

DATA TRANSMISSION OVER A BUS WITH PEAK-LIMITED TRANSITION ACTIVITY

Vijay Sundararajan

Keshab K. Parhi

Dept. of ECE University of Minnesota
Minneapolis, MN 55455

E-mail: vijay@ece.umn.edu, parhi@ece.umn.edu

ABSTRACT

Transitions on high capacitance busses in VLSI systems result in considerable power dissipation. Various coding schemes have been proposed in literature to encode the input signal in order to reduce the number of transitions. Reducing number of transitions comes in exchange for redundancy in data transferred over the busses. For a given amount of redundancy there exists a lower bound on the average number of transitions. In recent times noise and reliability problems have brought the peak/instantaneous power consumed in VLSI systems in to prominence. There has been limited study done on reducing the number of instantaneous transitions and hence the peak power consumed in busses. In this paper we model a bus with a limit on the maximum instantaneous transition activity as a constrained channel and derive an upper bound on the data-rate obtainable using the capacity of the underlying channel. We then demonstrate that some existing bus encoding schemes are near-optimal with respect to the derived bounds thus, perhaps, obviating the need to search for newer more complicated coding schemes. Also considered is a bus with a constraint on number of transitions in a fixed number of (k) bus transmissions. The capacity of such a bus is derived in the same manner as a bus with a constraint on maximum instantaneous transition activity.

1. INTRODUCTION

Busses constitute an important resource for addressing and data transfer in VLSI systems implementation. Reducing the power consumed in busses while transferring data is therefore a high priority objective in minimizing power consumption for the entire system. The fact that the power consumed in bus accesses account for a significant fraction of the total power consumed in VLSI systems has been independently established by many researchers, [1], [2]. The primary reason for this happens to be the fact that the capacitance of shared busses is quite large in comparison to the capacitance of other data-path units. There are essentially two ways to reduce average power consumption in busses. The first one involves minimizing bus accesses by either reducing the number of data-path units connected to large busses [2] or reducing the number of accesses of READ/WRITE busses for large memory units by algorithm transformations [3]. The second way to reduce power consumed in busses is to reduce the effective capacitance of busses by reducing bus transition activity. In this regard many researchers have studied reduction of bus transition activity by resorting to coding, much in the vein of error-correcting codes, [4], [5], [6].

While the previously stated methods except [5], [6] target average power consumed in data-busses another important metric coming in to prominence these days is peak power consumption. There are cases when the peak power can be much higher than the average power, and this leads to an undesired increase in simultaneous switching noise [7], metal electro-migration problems [8], and local physical deformations due to nonuniform temperatures on the die.

In [4] lower/upper bounds are established on the average power consumed in data-busses using Shannon's channel coding theorem and the concept of Entropy. However, these bounds do not give much information on the peak transition

activity and thereby the peak power consumption. In this paper we will model a bus with a limit on the maximum number of instantaneous transitions as a constrained channel. The capacity of the constrained channel then gives us the maximum data-rate possible for data transmission over this bus.

In section 2 past work is put in perspective. In section 3 an upper bound is derived on the maximum possible data transmission rate over a bus that has an arbitrary limit on the instantaneous switching activity. Also considered in section 3 is a bus with a constraint on number of transitions in a fixed number of (k) bus transmissions. The capacity of such a bus can be derived in the same manner as a bus with a constraint on maximum instantaneous transition activity. Analytical results showing maximum transmission rates for several busses with various limits on the instantaneous switching activity is tabulated in section 4. In section 5 it is shown that an existing coding scheme called limited-weight coding [5] is near-optimal with respect to the capacity bounds derived in this paper. Finally in section 6 the paper is summarized with conclusions.

2. PAST WORK

In the past there has been one significant contribution on studying the average power consumed in data-busses [4]. In this section, we will give a brief outline of this technique.

2.1. Bounds on Bus Transition Activity

2.1.1. Random Variables and Entropy

Let X be a discrete random variable with alphabet X and probability mass function $p(x) = \Pr(X = x)$, $x \in X$. A measure of the information content of X is given by its entropy $H(X)$, which is defined as follows [9],

$$H(X) = - \sum_{x \in X} p(x) \log_2 p(x) \text{ bits.} \quad (1)$$

This definition of the measure of information implies that the greater the *uncertainty* in the source output, the higher is its information content. In a similar fashion, a source with zero uncertainty would have zero information content and therefore its entropy, from (1), would be equal to zero.

In [4] it is shown that if we use M bits on an average to transmit data-words of size W from a source with entropy of H bits per word then the average transition activity T is bounded on both sides as,

$$H^{-1}\left(\frac{H}{M}\right)M \leq T \leq \left(1 - H^{-1}\left(\frac{H}{M}\right)\right)M, \quad (2)$$

and the bounds in (2) are asymptotically achievable.

3. CAPACITY OF A CONSTRAINED BUS

The bounds in [4] which were discussed in the last section are for average number of transitions for data-transmission over a data-bus. These bounds give no information about

the peak transition activity in the transmission process. The peak power consumed in transmitting data over a data bus is directly dependent on the peak transition activity. Also, this quantity is completely independent of the statistical characteristics of the data-source at the transmitting end. The peak transition activity is a function of the channel (bus) characteristics alone. In order to limit peak transition activity we must completely eliminate bus transitions which result in exceeding a desired number of transitions.

Imagine a data-bus which is W -bits wide, in addition, let the peak instantaneous transition activity on this data-bus be limited to d -bits with $d \leq W$. Such a bus will from now onwards be referred to as a d peak limited W -bit bus. Such a bus can be described by a constraint graph $G = (V, E)$ as follows. There is a vertex corresponding to every possible state of the bus. The state of a bus is determined by the data pattern being transmitted over it. A W -bit bus, for example, has 2^W states. Now there is an edge $e_{uv} \in E$ providing the states corresponding to nodes u and v have a Hamming distance $\leq d$. We number these vertices in such a manner that a vertex u corresponds to a unique number $i_u \in \{0, \dots, 2^W\}$. The adjacency matrix $A(G)$ of this constraint graph is a $|V| \times |V|$ (0,1) matrix that has an entry of 1 in the $(i_u, i_v)^{th}$ location provided there is an edge $e_{uv} \in E$ and an entry of 0 otherwise. It can be shown [10] that for this matrix the eigenvalue with largest absolute magnitude is real and positive. Let the largest positive real eigenvalue corresponding to $A(G)$ be λ . Then the maximum data-rate that can be sustained over this data-bus is upper bounded by $\log_2(\lambda)$.

Example 1 Consider data transmission over a 2-bit bus. Assume that there is a limit of one on number of peak instantaneous transitions. The constraint graph corresponding to this bus has 4 states as shown in Fig. 1. Also the adjacency matrix for this graph, shown in Fig. 1 has a maximum positive real eigenvalue of 3 so the channel capacity of this channel is $\log_2(3) = 1.585$ bits/transmission.

Another way of modeling a peak limited data bus is by means of a trellis diagram which can be derived from the constraint graph. The trellis diagram for the bus whose constraint graph is depicted in Fig. 1 is shown in Fig. 2.

We will now show that the capacity of a d peak limited W -bit bus is $\log_2(\lambda)$ where λ is the maximum positive eigenvalue of the adjacency matrix A of the constraint graph of this bus. Now the capacity of such a bus can be expressed in terms of bits per word (transmitted). Denoting by $N(n)$ the number of allowed word sequences of length n . We have the capacity C is given by,

$$C = \lim_{n \rightarrow \infty} \frac{\log_2 N(n)}{n}. \quad (3)$$

The matrix A^n can be used to find the capacity. Intuitively, we choose the number of stages n to be quite large and count the number of paths of length n through the trellis from any initial state to any final state. This is the sum of the elements in A^n . Therefore,

$$C = \lim_{n \rightarrow \infty} \frac{\log_2 \sum_{ij} (A^n)_{ij}}{n}, \quad (4)$$

where $(A^n)_{ij}$ is the ij th entry of A^n . Now we can express A as $A = UDU^{-1}$ using the Jordan Canonical form of A [11], with U being an unitary matrix and D a diagonal matrix. Hence, $A^n = UD^nU^{-1}$, which implies that the elements of A^n are fixed linear combinations of the elements of D^n .

The sum $\sum_{ij} (A^n)_{ij}$ will include the term $\mu\lambda^n$ where μ is a constant independent of n and λ is the largest eigenvalue of A . Asymptotically, the sum will be dominated by the term $\mu\lambda^n$. Then,

$$\begin{aligned} C &= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \sum_{ij} (A^n)_{ij}, \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \log_2 \mu \lambda^n. \end{aligned} \quad (5)$$

The terms $(1/n)\log_2(\mu)$ can be dropped as n goes to infinity. Thus $C = \log_2\lambda$.

A *maximal capacity* peak limited bus is one which allows all transitions except the ones where all the bits in the bus switch simultaneously. That is, a $W-1$ peak limited W -bit bus is a maximal capacity bus. The capacity of such a bus can be shown to be $\log_2(2^W - 1)$. Here we will just show that $2^W - 1$ is an eigenvalue of the adjacency matrix of the constraint graph corresponding to such a bus.

Example 2 The adjacency matrix, A , of the constraint graph for a $W-1$ peak limited W -bit bus has 1s every where except in the principle anti-diagonal. If we add all columns in A together then we get a column vector with all entries $= 2^W - 1$. Now, consider the matrix $A - (2^W - 1)I$ where I is the identity matrix of appropriate dimensions. If we add all columns of $A - (2^W - 1)I$ together then we get a 0 column vector as the sum of all columns for any row of A yields $2^W - 1$. Therefore the determinant of $A - (2^W - 1)I = 0$. Hence, $2^W - 1$ is an eigenvalue of A .

A *minimal capacity* peak limited bus is one which allows only those transitions where at most a single bit on the bus switches in successive transitions. That is, a 1 peak limited W -bit bus is a minimal capacity bus. The capacity of such a bus can be shown to be $\log_2(W + 1)$. The bus whose constraint graph is depicted in Fig. 1 is both a maximal and a minimal capacity peak limited bus.

Example 3 The adjacency matrix, A , of the constraint graph for a 1 peak limited W -bit bus has precisely $W + 1$ 1s in every row. If we add all columns in A together then we get a column vector with all entries $= W + 1$. Now consider the matrix $A - (W + 1)I$ where I , as before, is the identity matrix of appropriate dimensions. If we add all columns of $A - (W + 1)I$ together then we get a 0 column vector as the sum of all columns for any row of A yields $W + 1$. Therefore the determinant of $A - (W + 1)I = 0$. Hence, $W + 1$ is an eigenvalue of A .

In general a d peak limited W -bit bus can be shown to have a capacity of $\log_2(\sum_{k=0}^d \binom{W}{k})$. As before the adjacency matrix, A , of the constraint graph for a d peak limited W -bit bus has precisely $\sum_{k=0}^d \binom{W}{k}$ 1s in every row. If we add all columns in A together then we get a column vector with all entries $= \sum_{k=0}^d \binom{W}{k}$. Hence as for the case with a maximal and minimal capacity W -bit bus $\sum_{k=0}^d \binom{W}{k}$ is an eigenvalue of A .

So, we have completely characterized the maximum data-rate for all peak limited data-busses.

3.1. Minimizing Transition Activity over a few Transmissions

Sometimes it is desirable to limit the power consumed over a few clock cycles than the peak instantaneous power. This is particularly true if limiting the peak power results in extremely low data-rates.

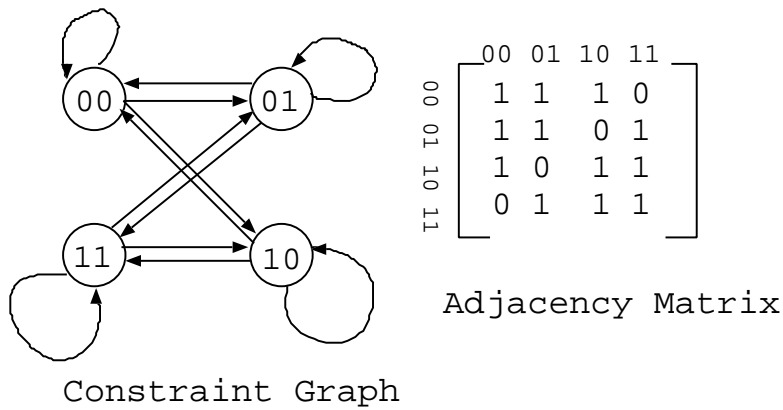


Figure 1. The constraint graph and adjacency matrix of a 2-bit bus with a peak limit of 1 transition between successively transmitted data words.

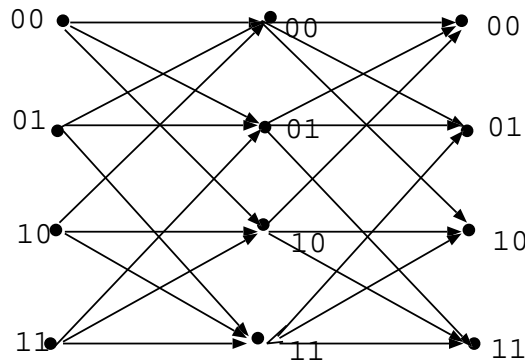


Figure 2. Trellis for corresponding to the constraint graph in Fig. 1.

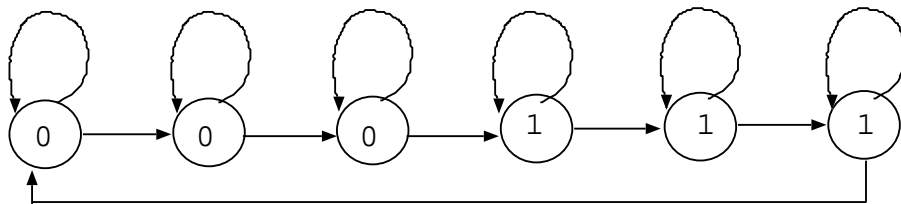


Figure 3. Constraint graph of a 1-bit bus with a limit of 1 transition in four successively transmitted data words.

Table 1. Capacity of busses for a variety of peak limits. The first column tabulates the bus width. The second to ninth columns show the peak limit (P. L.) in bits and the capacity (C) in bits/transmission. Bus widths ranging from 4-64 and peak limits ranging from 1-8 have been tabulated.

Bus Width bits	P. L./ C bits/Trans.	P. L./ C	P. L./ C	P. L./ C	P. L./ C	P. L./ C	P. L./ C	P. L./ C
4	1/2.3219	2/3.4594	3/3.9069					
8	1/3.1699	2/5.2095	3/6.5392	4/7.3487	5/7.7748	6/7.9484	7/7.9944	
16	1/4.0875	2/7.0980	3/9.4450	4/11.2975	5/12.7492	6/13.8623	7/14.6846	8/14.7609
32	1/5.0444	2/9.0471	3/12.4223	4/15.3390	5/17.8896	6/20.1320	7/22.1063	8/23.8416
64	1/6.0224	2/11.0231	3/15.4168	4/19.3733	5/22.9853	6/26.3114	7/29.3920	8/32.2565

Rather than limiting the number of transitions between successive words transmitted on a bus to be $\leq d$ we can limit the number of transitions over a few bus transmissions (say k) by requiring that the number of bus transitions for k consecutive bus transmissions is always $\leq k \times d$ (a fixed number). The constraint graph for such busses can be constructed in the same manner as for busses restricting peak transition activity. Calculating the number of states and the edges is, however, more complicated. An example of the partial constraint graph for a 1-bit bus having a limit of at most 1 transition in four successive bus transmissions is shown in Fig. 3. Also, to be noted is the fact that the capacity of such busses can once again be computed as for the busses with peak limited transition activity.

4. ANALYTICAL RESULTS

In Table 1 we show the capacity of busses for several bus widths and various peak limits. As can be observed even for severe peak limits amazingly high maximum data-rates are possible, e.g., for a peak limit of 8 in a 64-bit bus maximum data-rate/capacity = 32.2565. While the data in Table 1 gives very useful information about maximal data-rates achievable for various peak limits on busses, we now show that the limited weight coding scheme described in [5] are near-optimal codes in terms of efficiency.

5. NEAR-OPTIMAL LIMITED WEIGHT CODES

This method uses a M bit bus to transmit W bit data with $M - W$ bits serving as steering bits which serve to reduce the Hamming distance between successively transmitted words to be under a predetermined constant d . In order to limit the Hamming distance between successively transmitted words to be $\leq d$ we must satisfy the following inequality,

$$\sum_{i=0}^d \binom{M}{i} \geq 2^W. \quad (6)$$

The minimum value M satisfying (6) dictates the most efficient word size to be used for transmitting the data-words over the data-bus. Clearly a limited weight code is optimal with respect to the bus capacity derived previously if and only if the capacity of the bus = W . The capacity of the bus is given by $\log_2(\sum_{i=0}^d \binom{M}{i})$. It turns out that the only limited weight coding scheme that is efficient is the so called bus-invert coding scheme derived in [6] which leads to a $W/2$ peak limited $(W + 1)$ -bit bus. This follows from the fact that for even values of W , $\sum_{i=0}^{\frac{W}{2}} \binom{W+1}{i} = 2^W$. All other limited weight coding schemes are sub-optimal.

6. CONCLUSIONS

In this paper we have derived the maximal transmission rate for a bus of arbitrary width that has an arbitrary limit on the number of bits that can switch between successively transmitted words. Analytical results shown in this paper seem to indicate that even with very tight limits on the instantaneous switching activity a high data transmission rate is possible. We then demonstrated that an existing coding scheme called limited-weight coding [5] does indeed come close to achieving the capacity bounds derived in this paper, thus, perhaps, obviating the need to design more complicated coding schemes.

REFERENCES

- [1] D. Liu and C. Svensson, "Power Consumption Estimation in CMOS VLSI Chips," *IEEE Journal of Solid State Circuits*, vol. 29, no. 6, pp. 663–670, 1994.
- [2] R. Mehra, L. M. Guerra, and J. Rabaey, "Low Power Architectural Synthesis and the Impact of Exploiting Locality," *Journal of VLSI Signal Processing*, vol. 13, no. 2-3, pp. 239–258, 1996.
- [3] S. Wuytack, *et al.*, "Global Communication and Memory Optimizing Transformations for Low Power Systems," in *Proceedings of International Workshop on Low Power Design*, (Napa, CA, USA), pp. 203–208, April 1994.
- [4] S. Ramprasad, *et al.*, "Achievable Bounds on Signal Transition Activity," in *Proceedings of ICCAD'97*, (San Jose, CA, USA), pp. 126–129, Oct. 1997.
- [5] M. Stan and W. Burleson, "Limited-Weight Codes for Low-Power I/O," *Proceedings of International Workshop on Low Power Design*, pp. 209–214, April 1991.
- [6] M. Stan and W. Burleson, "Bus-Invert Coding for Low-Power I/O," *IEEE Transactions on VLSI Systems*, vol. 3, pp. 49–58, March 1995.
- [7] M. Dolle, "Analysis of Simultaneous Switching Noise," in *Proceedings of ISCAS-95*, (Seattle, WA, USA), pp. 904–907, June 1995.
- [8] A. Dasgupta and R. Karri, "Hot-carrier Reliability Enhancement via input reordering and transistor sizing," in *Proceedings of the 33rd ACM/IEEE Design Automation Conference*, pp. 819–824, Jun. 1996.
- [9] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, 1991.
- [10] R. E. Blahut, *Digital Transmission of Information*. Addison-Wesley Publishing Company, 1990.
- [11] G. Strang, *Linear Algebra and its Applications, Third Edition*. San Diego, CA: Harcourt Brace Jovanovich, 1988.