\{ \int_{0}^{t^*} \lambda e^{-\xi / b} e^{-(L - bt)/b} dt + \int_{t^*}^{U} \lambda e^{-\lambda t} e^{-\xi} \frac{a(t - t^*)}{b} dt \right\} \\
= \frac{1}{1 - e^{-\lambda \xi / b}} \frac{\lambda}{\lambda - \xi} e^{-\lambda \xi / b} (1 - e^{-t^*(\lambda - \xi)}) \\
+ \frac{\lambda}{\lambda - \xi / b} e^{-\xi} \left( e^{-\xi} (\xi - \xi) - e^{-\xi} \right) \quad (5.6)
\]

Since arrivals and departures occur singly, the same distribution of \( X \) as computed above is also seen by arriving customers.

The LST of the response time of the device is given by

\[ W(\xi) = \mathcal{X}(1 - \frac{\lambda}{\xi}) \quad (5.7) \]

The moments of \( X \) (and \( W \)) are simple to extract from formulas (5.2) and (5.6). Thus for example

\[ E(X) = \frac{(1 - \rho)I''(1) + \rho(2 - \rho)I'(1)}{2(1 - \rho)I'(1)} \quad (5.8) \]

where

\[ I'(1) = [\lambda L + \alpha e^{-\lambda \xi} + (b - a)e^{-\lambda t^*}] / b \]

\[ I''(1) = [(\alpha L)^2 + 2\alpha(b - a)e^{-\lambda \xi} + [(2b - 2a) - (a - \alpha \xi)^2]e^{-\lambda t^*}] / [2b] \quad (5.9) \]

6. DISCUSSION

We have demonstrated that the popular notion about the optimal control of a CCD-like shift register is indeed optimal when the control is continuously and instantaneously affected, and arrivals obey the Poisson law. When control actions can only be taken at some discrete set of instants, the bang-bang policy is optimal up to one decision point, but even then it is \( \varepsilon \)-optimal. We chose to consider such discrete instants that correspond to equally separated positions of the register state descriptor, for technical reasons. It does not appear that a different choice (e.g. instants that are temporally equi-separated) would give a
CCD shift registers ...

different answer, but the details were not worked out for this case.

Another assumption we have held throughout is that of the time-homogeneous exponential distribution of the time till the next request arrival. Any deviation from that would result in a considerably more complicated structure; but even then we propose the following:

**Conjecture:** For any time-homogeneous distribution of the time to next request arrival the optimal continuous-time control is bang-bang, with the number of switching points less than or equal to the number of local maxima of the probability density function. A corresponding result holds for the discrete-time problem.

---

**ACKNOWLEDGEMENT**

The remarks of careful referees improved the presentation of the paper. In particular, the Appendix was prepared in response to one of their queries.

**APPENDIX.**

The purpose of this Appendix is to support the claim made in Section 1, that in order to minimize the expected request delay it is optimal to use the maximum rotation rate as long as there are requests to be processed. The claim is not trivial; one can manufacture realizations of the system evolution where such a service is not optimal: suppose when $s = L$, one request is present, and its service starts. The next request is due to arrive just $L / b + \varepsilon$ from now; using $u = b$ will produce a latency of $(L - a\varepsilon)/b$ for that request, whereas slightly reducing the rotation rate would clearly reduce the total delay. However, while this may be the case for certain realizations the optimum in expectation satisfies

**Theorem A** Service of the CCD register should always be performed at the rate $b$ (the highest possible).

**Proof:** We give the essence of the proof. Let $W(i, s),$ $s \geq 0$, $i \geq 1$ denote the expected aggregate sojourn time experienced by requests at the system, beginning at time $t = 0$ to the end of the current busy-period, under the optimal $\Delta$-policy, given that at $t = 0$ bit 0 was at position $s$ and $i$ requests
CCD shift registers ...

were present. A $\Delta$-policy is a policy as introduced in Section 3, that makes decisions when $s = k\Delta$ for $0 < k \leq 2^N$.

Let $W_\Delta(i,s,u)$ be defined as similar to $W_\Delta(i,s)$, except that the action for the first $\Delta$ segment is known to be $u$ (not necessarily optimal) and thereafter the optimal policy is to apply. An immediate calculation yields

$$W_\Delta(i,s,u) = \int_0^{\Delta u} \lambda e^{-\lambda t} \left[ \frac{\Delta u}{u} - t + W_\Delta(i+1,s-\Delta) \right] dt + e^{-\lambda t\Delta u} W_\Delta(i,s-\Delta)$$

$$= (i + 1) \frac{\Delta u}{u} - \frac{1}{\lambda} + (1 - e^{-\lambda t\Delta u}) W_\Delta(i+1,s-\Delta) + e^{-\lambda t\Delta u} \left[ \frac{1}{\lambda} + W_\Delta(i,s-\Delta) \right], \quad (A.1)$$

Noting that $W_\Delta(i,s-\Delta)$ does not depend on $u$, we differentiate equation (A.1) with respect to $u$:

$$\frac{\partial}{\partial u} W_\Delta(i,s,u) = \frac{\partial}{\partial u} \left[ - (i + 1) \frac{\Delta u}{u^2} + \frac{\Delta u}{u^2} e^{-\lambda t\Delta u} \frac{\Delta u}{u} e^{-\lambda t\Delta u} [W_\Delta(i+1,s-\Delta) - W_\Delta(i,s-\Delta)] \right] \quad (A.2)$$

Since the value in the brackets is positive the right-hand-side of (A.2) is negative for $i \geq 0$ and $\lambda \geq 0$. This establishes the Theorem. $\Box$.

REFERENCES