ISCAS 2000 - IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland

COMBATING DIGITAL NOISE IN HIGH SPEED ULSI CIRCUITS USING BINARY BCH ENCODING

Dinesh Pamunuwa, Li-Rong Zheng and Hannu Tenhunen

ESD Lab, Department of Electronics, Royal Institute of Technology
Electrum 229, Isafjordsgatan 22-26, SE-164 40 Kista, Sweden
dinesh | lrzheng | hannu @ele.kth.se

ABSTRACT

Increased integration in deep submicron (DSM) technologies has caused very high increases in the RLC parasitics which affect the coupling of noise to signals propagating over interconnect. Error free transmission on-chip will no longer be guaranteed. This paper examines the issue of high speed signaling in DSM and proposes the use of particular BCH codes to improve the bit error rate in the face of noise. We conclude from our results that it is possible to achieve a considerable coding gain by choosing the code properly.

1. INTRODUCTION

Advances in IC fabrication technology have led to the Moore law scaling of device density in VLSI chips and have resulted in equally dramatic increases in device speed. This has a profound impact on the signaling that takes place on-chip and off-chip. Maintaining signal integrity when moving a bit from one point to another presents as great a challenge, if not more, as that involved in the high speed design of the logic itself. In fact it is the signaling delay and noise considerations that will impose the primary limitation on the clock speed of future circuits. When moving towards 0.1 μm technologies it is anticipated that signaling from microprocessor chip edge to chip edge will require 12 clock cycles, which will severely restrict the target implementation architectures. One proposed model architecture style (Ref. [1] and [2]) consists of a large number (from 10s to 100s) of smaller processing elements (PEs), each of size 50-100 k gates (exact size defined by the interconnect delay and buffering needs) interacting via global communication channels. These smaller blocks will prevent customary, well established design methodology from breaking down in DSM technologies because interconnect delay will not have disastrous effects within these blocks. However these PEs still need to communicate with each other in a very noisy environment over relatively long interconnects. In this paper we focus on signaling techniques and issues relevant to such communication needs (edge-to-edge within a PE, or PE to another close-by PE) of a processor.

The noise consists of various components and three of the most significant sources are interconnect delay (causing inter-symbol interference ISI), cross-talk, and power supply noise. Various remedies have been proposed to deal with these different kinds of noise. They include techniques which try to overpower the noise (as is currently done in standard ASIC CMOS design) and techniques which attempt to isolate the signal from the noise source. This paper reports the results of using error-control coding to improve the bit error rate (BER) under the influence of noise in general. It is a logical consequence of Shannon's famous theorems showing the increase of the data rate over a noisy channel beyond a certain limit necessitates the use of error control codes (ECCs).

This paper is organised as follows. In part 2 our channel model is introduced and it is shown that it accurately models the effects of the different noise sources mentioned above. The next section shows the coding gain from particular primitive narrow sense binary BCH codes. Finally we summarize our results and outline our future work.

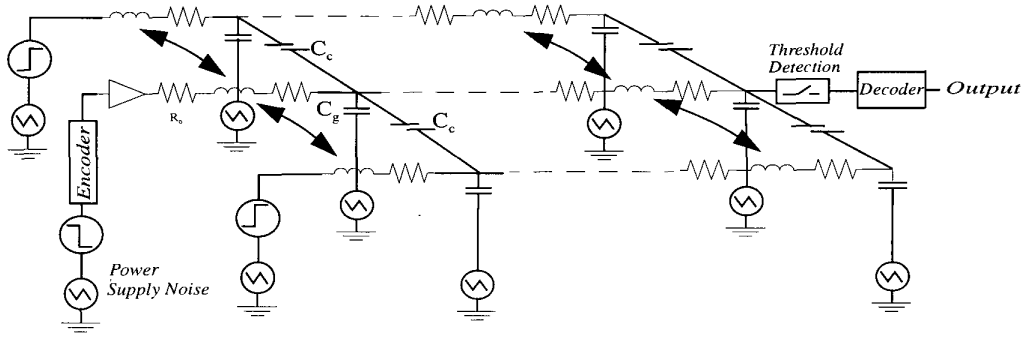
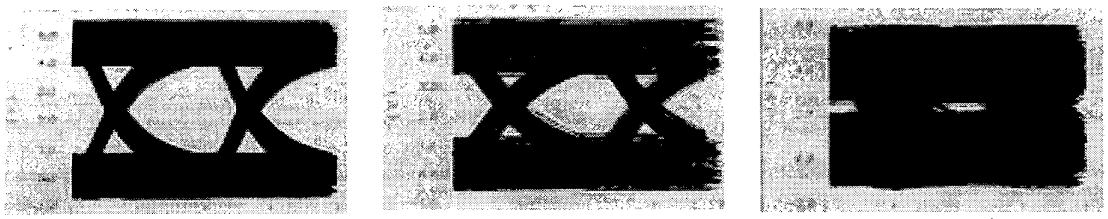


Figure 1. Driven line (victim) coupled to two lines (aggressors) switching in random fashion

0-7803-5482-6/99/$10.00 ©2000 IEEE



Aggressors switch in same direction | Adjacent interconnect grounded | Aggressors switch in random fashion

Figure 2. Eye diagrams for 700 MHz signaling Over 1.5 mm metal 1 interconnect for parasitics extracted for a 0.05 micron technology with minimum spaced wire geometry.

2. NOISE ANALYSIS AND MODELING

We use for our simulations, a distributed RLC line which models the effects of delay, cross talk and power supply noise. We encode a pseudo random bit sequence (PRBS) and feed it into the central line while two other PRBSs are fed to the adjacent lines. Sampling at the output is by means of threshold detection at half amplitude. The output from the detector is then decoded to give the final output. The following paragraphs discuss various noise sources and how accurately they are modelled in this channel.

The smaller cross sections of interconnects in DSM technologies result in a big increase in the resistive parasitics. This in turn means that the channel shows predominantly RC or *diffusive* behaviour, in spite of the inductive parasitics. A characteristic of diffusive propagation is signal dispersion which is much worse than a pure delay, when bit pipelining is possible. Now if the transmitted pulse spreads out into other time slots it causes ISI (the result of memory in the channel which causes consequent symbols to overlap). Since there should be sufficient time for the signal to attain steady state before the next bit is dumped onto the line, dispersion imposes a fundamental limitation on the bandwidth. ISI is a very serious problem if not countered and can contribute to a significant increase in the BER. Our distributed channel models the interconnect delay very accurately.

Crosstalk occurs primarily because of capacitive coupling between interconnect lines. In DSM technology, inductive coupling is also a factor (see Ref. [3]). Common terminology is to call the line under consideration the victim, and the coupled adjacent switching lines the aggressors. Depending on whether either or both of the aggressors switch in the same direction as the victim or not, the coupling coefficient will change. It can be seen from Fig. 1, that both capacitive and inductive coupling are modelled in our channel. If adequate measures are not taken to combat cross talk, the system can completely fail as will be shown in the next section.

Power supply noise is principally the difference in voltage caused by the drop across the parasitic impedances of the power supply network and is essentially a deterministic signal, since its spectrum depends to a large extent on the current profile of the switching logic blocks. The charging/discharging currents create a peak considerably higher than the average, causing inductive drops. A statistical approximation similar to that given in Ref. [4] is used to determine the logic current profile. The chip is assumed to be composed of 50k gate modules (in agreement with our original premise as given in the introduction), and a certain fraction of these gates is assumed to switch in any given clock cycle. These switching gates are distributed in a triangular manner over several stages, resulting in a triangular current profile, the dimensions of which depend on the rise time and the switching load. The power supply network can also be subject to LC ringing. Typical values result in peaks in the spectrum at around 85 MHz corresponding to the package resonance frequency and at around 370 MHz corresponding to tank ringing. In our channel, the effects due to the transient load currents -calculated as explained above, and LC ringing are modelled by adding components at the relevant frequencies to random noise. Hence the white spectrum of random noise is effectively coloured by the spectral content of the noise caused by the charging discharging transients and by the package resonance frequency components. Different composite noise files corresponding to different random distributions corrupt the ground taps.

3. ERROR CONTROL CODING TO IMPROVE BIT ERROR RATE

3.1 Signaling over Medium Length Interconnect in DSM Technology

The distances that can be considered as medium length in the context of the modular architecture described above are in the range of a few millimetres. Given in Fig. 2 are three eye diagrams from simulations run over 1.5 mm interconnect in a hypothetical 0.05 μm technology at a signaling frequency of 700 MHz. The effect of delay and signal dispersion can easily be seen. The rounding of the edges and a horizontal shift of the eye are clearly discernible. The effect of power supply noise which has corrupted the ground and Vdd rails and caused a reduction in the eye opening can also be observed. It is however the effect of cross-talk that is most dramatic. The first diagram in Fig. 2 shows the simulated eye diagram of the response to an input PRBS with the aggressors "helping" the victim so to speak, by carrying the same data stream, and hence switching in a similar manner. A very clear and well defined eye-diagram is discernible. The second figure shows the case where the two adjacent lines are grounded and a slight decrease in the eye opening can be seen. The third figure shows the eye diagram where the two aggressors run different PRBSs. The difference is very evident: the eye opening has almost completely closed.

The driver in the above simulations used transistors sized to 30 times the minimum size. When the length is increased to 3mm, the eye opening closes completely regardless of the driver transistor siz-

ing. It is a combination of ISI and cross talk which kills the signal. The well known technique of combating delay and hence ISI, is to break up the interconnect into smaller segments and insert repeaters. Repeater insertion is universally accepted as reducing the delay from a quadratic to a linear dependence on length. (Ref. [5] presents for the first time an analysis for a distributed line model taking into account inductive parasitics. It shows that the traditional treatment of the line as a pure RC channel can result in oversizing of the buffers.) The eye diagram of the output over a single interconnect can be greatly cleaned up by buffering which reduces the effect of delay degradation and hence ISI.

However, if the adjacent interconnects are buffered as well, which would be the case in a parallel bus, the effect of cross talk is compounded and we are only marginally better off with regard to the eye opening than with the unbuffered case.

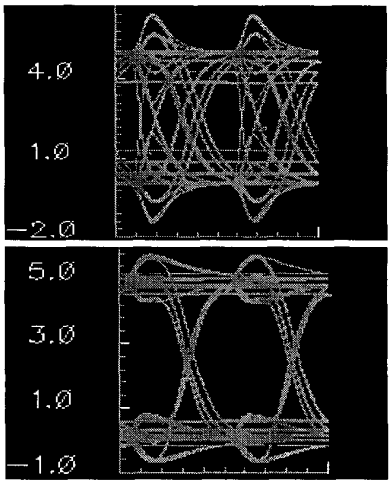


Figure 3. Eye diagrams for signaling with aggressors buffered and unbuffered

Fig. 3 shows the eye diagrams for signaling over 2mm interconnect, where in the first case just the central line is buffered, and in the second, the flanking lines are also buffered. (The sizing and numbering is according to Ref[5]). The second eye opening is much reduced and hence sampling errors are more likely.

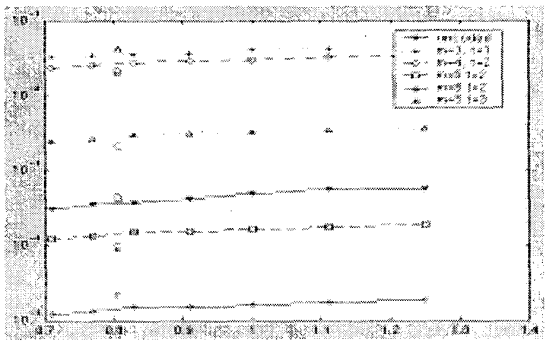
The most straightforward estimate of the signaling integrity of a system is the BER. In the next section we show how we have used ECCs to improve the BER for signaling on inter module length interconnect.

3.2 Signaling with Binary BCH Encoding

One of the most important families of ECCs is BCH codes, which are a powerful and popular class of linear cyclic block codes well suited to coping with random errors (see Refs. [6], [7] and [8]). Their popularity is due in fairly large part to the existence of computationally efficient and easily implementable decoding procedures. A background in finite field arithmetic is necessary to understand the genesis of BCH codes. The basic idea is that the field consists of q elements, and the operations of addition and multiplication are defined to conform to certain rules. The field of q elements is depicted as GF(q). When q is the integer power of 2, $GF(2^m)$ are formed and these extension fields are the basis for working with BCH codes. Now the elements are defined in terms of a primitive element α which is a root of a primitive polynomial.

For a linear block code to be capable of correcting τ errors per block it must have a minimum Hamming distance of $2\tau+1$. Binary BCH codes are built by constructing the generator polynomial so that its roots contain 2τ consecutive powers of β where τ is the number of errors per block to be corrected and β is an element of order n from $GF(2^m)$. Here n is the number of bits per code word. The encoder takes the input word and imparts the necessary redundancy. The decoder receives the transmitted word and calculates its syndrome, and maps it to an error pattern. It is the second step in the decoding procedure which is computationally heavy and poses area and speed constraints. It involves formulating a set of simultaneous equations from the syndrome, the solution of which gives a connection polynomial, which in turn gives the location of the errors by its roots. It is possible to use FFT techniques to perform coding and decoding, which lend themselves to fast implementations. Also there has been a lot of investigation recently into area and power efficient finite field arithmetic implementations (see Refs. [9] and [10]).

The increase in the number of bits introduced by the code reduces the overall information throughput. To maintain the same effective information rate it is necessary to increase the energy per information bit E_b by an amount proportional to $1/\rho$, the code rate. Here we present in Fig. 4, curves of BER for different generator polynomials plotted against the pk-pk power supply noise for a normalized E_b.



A: 4/7 single ECC; **B**: 7/15 double ECC; **C**: 16/31 triple ECC; **D**: uncoded data; **E**: 21/31 double ECC; **F**: 63/51 double ECC

Figure 4. Curves of BER with different codes in the face of ISI, cross talk and power supply noise.

It can be seen that a properly designed code can result in significant coding gains. The BER for uncoded data over a 2mm long interconnect in a 0.05μm technology with minimum spacing at a frequency of 500 MHz, is approximately 4 in 10^4. When encoding with a rate 4/7 single ECC generated from an extension field of degree m=3, which is basically a Hamming code, it is necessary to transmit at a bit rate of 500MHz times 7/4 (reciprocal of ρ), to maintain the same information rate. This results in a much worse performance, with the BER increasing to almost 3 in 10^2. A rate 7/15 double ECC over $GF(2^4)$ results in a marginal improvement, but is still much worse than the uncoded BER. However a rate 21/31 double ECC over $GF(2^5)$ results in an improvement of the BER to approximately 1.5 in 10^4. This is a three fold improvement over the uncoded BER. This can be explained by the fact that the increase in ρ caused by the

increase in complexity of the extension field allows the same information rate to be maintained at a signaling frequency lower than with the rate 7/15 code. The most dramatic improvement is given by the rate 51/63 double ECC which causes a drop in the BER to almost 3 in 10^5. This is again a consequence of the increase in ρ.

It is also interesting that a rate 16/31 triple ECC performs better than a rate 7/15 double ECC with approximately the same ρ. This seems to indicate that errors are more likely to occur in groups of three rather than two and serves to emphasise the fact that it is rather difficult to formulate a mathematical model for the channel and accordingly select a code, as is usually done in communication applications. The nature of the errors depends very much on the layout and on the correlation between bit streams. Simulation is thus an invaluable tool in selecting a proper code. Deciding on the appropriate code depends very much on the application.

4. SUMMARY

DSM technology poses a lot of challenges to reliable high speed signaling. Repeater insertion reduces delay and hence ISI, but it must be remembered that there are other sources of noise which can corrupt the signal. In particular cross talk in closely spaced parallel buses cannot be countered except with insertion of repeaters far in excess of the optimum with respect to minimizing delay. This increases both the total delay and power consumption, introduces additional jitter and complicates the design of the power supply net. The obvious solution is to run the lines far apart or to shield them with grounded lines, but with any sizeable bus, this would entail a lot of wasted area and the purpose of integration would be defeated. Also there will be fairly severe restrictions on routing due, among other things, to the electromigration of Al wires. Transistors however are plentiful and a relatively complex block of logic will be almost free.

We have shown that binary BCH ECCs in such noisy channels can greatly improve the BER with a judicious choice of Galois field complexity and ρ. It is not feasible to derive a mathematical, frequency dependent model which mirrors the myriad types of noise present in an on-chip environment to select the code with the necessary gain at the required BER. However simulations allow the choice of a code that gives a decent coding gain. Now since CMOS is increasingly cheap, and maintaining signal integrity over interconnect increasingly dear, the area and power overheads of the encoding and decoding circuitry no longer present unacceptable bounds for special data buses. It must be noted that this overhead must be offset against the considerable saving in area and power resulting from the reduced numbers and sizes of drivers.

It is important that ECCs are a sledgehammer method to reduce the BER which does not care about the roots or causes of the noise, and only works on the bits in error. Hence it is especially useful to reduce the incidence of spurious errors. Obviously the coding and modulation methods depend on the low level signaling scheme used. For the research presented in this paper, voltage mode signaling has been used. Some of the results may change for current mode signaling, when for example the impact of power supply noise is reduced through isolation. Thus more detailed analysis into architectural and circuit models is necessary, and more effort in establishing robust design guidelines and rules is needed. A new interconnect centric design methodology needs to be developed.

5. REFERENCES

[1] D Sylvester et al., "Getting to the Bottom of Deep Submicron", IEEE/ACM International Conference on Computer Aided Design, 1998, Digest of Technical Papers pp. 203-211

[2] A Hemani et al., "Lowering power consumption in clock by using Globally Asynchronous Locally Synchronous design style", in proc. 36th Design Automation Conference 1999, pp. 873-878

[3] L R Zheng et al., "Transient and Crosstalk Analysis of Interconnection Lines for Single Level Integration Packaging Modules", in proc. IEEE 7th Topical Meeting on Electrical Performance of Electronic Packaging, 1998

[4] W J Dally et al., "Digital Systems Engineering", Cambridge University Press, 1998

[5] Y I Ismail et al., "Effects of Inductance on the Propagation Delay and Repeater Insertion in VLSI Circuits", in proc. 36th Design Automation Conference 1999, pp. 721-724

[6] C B Rorabaugh, "Error Coding Cookbook", McGraw-Hill, 1996

[7] S Lin et al., "Error Control Coding: Fundamentals and Applications", Prentice-Hall, 1983

[8] E R Berlekamp, "Algebraic Coding Theory", McGraw-Hill, 1968

[9] S K Jain et al., "Efficient Semisystolic Architectures for Finite-Field Arithmetic", in IEEE Transactions on VLSI Systems, March 1998, Vol. 6 1 pp.101-113

[10] L L Song et al., "Low-Energy Programmable Finite Field Data Path Architectures", in proc. IEEE International Symposium on Circuits and Systems 1998, Vol. 2, pp. 406-409