

Achievable Bounds On Signal Transition Activity*

Sumant Ramprasad, Naresh R. Shanbhag, and Ibrahim N. Hajj
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Urbana, IL 61801

Abstract

Transitions on high capacitance busses in VLSI systems result in considerable system power dissipation. Therefore, various coding schemes have been proposed in the literature to encode the input signal in order to reduce the number of transitions. In this paper we derive achievable lower and upper bounds on the expected signal transition activity. These bounds are derived via an information-theoretic approach in which symbols generated by a source (possibly correlated) with entropy rate \mathcal{H} are coded with an average of R bits/symbol. These results are applied to, 1.) determine the activity reducing efficiency of different coding algorithms such as Entropy coding, Transition coding, and Bus-Invert coding, 2.) bound the error in entropy-based power estimation schemes, and 3.) determine the lower-bound on the power-delay product. Two examples are provided where transition activity within 4% and 8% of the lower bound is achieved when blocks of 8 and 13 symbols respectively are coded at a time.

1 INTRODUCTION

The on-chip dynamic power dissipation of CMOS circuits at a node is given by, $P_D = \frac{1}{2}TC_LV_{dd}^2f$, where T is the transition activity at the node, C_L is the capacitance, V_{dd} is the supply voltage, and f is the frequency of operation. At the system level, off-chip busses have capacitances, C_L , that are orders of magnitude greater than those found on signal lines internal to a chip. Therefore, transitions on these busses result in considerable system power dissipation. To address this problem, various signal encoding techniques have been proposed in the literature to encode the data before transmitting it on a bus so as to reduce the expected and the peak number of transitions. Hence, the signal encoding approaches in literature achieve power reduction by reducing T while keeping C_L more or less unaltered. In this paper we derive achievable lower and upper bounds on the expected transition activity for any coding algorithm. These bounds are derived via an information-theoretic approach in which each symbol of a source (possibly correlated) with entropy rate \mathcal{H} is coded with R bits on average. The concept of entropy, \mathcal{H} , from information theory was first employed for single-bit signals, albeit in an empirical manner, in the area of high-level power esti-

mation [2,3]. In contrast, the work presented here is non-empirical, applicable to multi-bit signals, independent of the coding algorithm, and completely unravels the connection between the bounds on transition activity and entropy. This work is a continuation of our effort in developing an information-theoretic view of VLSI computation [5], whereby equivalence between computation and communication is being established.

In section II, we present the preliminaries necessary for the development in the rest of the paper. In section III, the main result is presented in the form of *Theorem 1*. In section IV, we employ *Theorem 1* to, 1.) derive lower and upper bounds for different coding algorithms, 2.) bound the error in entropy-based power estimation schemes, and 3.) determine the lower-bound on the power-delay product. We also present two examples where transition activity within 4% and 8% of the lower bound is achieved.

2 PRELIMINARIES

Let X be a discrete random variable with alphabet \mathcal{X} and probability mass function $p(x) = \Pr(X = x)$, $x \in \mathcal{X}$. A measure of the information content of X is given by its entropy $H(X)$, which is defined as [1], $H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x)$ bits. The joint entropy $H(X_1, X_2, \dots, X_n)$ of a sequence of discrete random variables (X_1, X_2, \dots, X_n) with a joint distribution $p(x_1, x_2, \dots, x_n)$ is defined as, $H(X_1, X_2, \dots, X_n) = -\sum p(x_1, x_2, \dots, x_n) \log_2 p(x_1, x_2, \dots, x_n)$ bits. The entropy rate of a stochastic process $\{X_i\}$ is defined by, $\mathcal{H} = \lim_{n \rightarrow \infty} \frac{1}{n} H(X_1, X_2, \dots, X_n)$ bits, when the limit exists. The function $H(x)$ is defined on the real interval $[0, 1]$ as, $H(x) = -x \log_2 x - (1-x) \log_2 (1-x)$ bits. The function $H(x)$ maps the probability of a binary-valued, independent variable to its entropy. The inverse, $H^{-1}(y)$, of H , is defined on the real interval $[0, 1]$ as, $H^{-1}(y) = x$, if $y = H(x)$ and $x \in [0, \frac{1}{2}]$. The function $H^{-1}(\cdot)$ maps the entropy of a binary-valued, independent variable to a probability value that lies between 0 and $\frac{1}{2}$. The transition activity of a bit-level signal, b_n , is defined as, $t = \Pr(b_n = 0 \text{ and } b_{n-1} = 1) + \Pr(b_n = 1 \text{ and } b_{n-1} = 0)$. We define a level encoding algorithm to be the case where the symbol '0' is coded with the bit '0' and the symbol '1' is coded with the bit '1'. In a transition encoding algorithm, the symbol '0' is coded by transmitting the previous transmitted bit and the symbol '1' is coded by trans-

*This work was supported by National Science Foundation CAREER award MIP-9623737 and Joint Services Electronics Program under contract #00014-96-1-0129.

mitting the complement of the previous transmitted bit. Hence, a ‘1’ is coded with a transition and a ‘0’ with no transition. Table 1 shows an example of transition coding.

3 ACHIEVABLE BOUNDS

In order to derive bounds on transition activity, we will employ Lemmas 1 and 2 presented below. Lemma 1 bounds x given $y \leq H(x)$. Lemma 2 employs Lemma 1 to bound the expected number of 1’s in a sequence of bits with a certain entropy rate \mathcal{H} . Theorem 1 employs Lemma 2 to bound the number of transitions/symbol of a source with a certain entropy rate \mathcal{H} given that each symbol is coded employing an expected number of R bits. The proofs are presented in the Appendix.

Lemma 1 For all (x, y) such that $x \in [0, 1]$ and $y \in [0, 1]$, if $y \leq H(x)$ then, $H^{-1}(y) \leq x \leq 1 - H^{-1}(y)$.

Lemma 2 If $\{B_i\}$ is a 0-1 valued stochastic process with entropy rate greater than or equal to \mathcal{H} and if $p_b = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n B_i$ exists then, $H^{-1}(\mathcal{H}) \leq p_b \leq 1 - H^{-1}(\mathcal{H})$.

Theorem 1 Let,

1. \mathcal{H} be the entropy rate of a stochastic source $\{X_i\}$,
2. the symbols be coded in a uniquely decodable manner into bits $\{B_i\}$ employing an expected number of $R (> \mathcal{H})$ bits/symbol,
3. the bits be transmitted in some arbitrary manner over a finite set of wires such that a receiver can uniquely decode the bits, and
4. T be the expected number of transitions in the bits on the $\frac{1}{n} \sum_{i=1}^n B_i$ wires per symbol (i.e. $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n B_i \oplus B_j$ exists, where B_j precedes B_i on the same wire and \oplus is the exclusive-or operator) then, $H^{-1}(\frac{\mathcal{H}}{R})R \leq T \leq (1 - H^{-1}(\frac{\mathcal{H}}{R}))R$.

The lower and upper bounds on transition activity computed by Theorem 1 for different values of R are shown in Figure 1. Any coding algorithm will need to reside in the region shown in Figure 1. The transition activity can be made arbitrarily close to 0 by increasing R . In practice, however, R will typically be less than approximately $10\mathcal{H}$ because most of the reduction in the lower bound is achieved by the time $R = 10\mathcal{H}$.

4 APPLICATIONS OF BOUNDS

In this section, we provide applications of Theorem 1. We first employ Theorem 1 and the value of R specified by a coding algorithm to derive bounds for that algorithm.

4.1 Entropy Coding

An entropy code is one for which the expected number of bits/symbol is equal to the entropy of the source.

Corollary 1 For an entropy coder, $T = \frac{\mathcal{H}}{2}$.

Proof: From Theorem 1 and since $R = \mathcal{H}$ for an entropy coder we have, $H^{-1}(1)\mathcal{H} \leq T \leq (1 -$

$H^{-1}(1))\mathcal{H}$. Substituting $H^{-1}(1) = \frac{1}{2}$, we get $T = \frac{\mathcal{H}}{2}$. \blacktriangleleft

Therefore, as shown in Figure 1, the lower and upper bounds on T are identical for entropy coding.

4.2 Transition Coding

We will redefine transition coding slightly in this subsection as coding the less probable symbol (‘0’ or ‘1’) with a transition and the other symbol with no transition.

Corollary 2 Transition coding achieves the lower bound on the transition activity for an i.i.d. source with alphabet $\mathcal{X} = \{0, 1\}$ when $R = 1$ bit/symbol.

Proof: From Theorem 1 and since $R = 1$, we have $H^{-1}(\mathcal{H}) \leq T \leq 1 - H^{-1}(\mathcal{H})$. Since the source is i.i.d. with alphabet $\{0, 1\}$, $H^{-1}(\mathcal{H})$ is also the probability of a ‘0’ or the probability of a ‘1’, whichever is less. Hence we can achieve the lower bound on the transition activity by coding the less probable symbol (‘0’ or ‘1’) with a transition and the other symbol with a no transition. \blacktriangleleft

The upper bound on transition activity of $1 - H^{-1}(\mathcal{H})$ is greater than or equal to the upper bound of $\frac{\mathcal{H}}{2}$ in [2]. This is because, in the proof of the upper bound in [2], the implicit assumption is made that the symbol ‘0’ is coded with the bit ‘0’ and the symbol ‘1’ is coded with the bit ‘1’, i.e., level coding is employed. It is possible to achieve the higher transition activity of $1 - H^{-1}(\mathcal{H})$ for the same entropy if the more probable symbol is coded with a transition.

4.3 Bounds For 1-bit Redundant Codes

In algorithms such as Bus-Invert coding [6], the transition activity on the bus is reduced by employing an additional bit. We now calculate the lower bound for any coding algorithm that uses 1 bit of redundancy. Thus $R = \mathcal{H} + 1$ and from Theorem 1, the expected transition activity is bounded by, $(\mathcal{H} + 1)H^{-1}(\frac{\mathcal{H}}{\mathcal{H} + 1}) \leq T \leq (\mathcal{H} + 1)(1 - H^{-1}(\frac{\mathcal{H}}{\mathcal{H} + 1}))$. The extra bit may be either an extra line on the bus, or an extra clock to transfer the data. If $\mathcal{H} = 8$ then the lower bound is 2.4506 transitions/symbol. For uniformly distributed, temporally independent data, Bus-Invert coding achieves a transition activity of 3.269 transitions/symbol which is 33.40% above the lower bound. The lower bound can be approached by coding larger and larger blocks of bits. Now, assume the source entropy, \mathcal{H} , is increased and $R = \mathcal{H} + 1$. The ratio $\frac{\mathcal{H}}{\mathcal{H} + 1}$ approaches 1 and T approaches $\frac{\mathcal{H}}{2}$. Thus, as the entropy increases, the benefit of Bus-Invert coding or any 1-bit redundant code is reduced. The above analysis can be extended for a k -bit redundant code.

4.4 Error Bounds For Entropy-Based Power Estimation Schemes

The bounds on transition activity from Theorem 1 can be employed to calculate the maximum error in schemes that estimate power dissipation from entropy [2, 3]. For given \mathcal{H} and R , the error in estimating T , will be less than or equal to the maximum of

the difference between the upper bound and the estimate of T , and the difference between the estimate of T and the lower bound. As an example, assume that a source with entropy $\mathcal{H} = 6$ bits is transmitted over an 8-bit bus. Since $R = 8$, the lower bound on transition activity at the bus can be calculated from Theorem 1 to be 1.716. The upper bound, assuming no glitching, is 6.284. Assuming an estimate, T_e , for transition activity, the error in the estimate is less than $\max(T_e - 1.716, 6.284 - T_e)$. If additional information beyond \mathcal{H} and R is not available, then it is not possible to obtain tighter bounds on the error in estimating transition activity. This is because the bounds on transition activity are achievable.

4.5 Lower Bound On Power-Delay

If the capacitance C_L , the supply voltage V_{dd} , and the frequency of operation, f , are given, then the minimum average power dissipation is proportional to the lower bound on the transition activity. The delay (for instance, for transmitting the data on a bus) is proportional to R . Hence the lower bound on the power-delay product, $PowerDelay_{min}$, given \mathcal{H} and R , is given by, $PowerDelay_{min} = KH^{-1}(\frac{\mathcal{H}}{R})R^2$, where K is a constant of proportionality. The graph of $PowerDelay_{min}$ versus R for a given value of \mathcal{H} is shown in Figure 2. For given \mathcal{H} , we can find the R that minimizes $PowerDelay_{min}$ by equating its derivative with respect to R to 0. The value of R that minimizes $PowerDelay_{min}$ is found to be, $R_{min,power-delay} = 1.25392 \mathcal{H}$. Thus, a source with entropy rate, \mathcal{H} , requires approximately an average of $1.25\mathcal{H}$ bits/symbol to encode for minimum power-delay product. If $R > 1.25\mathcal{H}$, then the delay increases resulting in a non-optimal power-delay product. Similarly, if $R < 1.25\mathcal{H}$ then the power component increases because less redundancy is being added.

4.6 Bounds For An i.i.d. Source

Consider an i.i.d. source with a 5 symbol alphabet $\mathcal{X} = \{A, B, C, D, E\}$ with probabilities $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, and $\frac{1}{16}$ respectively. The entropy rate is equal to $\frac{15}{8}$ bits. If an average of $R = 3$ bits are employed to code a symbol then from Theorem 1, the bounds on transition activity are, $0.468426 \leq T \leq 2.531574$. We now calculate the actual transition activity that is achieved by various coding algorithms and compare them with the bounds. To simplify the calculation of transition activity, we assume transition coding is employed to transmit the bits, i.e., a '1' is transmitted with a transition and a '0' is transmitted with no transition. Thus, the number of transitions is equal to the number of 1's. We can make the assumption of transition coding in the examples because the purpose of the examples is to show the existence of coding algorithms that approach the lower bound.

Entropy Coding Followed By Spatial Redundancy Coding: In this algorithm, we initially employ an entropy coder to code the symbols employing the minimum expected number of bits/symbol. A code that achieves the entropy is A = 0, B = 10, C = 110, D = 1110, and E = 1111. At the output of the entropy coder we have an average of $\frac{15}{8}$ bits/symbol.

Since $R = 3$ bits/symbol, if we code a block of 15 bits from the entropy coder with 24 bits, we can employ the redundancy to reduce the number of transitions. This is an extension of Bus-Invert coding and results in a code in which the expected number of transitions (or 1's) is 0.565499 transitions/symbol, which is within 18% of the lower bound.

Probability Based Coding: An alternative algorithm to reduce the number of transitions is by coding the most probable symbol A as 000 or no transitions, B = 001, C = 010, D = 100, and E = 011. The expected number of transitions/symbol is 0.5625, which is within 17% of the lower bound. We can further reduce the number of transitions by coding blocks of symbols. The transition activity/symbol for different block sizes is shown in Figure 3. Thus, we can achieve a transition activity within 4% of the lower bound by employing a block size of 8.

4.7 Bounds For A Markov Process

Consider the 3-state stationary Markov process U_1, U_2, \dots having the transition matrix P_{ij} in Table 2 [1]. Thus the probability that S_1 follows S_3 is equal to zero. An algorithm to encode the process will consist of 3 codes C_1, C_2 , and C_3 (one for each state S_1, S_2 , and S_3), where C_i is a code mapping from elements of the set $\{S_1, S_2, S_3\}$ into a code-word in C_i (see Table 3 for an example). To encode the current symbol S_j , we note the previous symbol S_i and select code C_i . We send the code-word in C_i corresponding to S_j . This is repeated for the next symbol. The stationary distribution of this Markov chain is $\mu = [\frac{2}{9}, \frac{4}{9}, \frac{1}{3}]^T$. The entropy rate of the stationary Markov process is given by [1], $H(\mathcal{X}) = -\sum_{ij} \mu_i P_{ij} \log_2 P_{ij} = \frac{4}{3}$ bits. A code that achieves the entropy rate is shown in Table 3. If $R = 2$ bits/symbol, then from Theorem 1, $0.347904 \leq T \leq 1.652096$ transitions/symbol. The transitions/symbol for entropy coding followed by redundancy coding and probability based coding is shown in Figure 4 for different block sizes. We can achieve a transition activity within 8% of the lower bound with a block size of 13 symbols.

APPENDIX

We now present proofs of Lemma 2 and Theorem 1. The proof of Lemma 1 is omitted because of lack of space and because it can be verified from a plot of $H(\cdot)$. The proofs of achievability of Lemma 2 and Theorem 1 [4] are also omitted due to lack of space.

Proof Of Lemma 2

From the definitions of entropy rate and \mathcal{H} in the statement of Lemma 2, $\mathcal{H} \leq \lim_{n \rightarrow \infty} \frac{1}{n} H(B_1, B_2, \dots, B_n) \Rightarrow \mathcal{H} \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n H(B_i)$ [Independence bound on entropy] $\Rightarrow \mathcal{H} \leq H(\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n B_i)$ [Jensen's inequality and concavity of H] $\Rightarrow \mathcal{H} \leq H(p_b)$ [Since $p_b = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n B_i$]. Thus, we can substitute \mathcal{H} for y and p_b for x in Lemma 1 to obtain Lemma 2. \blacktriangleleft

Proof Of Theorem 1

Let (B_1, B_2, \dots, B_n) be the n bits that encode a block of N symbols (X_1, X_2, \dots, X_N) , where this mapping from the symbol sequence (X_1, X_2, \dots, X_N)

to the bit sequence (B_1, B_2, \dots, B_n) can only be either one-to-many or one-to-one (and not many-to-one or many-to-many) because the coded bitstream needs to be uniquely decodable. Since the symbols are coded employing an expected number of R bits/symbol, for large N , $n = NR$. The entropy rate, \mathcal{H}_b , of the code bits is given by, $\mathcal{H}_b = \lim_{n \rightarrow \infty} \frac{1}{n} H(B_1, B_2, \dots, B_n)$
 $\Rightarrow \mathcal{H}_b = \lim_{n \rightarrow \infty} \frac{1}{n} \sum p(b_1, b_2, \dots, b_n) \log_2 p(b_1, b_2, \dots, b_n)$ [From the definition of joint entropy]
 $\Rightarrow \mathcal{H}_b \geq \lim_{N \rightarrow \infty} \frac{1}{NR} \sum p(x_1, x_2, \dots, x_N) \log_2 p(x_1, x_2, \dots, x_N)$ [From $n = NR$ and the possibility of a 1-many mapping. The inequality will be an equality if the mapping from symbols to bits is one-to-one.]
 $\Rightarrow \mathcal{H}_b \geq \frac{\mathcal{H}}{R}$.

Define a function g on (B_1, B_2, \dots, B_n) as, $(C_1, C_2, \dots, C_n) = g(B_1, B_2, \dots, B_n)$, where, $C_i = B_i \oplus B_j$, and B_j and B_i are transmitted on the same wire and B_j immediately precedes B_i . If B_i is the first bit transmitted on the wire, then B_j is '0'. Bit C_i is transmitted on the same wire and in the same order as bit B_i . Clearly, we can compute B_i given C_i as, $B_i = C_i \oplus C_j$, where C_j and C_i are transmitted on the same wire and C_j immediately precedes C_i . If C_i is the first bit transmitted on the wire, then C_j is '0'. Since (B_1, B_2, \dots, B_n) and (C_1, C_2, \dots, C_n) are functions of each other and $\mathcal{H}_b \geq \frac{\mathcal{H}}{R}$ we have, $\lim_{n \rightarrow \infty} \frac{1}{n} H(C_1, C_2, \dots, C_n) = \lim_{n \rightarrow \infty} \frac{1}{n} H(B_1, B_2, \dots, B_n) = \mathcal{H}_b \geq \frac{\mathcal{H}}{R}$. Let $p_c = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n C_i$, where p_c is the probability of C_i being a '1'. The limit exists because of assumption (4) in Theorem 1. Since C_i is a binary random variable, $p_c n$ is the expected number of 1's in (C_1, C_2, \dots, C_n) for large n . Substituting p_c for p_b and $\frac{\mathcal{H}}{R}$ for \mathcal{H} in Lemma 2 we have, $H^{-1}(\frac{\mathcal{H}}{R}) \leq p_c \leq 1 - H^{-1}(\frac{\mathcal{H}}{R})$. Since there are on the average R bits/symbol and C_i is '1' iff there was a transition at B_i , $T = p_c R$. Multiplying the inequality $H^{-1}(\frac{\mathcal{H}}{R}) \leq p_c \leq 1 - H^{-1}(\frac{\mathcal{H}}{R})$ by R and substituting $p_c R$ with T we have Theorem 1. \blacksquare

References

- [1] T. M. Cover and J. A. Thomas, "Elements of Information Theory," John Wiley & Sons, 1991.
- [2] D. Marculescu, R. Marculescu, and M. Pedram, "Information theoretic measures for power analysis," *IEEE Trans. CAD*, pp. 599–610, June 1996.
- [3] M. Nemani and F. Najm, "Towards a High-Level Power Estimation Capability," *IEEE Trans. CAD*, pp. 588–598, June 1996.
- [4] S. Ramprasad, N. R. Shanbhag, and I. N. Hajj, "Achievable bounds on signal transition activity," *submitted to IEEE Trans. VLSI*.
- [5] N. R. Shanbhag, "A mathematical basis for power-reduction in digital VLSI systems," *IEEE Trans. CAS II* (to appear).
- [6] M. R. Stan and W. P. Burleson, "Bus-Invert Coding for Low-Power I/O," *IEEE Trans. VLSI*, pp. 49–58, March 1995.

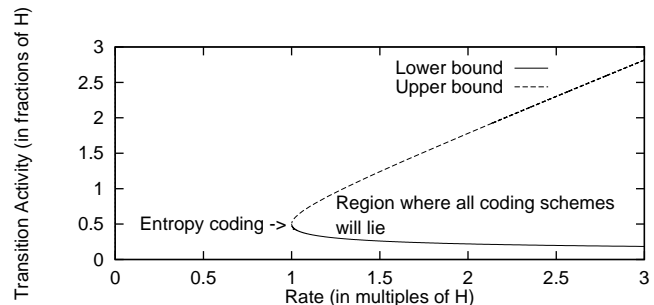


Figure 1: Bounds on Transition Activity vs. R

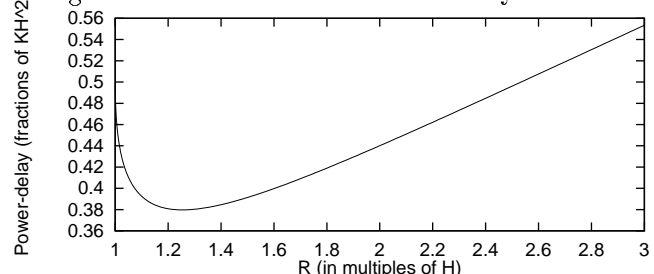


Figure 2: Lower bound on power-delay

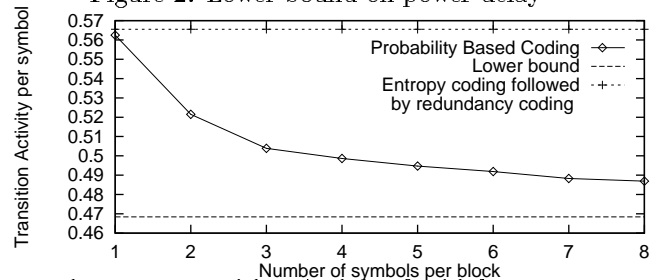


Figure 3: Transition Activity for i.i.d. source

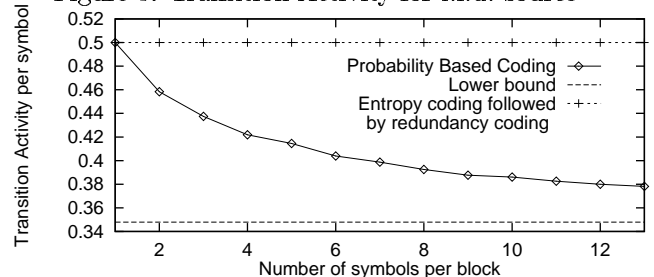


Figure 4: Transition Activity for Markov Process

Table 1: Transition Coding

Time	Level Code	Transition Code
0	000000	000000
1	000011	000011
2	000010	000001
3	010000	010001

Table 2: Transition Matrix

Previous state	S_1	S_2	S_3
S_1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
S_2	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
S_3	0	$\frac{1}{3}$	$\frac{1}{3}$

Table 3: Entropy Code for Markov Process

	S_1	S_2	S_3
C_1	0	10	11
C_2	10	0	11
C_3	—	0	1