Global Optimization Approach to Transistor Sizing for High Performance CMOS VLSI circuits [1]

Sharad Mehrotra, Paul Franzon[2], Wentai Liu[3]

NCSU-VLSI 93-10

November 1993

Picosecond Digital Electronics Laboratory
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh NC 27695-7911

Abstract

A stochastic global optimization approach is presented for skew minimization in CMOS VLSI circuits. This is a direct search strategy for the best design among feasible ones, with the designer determining when the search is stopped. Through examples, we show the power of this technique in quickly obtaining very good designs, even for constrained problems.

[1] This report has been submitted to *31st ACM/IEEE Design Automation Conference*,, October 1993

[2] partially supported by MIP-901704 and an NSF National Young Investigator award

[3] partially supported by MIP-9212346

1 Introduction

In designing high performance CMOS circuits, it is often necessary to properly size the various transistors in a skeletal cell, in order to meet performance requirements. For example, transistor sizing has been extensively used for delay and power optimization of digital CMOS circuits. [3], [8]. In this paper, we describe a sizing program based on stochastic optimization of a model function. This is a direct search method which is more accurate than traditional nonlinear optimization programs, and is considerably faster than simulated annealing.

There are two basic approaches for transistor sizing that have been explored by various researchers. The first approach involves developing a simplified model of signal delay through a CMOS gate, either analytically [3], or by macromodels based on simulations [8]. Then, the delay of the entire cell is computed using this model. The transistor sizing problem is then formulated as an optimization problem with the objective of minizing the delay, as predicted by the model. Some optimization problems have a special form to guarantee efficiency and accuracy of the optimization process, e.g. the posynomial objective function, used in [3], [5]. Otherwise, a nonlinear optimization tool is required.

The second approach involves coupling a circuit simulator to a nonlinear optimization tool (or tool set). For example, Ochotta et. al. [13] employ an augmented asymptotic waveform evaluation technique to evaluate the behavior of each circuit visited by a simulated annealing program In Delight.Spice [12], a set of nonlinear optimization programs are integrated with the SPICE circuit simulator.

The first, 'equation based', approach, though effective for delay optimization, is difficult to generalize to other problems, such as delay skew optimization, as is required for wave-pipelined circuits. Skew, in general, cannot be estimated using a single process corner or data vector. Maintaining accuracy in estimating the spread in delay over input data vector and process variations using a simple model is very difficult. The only solution is to evaluate each sizing scheme through detailed circuit simulations, to account for data and process variations. This makes objective computation a very expensive task, even for very small sized circuits.

The second, 'simulation based' approach has the combined computational burden of running a non-linear optimization program and running at least one full circuit simulation each time the objective function needs to be evaluated. There are some additional hindrances in employing a conventional non-linear optimization program for this task:

1. Gradient information is very difficult to obtain. Though there are numerical optimization techniques which do not require explicit gradient information, these techniques tend to be slow. Also, they try to evaluate the gradient through perturbation. This implies a further increase in the number of objective evaluations.
2. Transistor sizes can only be varied in certain quanta. Most numerical optimization techniques operate on a continuous parameter range. Hence the final solution might be an infeasible sizing scheme. Moving the solution to the closest feasible size may lead to a sub-optimal solution.
3. Constraints in the optimization problem, e.g. area and power, further complicate the optimization task.
4. The user has no direct control over the optimizer, i.e., the optimization task is not interactive, and it is difficult for the engineer to use his or her judgement in guiding the optimizer.
5. The optimization routines look for strict local minima. Usually, the designer is interested only in obtaining a rough approximation to a globally optimal solution. To achieve the global minimum, the optimizer has to be run from multiple, random, initial solutions. Even then, there are not even theoretical guarantees of achieving a globally optimal solution, except for some very restricted problems.

One way of avoiding gradient evaluation and restricting the solutions to feasible sizings is by employing simulated annealing [6] for optimization. However, simulated annealing programs tend to be prohibitively slow, specially when each objective evaluation is so expensive.

In view of these difficulties, we present a new approach to transistor sizing in small, high performance circuit blocks. This approach is based on stochastic modeling of the circuit responses of interest. It is a direct search for the best design among feasible ones. No gradient computations are required. The designer has direct control over the number of simulations conducted, and the search process can be stopped any time the designer is satisfied with the best solution produced thus far. The stochastic model helps in identifying the most promising design based on the existing information about the problem. Only the most promising designs are simulated. Hence simulations are organized naturally and efficiently. The method is capable of identifying the global minimum much faster than exhaustive search.

Recently, global optimization methods have been proposed for optimization of continuous functions based on stochastic modeling [17], [10]. Since this approach is relatively unknown in the circuit optimization domain, we will briefly review its important features in the next section. In section III we will propose an optimization algorithm consistent with the stochastic modeling philosophy. In section IV we present results obtained on two example circuits. Finally, section V is devoted to a discussion of the scope and limitations of this method.

2 Optimization by stochastic modeling of objective function

Consider the following unconstrained optimization problem

$$\min_x \quad \zeta(x) \quad x \in A \subset R^d \tag{1}$$

where x is a d dimensional vector, A is a finite subset of R^d. $\zeta(x)$ is the objective function whose value at any $x \in A$ can be determined only through an expensive simulation. Besides this, there is very little information about the objective function. Suppose the objective function is perceived to be continuous and "smooth", but not unimodal. In such a situation, it is reasonable to approximate $\zeta(x)$ by a simpler function and to perform an optimization on this simplified function. However, a global polynomial approximation is inappropriate; information is lost in fitting the model to data, unless the degree of the polynomial is as large as the data set. A better approach to function approximation is needed.

Recently, stochastic models have been proposed to capture complex objective functions [14]. With this approach, the value of the unknown function at each point in A is assumed to be a random variable. Then, the unknown function itself is a *sample path of a stochastic function*. In the general case, a stochastic function $\phi(x)$ is defined by a family of multidimensional probability distributions $F_{x_1,\ldots,x_m}(y_1,\ldots,y_m) = P(\phi(x_i) < y_i, i = 1,\ldots,m)$. For example, if this distribution is joint Gaussian, then the stocha
the *a priori* average function $\mu(x)$ and covariance $\sigma(x_i, x_j)$. If k values of the stochas
function are known, $\phi(x_i) = \zeta(x_i)$, $i = 1,\ldots,k$ then the *conditional distribution* of $\phi($
at any x is normal with the mean value

$$m_k(x \mid \phi(x_i) = \zeta(x_i), i = 1,\ldots,k) = \mu(x) + (\sigma(x, x_1),\ldots,\sigma(x, x_k)) R_k^{-1} (\zeta(x_1) - \mu(x_1),\ldots,\zeta(x_k) - \mu(x_k))^T \quad (2)$$

and variance

$$s_k^2(x \mid \phi(x_i) = \zeta(x_i), i = 1,\ldots,k) = \sigma(x, x) - (\sigma(x, x_1),\ldots,\sigma(x, x_k)) R_k^{-1} (\sigma(x, x_1),\ldots,\sigma(x, x_k))^T, \quad (3)$$

where R_k^{-1} is the inverse of the $k\times k$ covariance matrix of the randompro
where the function values are known.

Using a stochastic model, and a set of "measurements" on the ob
prediction of the value of $\zeta(x)$ at untried points can be made usin
bution of $\phi(x)$. This prediction is only probabilistic, i.e., at
distribution associated with the possible values for $\zeta(x)$, speci
given above. This allows a search strategy for points of small $\zeta($
on the conditional distribution. It seems more likely to find a p
value where $m_k(x \mid \ldots)$ is small. However, large values of $s_k^2(x \mid \ldots)$ indicate regions of g
uncertainity, i.e. regions where function values can differ great
Hence a rational choice has to be discriminate between points of s
ance or points of small variance but somewhat larger mean. We sha
proposed algorithms that fit the stochastic modeling paradigm.

stochastic modeling of unknown functions, see reference [14]. Gl
are further described by Mockus in [10].

2.1 Algorithms based on Statistical Modeling

Several algorithms for optimization using a stochastic model fun
We summarize some interesting approaches and finish with the P-Al
the basis of our optimization procedure. These approaches essent
model chosen and the method used for minimizing the model functio

In Groch et. al.[4], the model in the multidimensional case i
function. Instead, the conditional mean and variance of the one d
are generalized. The choice of the next point of evaluation is ma

$$r_i^{k+1}(x) = \eta_i^k(x) - c\sigma_i^k(x), \quad i = 1, \dots, n, \tag{4}$$

where the experimental region A is divided into N disjoint simpli
constant, determining the weight given to variance with respect t
of each new point causes the experimental region to be further su
feature of this method is that it is simple; r can be minimized ana
imation to the global optimum can be located by minimizing r. A lo
be then performed.

In Adachi et. al.[7], the model function is a stationary stoch
tional mean is an interpolating function. This interpolating fu
points and its derivatives are easy to compute. The variance of a
sponse by this function is also easy to compute. The optimal poin
the interpolating function starting from the smallest data poin
a constraint on the coefficient of variation. The coefficient of var
mean and variance of the conditional distribution. The procedure
leading to (hopefully) a different local optimum of the objective f
all local optima can be located. The auxiliary computations are
of a nonlinear optimizer at every iteration to find the minimum of t

Bernardo et. al. [1], employ a stochastic model function to pe
of electronic circuits. Their approach relies heavily on desig
significant parameters through parameter effect plots and also to
regions in the design space.

The P-algorithm was developed and characterized in [19], [20]
cedure. At each iteration, a new observation point x_{k+1} is chosen that has the highe
probability of $\phi(x_{k+1})$ being smaller than y_{ok} which is some chosen value smaller t
mean value of ϕ at each point in A i.e.,

$$x_{k+1} = \mathrm{Arg}\max_{x\in A} P_x(y_{ok}) \qquad (5)$$

is chosen as the next observation point where y_{ok} is less than $\min_{x\in A} m_k(x \mid$
$x_i, \zeta(x_i), i = 1, \ldots, k)$, and

$$P_x(y_{ok}) = \text{Probability}(\phi(x) \le y_{ok}) \qquad (6)$$

Based on rather intuitive axioms, it is shown in [17] that $\phi(x)$ can
sian random variable whose conditional mean $m_k(x)$ and variance $s_k^2(x)$ is given as:

$$m_k(x \mid x_i, \zeta(x_i), i = 1, \ldots, k) = \sum_{i=1}^{k} w_i^k \zeta(x_i) \qquad (7)$$

$$s_k^2(x \mid x_i, \zeta(x_i), i = 1, \ldots, k) = \gamma_k \sum_{i=1}^{k} (\sigma(x, x) - \sigma(x, x_i)) w_i^k \qquad (8)$$

where w_i^k are weights chosen such that $\sum_{i=1}^{k} w_i^k = 1$ and $m_k(x \mid x_i, \zeta(x_i), i = 1, \ldots, k) = \zeta(x_i)$
at the k observed points, i.e. the mean value interpolates the kno
proved that a sequence of points thus generated, converges to the

The P-algorithm is a general formulation of a strategy to max
gained by each function evaluation, and is quite easy to implemen
mentation of the P-algorithm [17], several decisions have to be
and the accuracy of the method, namely the following:

- The appropriate form of the weighting function w has to be chosen.

- The appropriate form of the covariance $\sigma(x, y)$ has to be chosen. that σ should be such that $(\sigma(x, x) - \sigma(x, z)) = \|x-z\|$ where $\|x-z\|$ is th norm.

- An appropriate search method for finding x_{k+1} must be used. This could be an multimodal optimization problem.

- An appropriate value of y_{ok} has to be chosen. It has been shown in [20] th a value of y_{ok} leads to the points of greatest uncertainty. If y_{ok} is greater than or equal to $min_{x \in A} m_k(x \mid x_i, \zeta(x_i), i = 1, \ldots, k)$, then the next point chosen will values that attain this minimum.

  In the next section, we detail our implementation of the P-alg iterative, hence the user can stop the iterations any time the satisfactory. The number of simulations to be run are directl The algorithm only identifies the most promising points for si

3 Implementation

The P-Algorithm is a formalization of an intuitive search strat a framework for devising global optimization algorithms. We nov implementation of the P-Algorithm.

1. Choose k points x_i, $i = 1, \ldots, k$ uniformly from A using Latin Hypercube S and compute $\zeta(x_i)$ by simulation. Start iteration $l = 1$.

2. Using the BLUP and MSE expression in [15] (see Appendix 1, eq find the mean $m(x_j)$ and variance $s(x_j)$ at $N \gg k$ uniformly distributed point

3. Find the smallest value of $m(x_j)$, i.e.,

$$m_l(x) = \min_{j \in 1 \ldots N} m(x_j) . \qquad (9)$$

   Let $y^l_{ok} = m_l(x) - \epsilon_l$. At each x_j find the probability $P_{x_j}(y^l_{ok})$.

4. Choose n_l points with largest probability $p(x \mid y_k)$ from the N points.

5. Compute $\zeta(x)$ at the n_l points found above. If $\min_{j \in 1 \ldots n_l} \zeta(x_j)$ is satisfactory, the
stop, else continue

6. $k = k + n_l$ If $k > K_{max}$, then stop, else $l = l + 1$, go to step 2.

This algorithm is parameterized by k, n_l and K_{max}. These parameters have to be adapted to the specific problem or left to the designer' $k = 10 * d$, $n_l = 2d$, where d is the dimensionality of the design space A, good values for problems below. In this way the designer directl simulations to be run. The choice of N search points can be suitabl judgement, and can account for constraints on the design space, e a polytope constrained by linear inequalities on the design vari the mean and variance has to be estimated at the N points, which i

This leads us to another very important question, namely, hand design. If the constraints are linear, they can be handled very m of the N points for violation. When the constraints are implicit a after simulation, e.g. a maximum delay or power restriction when d the procedure needs to be modified. There are two immediate possib can be evaluated through the same simulation, then another stoch to model the constraint. A certain tradeoff has to be established secondary model and the actual objective. If a penalty method is true objective and the constraint in a single objective function is a little more difficult, as the overall model might have to accour Alternatively, the constraint can be modeled piece-wise linearl experiment [16]. Then the optimization task is considerably simp depends on the severity of the constraints. For a very tight con pay close attention to the constraining function and use the firs For a loose constraint, the second approach would be more suitab given below, the first approach was adopted for meeting a maximum addition to optimizing the delay skew.

In the next section, the algorithm described above is executed o
to be optimized for maximum delay variation. In the first example,
with a maximum delay constraint. The second is an essentially unc
problem.

4 Optimization examples

4.1 Delay Controlled Elements for Wave-pipelined circuits

The design of wave-pipelined circuits involves very careful cont
in the combinational blocks. Techniques have been proposed by I
for balancing the path delays by inserting active delay elements
have shown how the delay of each gate can be accurately controlled
For CMOS gates, however, the delay is data dependent. For the CM
example, the rising delay is substantially smaller when both inp
opposed to one input being fixed at 1 and the other switching fro
avoiding this data dependence is to use the cross coupled biased-
in figure 1 [2]. This gate, however, consumes substantial static
low. Another gate structure suitable for wave pipe-lining is sh
transistor M3 is used to add extra resistance to the pull-up chai
simultaneous switching of both inputs. It also has the deleterio
circuit. Hence a proper balance has to be struck between the maxi
gate, as well as the data-dependent spread [11]. The delay spread
process variations also. Of course, the easiest parameters to co
the transistor sizes. Hence the goal of the optimization is to obt
such that the delay spread through each circuit block is minimize
on the maximum delay through the circuit.

The optimization problem is formalized as follows:

$$\text{Find}^{*} x = \text{Arg}\min_{x \in A} \max_{V} \delta^{*}(x) \tag{10}$$

$$\text{subject to} \quad {}_{V}\,\text{dmealxay}(x)_{max} \leq D \tag{11}$$

Here V denotes the nominal and the four process corner MOSFET
hypercube formed by restricting the widths of M1-M3 bet_{bias}ween 3.6μ
between 0.0 and 2.0 V. The widths of N1 and N2 are constrained to b
of M1 and M2 respectively. x is an arbitrary vector of feasible t
voltage. Note that the minimum allowed feature size is 0.6μm and
and N2 were restricted to vary in quanta of 0(.x6) μimsoddefyne dThæs tkheew δ
variation in delay through the circuit shown in figure 2 over the si
(see figure 3), and the *delay*(x) is the largest delay ov$_{max}$eritshtelse input
maximum delay constraint.

The models for delay and skew were initially established by sim
sizing schemes, selected randomly using Latin Hypercube Sampling
1 shows the sizing scheme with the best skew value, satisfying th
these 100 points. This sizing scheme is not feasible. The secon
feasible point to this sizing and the delay and skew value for tha

Since the number of possible sizings is small, all the feasibl
sizing schemes) were evaluated with five values of bias voltage ra
equations 17 and 19. This constitutes an exhaustive search of th
models. Since the smallest possible va$_{sk}$lwaesfcohrotheskebweizse 0 ,ρ .y 20
feasible sizes and bias voltages with the largest probability tha
D_{max} $\leq 1 ns$ were chosen for resimulation. Table 1 shows the resul
simulations. The best sizing in the second set was considerably
the first 100 samples and was considered quite suitable for desi
simulations were performed. The total time taken for simulation
DECstation 5000 while the overhead of model building and searchi
second.

This example illustrates how the optimization procedure is em

is pruned by the designer's judgement and a good solution is found
In the next example, the methodology is further expanded to incl
with a very different objective formulation.

4.2 CMOS Clock Driver Circuit

The second example we present here is the skew optimization of a si
clock driver circuit shown in figure 4. It is desired to obtain a s
from this circuit's outputs such that there is minimum skew betw
to minimize $z^* = \max(\delta_h, \delta_l)$, as shown in figure 5. Again, this skew has to b
over the process variations. This optimization has to be done u
sizing scheme. The absolute delay through this circuit is not a
consumption as this is a custom circuit block used only sparingl
optimization is essentially unconstrained. To test the scope of
and power supply variations were also considered. The model wa
of transistor sizes, process, temperature and power supply vari
example, process variations were considered by simulating each
process corners and the nominal process. The problem is formaliz

$$\text{Find } x^* = \text{Arg} \min_{x \in A} \max_{V} \delta^*(x) \tag{12}$$

Here, A is the hypercube formed by restricting the widths P1-
12μm, the temperature variation between 25-85 and V_{dd} between 4.75-5.25 V. As befo
the widths of N1-N6 are constrained to be one-half the widths of
represents the process variations considered. The optimization
the worst delay skew over all *external noise* factors, i.e., temperat
variations, has to be minimized over the *internal noise* factors, i.e.
For this problem, the sizing provided by the resident circuit des
skew of 290ps (row 1 of table 2). For optimization purposes, th
using $k = 100$ simulations, selected randomly using Latin Hypercub
the effect of temperature and supply variation was factored out.
points were sampled in the 6-dimensional space (w) of the transistor After each

these 1000 points, the *model* in equations (17) and (19) (see Append
different combinations of the supply voltage and temperature vari
of the probability $P(y_{ok})$ (equation 6) over these 9 combinations was found
1000 points. This value was used to estimate the likelihood of a s

$$x_w^* = \text{Arg}\min_{x_w \in A_w} \max_E P_{x_w}(y_{ok}) \qquad (13)$$

was the target for further simulation. Here x_w is the transistor width vector, A_w is the space
of permissible transistor widths and E represents the temperatu
Again, y_{ok} was chosen to be 0.0 which is the minimum possible value o
the 1000 sizing schemes evaluated, the 10 sizing schemes (instea
by equation 13) with the largest probability were chosen for fu
schemes were verified using the 5 process parameters and the 4 corn
temperature fluctuations. The smallest worst-case skew among thes
a significant improvement over the expert's design (row 2 of table
time was 640 cpu seconds on a DECstation 5000 and the overhead of
optimization was less than 10 cpu seconds.

These examples illustrate the power of this approach in eval
schemes efficiently for different optimization problems. The sto
capture the relationship of any performance measure to the tran
accuracy is reflected in the few subsequent evaluations required t

5 Discussion

When optimizing high performance combinational sub blocks, the
knowledge about the objective function, e.g., the difference betw
imum delay will always be greater than zero and less than some upp
through the block. Also, the objective function is expected to b
sides this, little can be said. Gradients are very hard to obtai
difficult. Each objective function evaluation is going to be expe
made through a full circuit simulation, especially for reasonab

of input parameters is usually fairly small, in the 5-20 range, co
transistors that can be sized independently. There might exist so
input parameters. The algorithms based on stochastic modeling fit
main idea behind these algorithms is to maximize the chances of
function after each evaluation. Almost any kind of a priori infor
up the search for optimal sizing. For example, it's quite easy t
performed by this algorithm to the part of the design space that is
to be most promising. The examples given in the previous section
capable of optimizing complex sizing objectives, based only on s
extremely attractive feature of this algorithm is its flexibilit
ation is guaranteed to improve the best solution found thus far,
investigation at any time when the attained solution is deemed to
this can also be perceived as a drawback of the algorithm since th
gence criterion, i.e. location of a stationary point. The only wa
successive iterations return the same solution. Another limitat
cost of model estimation grows rapidly with the number of data po
of the model equations (17 and 19) requires the inversion of a $n\times$
many model evaluations are made on the same data set (as in step 2 c
the matrix inversion needs to be done only once. This method sho
problems of dimensionality up to 25-30. Beyond that, the simple
2 is not sufficient to guarantee close to optimal solutions.

6 Conclusion and Future Work

We have demonstrated how an algorithm based on stochastic modeli
ing some very difficult transistor sizing problems. This algorith
of a search strategy for the optimal sizing scheme. The only inf
algorithm is the value of the objective function which can be cal
simulation. No gradient information is required. The main thrus
mize the probability of improving the best known solution with ea

objective function. This is consistent with the aim of obtaining
uated through circuit simulations, while minimizing simulation
is found at each step of the algorithm and so the search can be sto
solution found thus far is considered satisfactory. The example
solutions can be quickly found using this approach.

Further work is needed on the program interfaces. Process var
only using the process corner models. If more detailed models o
given, the procedure needs to be modified. For example, if the pr
specified by a normal distribution, then the stochastic model ca
process parameters as variables also. Further investigation is
analysis in this scenario.

7 Acknowledgements

The authors thank Tom Gray of IBM and Toby Schaffer of NCSU for hel
examples.

References

[1] M. C. Bernardo, R. Buck, L. Liu, W. A. Nazaret, J. Sacks, and W.
circuit design optimization using a sequential strategy. *IEEE*
Aided Design, CAD-11:pp. 353–360, March 1992.

[2] D. Fan, C. T. Gray, W. Farlow, T. Hughes, W. Liu, and R. Cavin
adder using wave pipelining. In *Proc. of MIT/Brown Conference on Advanced*
in VLSI and Parallel Systems, pages pp. 147–164, 1992.

[3] J. P. Fishburn and A. E. Dunlop. TILOS: a posynomial progra
transistor sizing. In *Proc. Int. Conf. Computer Aided Design*, pages pp. 3
1985.

[4] A. Groch, S. W. Director, and L. M. Vidigal. A new global opt
electronic circuit design. *IEEE Transactions on Circuits and Systems*, CAS-
1985.

[5] B. Hoppe, G. Neuendorf, D. Schmitt-Landsiedel, and W. Specks
speed CMOS logic circuits with analytical models for signal d
namic power dissipation. *IEEE Transaction on Computer-Aided Design*, CAD-9
236–247, 1990.

[6] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization
Science, 220(4598):pp. 671–680, 1983.

[7] Jin-Qin Lu and Takehiko Adachi. A parameter optimization m
circuit design using stochastic model function. *Electronics and*
Japan, 75(4):13–25, 1992.

[8] Mark D. Matson and Lance G. Glasser. Macromodeling and optimi
VLSI circuits. *IEEE Transactions on Computer Aided Design*, CAD-5(4):pp. 6
Oct 1986.

[9] M. D. McKay, R. J. Beckman, and W. J. Conover. A comparison o
for selecting values of input variables in the analysis of out
Technometrics, 21(2):239–245, May 1979.

[10] J. B. Mockus. *Bayesian approach to global optimization: Theory and applicatio*
Kluwer Academic Publishers, 1989.

[11] W. Van Noije, C. T. Gray, W. Liu, T. Hughes, and R. Cavin. C
1GBits/s bandwidth with 25ps resolution. In *Proceedings of the Cus*
Circuits Conference, pages 27.5.1–27.5.4, 1993.

[12] W. Nye, D. Riley, A. Sangiovanni-Vincentelli, and A. Tits.
Optimization-Based System for the Design of Integrated Circ
Computer-Aided Design, CAD-7(4):pp. 501–519, April 1988.

[13] Emil S. Ochotta, Rob A. Rutenbar, and L. Richard Carley. Eq
of high-performance analog circuits. In *Proc. of MIT/Brown Conferen*
Research in VLSI and Parallel Systems, pages 129–143, 1992.

[14] A. O'Hagan. Curve fitting and optimal design for predictio
Statisitcal Society, 40(1): pp. 1–42, 1978.

[15] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design an
experiments. *Statistical Science*, 4(4): pp. 409–435, 1989.

[16] S. Simovich, S. Mehrotra, P. Franzon, and M. Steer. Delay an
modeling for signal integrity management of PCBs and MCMs. *IEE*
Components, Packaging and Manufacturing Technology, to appear in, 1993.

[17] Aimo Torn and Antanas Zilinskas. *Global Optimization*. Springer-V

[18] D. Wong, G. De Micheli, and M. Flynn. Designing high-perfo
using wave pipelining. In *VLSI'89*, pages pp. 241–252, 1989.

[19] A. Zilinskas. Axiomatic approach to statistical models an
optimization theory. *Mathematical Programming*, 22: pp. 104–116, 1982.

[20] A. Zilinskas. Axiomatic characterization of a global optimi
tigation of its search strategy. *Operations Research Letters*, 4(1): pp

Appendix

The model used for the problem above was:

$$Y(x) = \beta_0 + \sum_{j=1}^{d} \beta_j x_j + \phi(x) \qquad (14)$$

where x is the d dimensional parameter vector. $\beta_0, \ldots, \beta_d$ is a $(d+1) \times 1$ vector of unknown coefficients. $\phi(.)$ is a random process with mean zero and co

$$V(w, x) = \sigma^2 \Pi_{i=1}^{d} exp(-\theta | w_j - x_j |) \qquad (15)$$

Suppose that n values of $Y(x)$ are given at sample points $s_1, \ldots, s_n$. These values are in the $n \times 1$ vector ζ_S. F is the $n \times (d+1)$ matrix of the n parameter vectors a
sites augmented by a unit vector, i.e.

$$F = \begin{bmatrix} 1 & s_{11} & \ldots & s_{1d} \\ \vdots & & & \\ 1 & s_{n1} & \ldots & s_{nd} \end{bmatrix} \qquad (16)$$

Then the Best Linear Unbiased Predictor $\hat{y}$ of $Y(.)$ is given as:

$$\hat{y} = X\hat{\beta} + r'(x) R^{-1} (\zeta_S - F\hat{\beta}) \qquad (17)$$

where

$$\hat{\beta} = (F'R^{-1}F)^{-1} F'R^{-1} \zeta_S \qquad (18)$$

where, as before, R is the $n \times n$ covariance matrix of the stochasti
sample locations, and r is the $n \times 1$ vector of covariances $V(x, s_k)$, $k = 1, \ldots, n$, and $X =$
$[1 \; x_1 \ldots x_d]$. The Mean Squared Error (MSE) is given as:

$$\text{MSE}(\hat{y}(x)) = \sigma^2 \left([X \; r'(x)] \begin{bmatrix} 0 & F' \\ F & R \end{bmatrix}^{-1} \begin{bmatrix} X' \\ r(x) \end{bmatrix} \right) \qquad (19)$$

	M1	M2	M3	vbias	delay	skew	
Best Random Sizi	n6g.15e-	80.679e-	80.655e-	0.698	9.9-1	07.7e-	10
Closest Feasible	S16z10ng	68.4e-	68.4e-	6.98	1.0e-	079 7e-	10
Best Sizing after Op	t71.m2iez-a	618.o4ne-	69.6e-	60.0	8.8e-	150 4e-	10

Table 1: Results for Delay Controlled Element

	M1	M2	M3	M4	M5	M6	skew	
Designer's	c9h6oiec-	46.8e	-46.8e	-96.6e	-46.8e	-46.8e	-26.9e-	10
Optimal Po	i7n.t2e	-36.6e	-86.4e	-96.6e	-96.6e	-46.8e	-16.1e-	10

Table 2: Results for Clock Driver Circuit

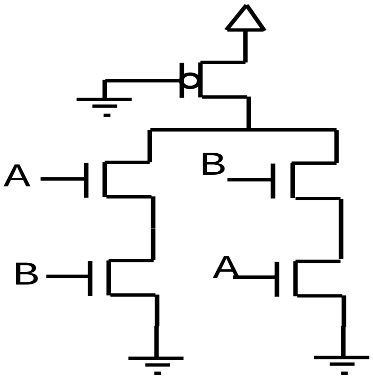
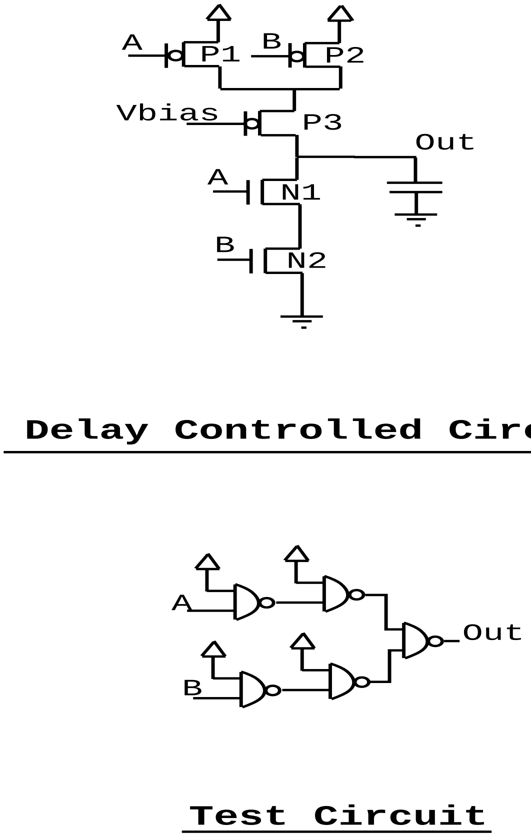


Figure 1: Cross coupled NAND gate



Delay Controlled Circuit Element

Test Circuit

Figure 2: Circuit Block for Wave-pipelining

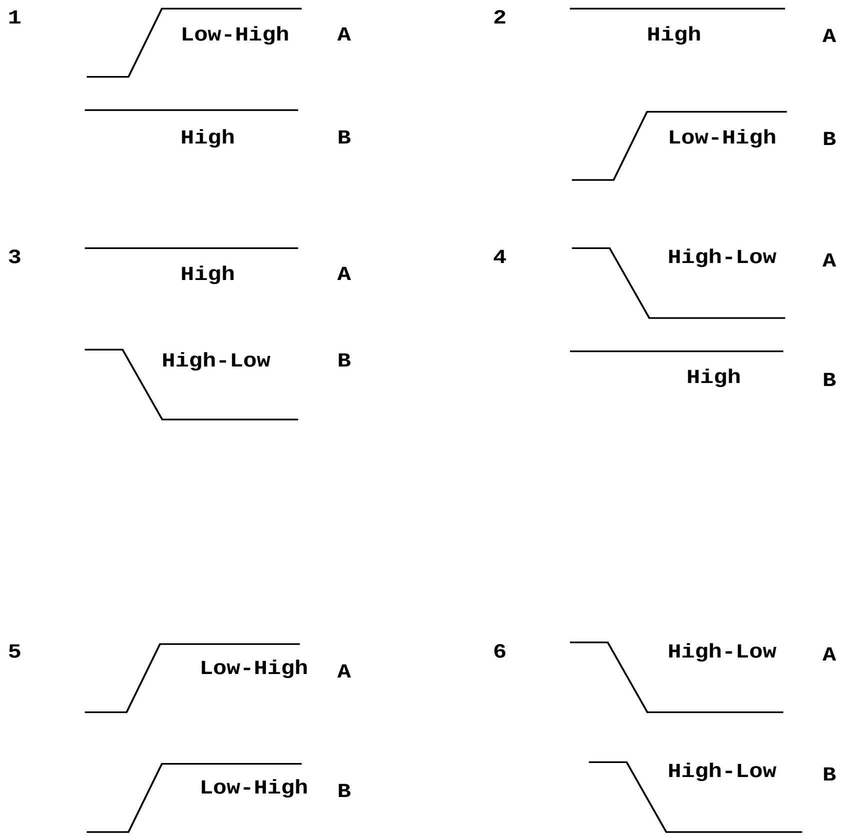


Figure 3: Possible input transitions

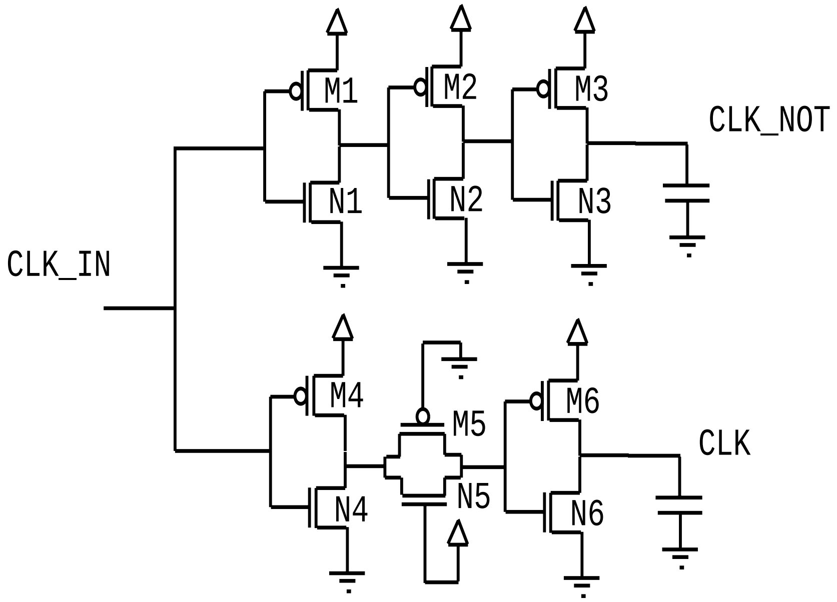


Figure 4: Clock Driver Circuit

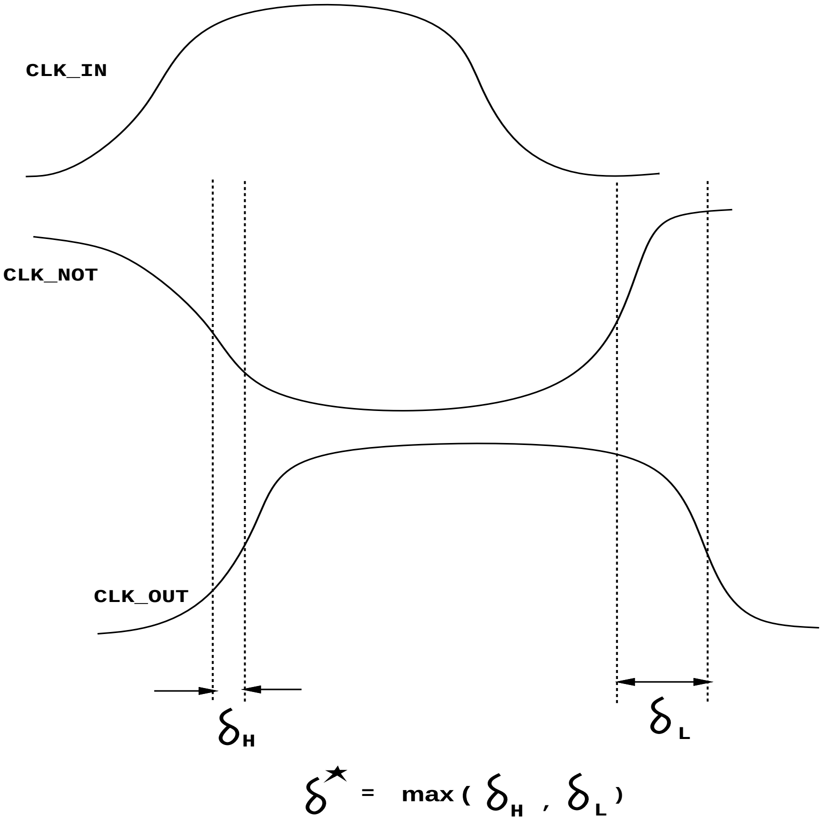


Figure 5: Skew Definition