Gate Sizing Using a Statistical Delay Model

E.T.A.F. Jacobs M.R.C.M. Berkelaar

Design Automation Section, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract

This paper is about gate sizing under a statistical delay model. It shows we can solve the gate sizing problem exactly for a given statistical delay model. The formulation used allows many different forms of objective functions, which could for example directly optimize the delay uncertainty at the circuit outputs. We formulate the gate sizing problem as a nonlinear programming problem, and show that if we do this carefully, we can solve these problems exactly for circuits up to a few thousand gates using the publicly available large scale nonlinear programming solver LANCELOT.

1. Introduction

This paper is about gate sizing under a statistical delay model. Gate sizing refers to the process of optimally assigning drive strengths to the individual gates of a circuit for a given cost function and constraints. For example, one might want to meet a certain delay value for minimum area and/or power penalty.

Gate sizing (as any kind of delay optimization) suffers from the problem that the delay model used might not accurately reflect the delays that will occur later on the chip. First of all, not all details of the final layout might be known yet, giving rise to uncertainty in wire delays. Also, gate sizing is usually based upon a static delay model, which assumes that a gate will always have the same delay, regardless of for example the boolean values on its input pins. In practice this assumption is not true as well. To try to alleviate these kinds of problems, statistical delay models have been introduced, which allow one to express the amount of uncertainty in delay values of gates and wires. In this paper we show that we can perform gate sizing under such a statistical model for gate and/or wire delays. This is to our knowledge the first time a statistical delay model is used in gate sizing.

Statistical delay analysis is basically a static delay analysis where each delay–inducing element (gate or wire) has an associated delay probability function. This function expresses delay uncertainty: either because not all details of the layout are yet known, or to express the fact that delay in a logic circuit is basically dynamic, it depends on things like state and local temperature of a gate, or cross talk in case of a wire. See [1] and [2] for a more detailed analysis.

The statistical treatment of delay uncertainty can replace the traditional best case / typical / worst case delay analysis, which is known to give very pessimistic estimates in many cases. Real statistical calculations can derive the delay probability functions at the circuit outputs as a function of the individual delay probabilities of the delay inducing elements, and of the circuit structure. As has been very clearly shown in [1] and [2], especially the effect of the circuit structure on statistical calculations will result in the fact that the uncertainty in the delay of the total circuit is often much smaller than the uncertainty in the delay of the individual delay elements. We can only deduce this when we do a real statistical delay analysis.

In the recent past, a few attempts have been made to perform static timing analysis in a statistical way. The first attempt known to the authors is described in [6] and [7]. Unfortunately, the details of how the statistical calculations were performed are not revealed in the papers. A more recent attempt is described in [9], where the statistical properties were obtained with Monte Carlo simulations. Monte Carlo simulations can however take a long time to complete and are therefore not practical in an environment directed at optimization, in which repeated delay evaluations are required.

The statistical delay analysis used in this paper is based on [1] and [2]. In these papers, the mean and standard deviation of the distribution resulting from applying the maximum operator on two normal distributions is obtained by sampling. This does not allow for a formal model of gate sizing, so in this paper we derive and use analytical expressions for them. Both mean and standard deviation of the distribution resulting from applying the maximum operator can be expressed as a function of only the means and standard deviations of the operandi. As we will see this analytical expression is key to enabling the use of the statistical delay method for gate sizing. An analytical expression enables us to derive analytical first and second order derivatives of the objective function and the constraints to the variables (among which are the drive strengths) of the gate sizing problem. This again makes it possible to solve the large scale nonlinear minimization problem, that gate sizing under a statistical delay model is, efficiently. We will solve the gate sizing problem using LANCELOT [5].

This paper is organized as follows: First we take a look at the delay model in section 2. Then we examine the theory of statistical calculations in section 3 and present the analytical result for the stochastic maximum–operation. Section 4 presents our gate sizing model including our new statistical delay model and formulations. A small example depicting the formulation of a gate sizing problem is given in section 5. We present results on circuits of up to a few thousand gates using both traditional as well as novel gate sizing objective functions and constraints in section 6. In section 6 we also present gate sizing results on a small tree–circuit in order to get a feeling for the meaning of and response to different optimization objectives. Finally, we draw conclusions and discuss future work in section 7.

2. Delay model

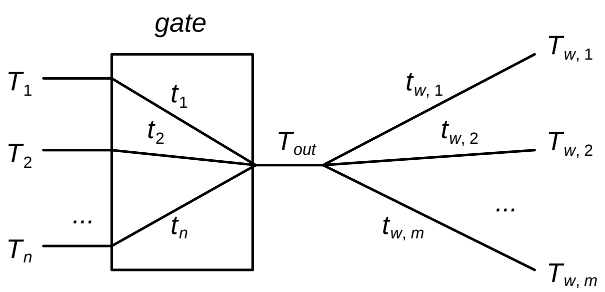

Figure 1. General Delay Model

We assume a delay model with both gate and wire delays. See figure 1 for illustration. Different delays from each input to each output are allowed, as well as different rise and fall times. We can calculate T_{out} and $T_{w,i}$ with:

$$T_{out} = \max_{i=1}^{n}(T_i + t_i) \tag{1}$$

$$T_{w,i} = T_{out} + t_{w,i} \tag{2}$$

Though the statistical delay model allows different delay times from the output of the gate to the inputs of the fanout gates [1] we will limit ourselves, for the purpose of clarity, to one statistical delay for the gate delay. However, we are able to deal with different gate delays. Our model of a sizable gate, taken from [3], which is used to relate the speed factors and the cell propagation delay (see also section 4), does not yet take into account different delays for several wiring segments. We therefore do not differentiate between different wire delays and assume one capacitance due to wiring.

3. Theory of statistical calculations

To model the basic uncertainty in the real delay value, we model the schedule time of a signal as a stochastic variable T, which we assume normally distributed with a mean μ_T and a standard deviation σ_T. We will also model the delay of a gate as a stochastic variable t, with mean μ_t and standard deviation σ_t. In [1] it is shown that, as long as we have a valid mean and standard deviation, the actual shape of the distribution for the delay elements is almost irrelevant if we are only interested in the total circuit delay distribution.

Traditional delay calculation would calculate the delay at the output of a two–input gate by:

$$T_{out} = \max(T_1, T_2) + t \tag{3}$$

This calculation involves two operations: a maximum–operation and an addition. For two statistically independent normally distributed stochastic variables A and B we can calculate the stochastic variable C, which is the addition of A and B, by:

$$\mu_C = \mu_A + \mu_B \tag{4}$$
$$\sigma_C^2 = \sigma_A^2 + \sigma_B^2$$

We also have to perform the maximum–operation with stochastic values. To see what happens in this case, we will concentrate on calculating $C = \max(A, B)$, with A, B and C stochastic variables. A and B are normally distributed with means μ_A and μ_B and standard deviations σ_A and σ_B. What is the distribution of C ? For any value x we can write:

$$P(C \le x) = P(A \le x) \cap P(B \le x) \tag{5}$$

Assuming statistical independence of A and B, we can write equation 5 as:

$$P(C \le x) = P(A \le x) \cdot P(B \le x) \tag{6}$$

Assuming statistical independence is an approximation in case of reconverging paths in the circuit. [2] shows this approximation only gives very small errors. We will use the notation F_A for the distribution function of the stochastic variable A. We will also use the notation:

$$P(A \le x) = F_A(x) = \int_{-\infty}^{x} f_A(u)du \tag{7}$$

in which the probability density function $f_A(x)$ for a normal distribution is given by:

$$f_A(x) = \frac{1}{\sigma_A\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_A}{\sigma_A}\right)^2} \tag{8}$$

Note that in case of a normal distribution no closed form for F_A exists. Using this notation, and taking the derivative left and right of equation 6, we get:

$$f_C(x) = f_A(x)F_B(x) + F_A(x)f_B(x) \tag{9}$$

We refer the reader to [1] and [2] where it is shown that the resulting probability density function $f_C(x)$ for stochastic variable C is very similar to, but not necessarily equal to, a normal distribution. We judge that the resulting probability density function approximates the normal distribution close enough for our purposes. Given any probability density function we can always calculate the mean and standard deviation [10]. In case of the normal distribution the probability density function is completely characterized by the mean and standard deviation. We now give μ_C which is a function of just μ_A, μ_B, σ_A and σ_B:

$$\mu_C = Ex_C = \frac{\sqrt{\sigma_A^2 + \sigma_B^2}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu_A - \mu_B}{\sqrt{\sigma_A^2 + \sigma_B^2}}\right)^2} + \tag{10}$$

$$\mu_A\phi\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)+\mu_B\phi\left(\frac{\mu_B-\mu_A}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)$$

in which $\phi(x)$ is given by:

$$\phi(x) = \int_{-\infty}^{x} e^{-\frac{1}{2}u^2}du \tag{11}$$

We also give Ex_C^2:

$$Ex_C^2 = (\mu_A+\mu_B)\frac{\sqrt{\sigma_A^2+\sigma_B^2}}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)^2} + \tag{12}$$

$$(\sigma_A^2+\mu_A^2)\phi\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)+$$

$$(\sigma_B^2+\mu_B^2)\phi\left(\frac{\mu_B-\mu_A}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)$$

We can now calculate σ_C by:

$$\sigma_C^2 = Ex_C^2-\mu_C^2 \tag{13}$$

We have now expressed μ_C and σ_C as functions of just μ_A, μ_B, σ_A and σ_B. Appendix A gives the derivation of equations 10 and 12. This analytical expression for μ_C and σ_C is an extension to [1] and [2], which is required to be able to perform gate sizing under a statistical delay model efficiently.

4. Gate sizing

We build upon the sizable model of a gate, which was introduced in [3]. The propagation delay of a gate as a function of its speed factor S_{cell} is given there by:

$$t_{cell} = t_{int} + c\cdot\frac{C_{load}+\sum_i C_{in,i}S_i}{S_{cell}} \tag{14}$$

In this equation, t_{int} is a constant denoting the delay due to capacitances internal to the gate, C_{load} is a constant denoting the capacitance loading the gate (mainly in wiring), C_{in} is a constant denoting the gate oxide capacitance of transistors driven by the gate and S_i is the speed (sizing) factor of those gates. The constant c relates propagation delay to capacitance. The internal delay t_{int} does not change while sizing, because the decrease in resistance is counteracted by the increase in internal capacitance. The remaining capacitances of wiring and successor gates, however, do not change due to the sizing of this gate. The speed factor S_{cell} can vary from 1, meaning no speed–up, to limit times speed–up.

In order to have fewer nonlinear terms to deal with, we multiply equation 14 by S_{cell}. This gives us:

$$t_{cell}S_{cell} = t_{int}S_{cell} + c\left(C_{load}+\sum_i C_{in,i}S_i\right) \tag{15}$$

In the gate sizing approach using the statistical delay model we take the mean of the gate delay μ_t equal to the delay t_{cell} in equation 15. We also want to change the gate's standard deviation while sizing. Therefore we define the standard deviation as:

$$\sigma_t = f(t_{cell}) \tag{16}$$

Now that we have discussed the equations relating delay to sizing, we need to discuss calculating the circuit delay. In our statistical approach we need to calculate the maximum arrival time at the inputs of a gate (see also section 2 equation 1). For each gate we calculate the mean and standard deviation of the maximum of the circuit delays at the inputs of the gate using equations 10, 12 and 13. Now we still have to take into account the gate delay. We add this gate delay (see also section 2 equation 2) to the calculated maximum circuit delay at the inputs using equation 4. We also calculate the delay distribution of the total circuit by taking the stochastic maximum over all the primary outputs of the circuit.

The total gate sizing formulation for minimal delay is given in equation 17. Note that also different objective functions are possible. In equation 17 $\max_\mu$ and $\max_\sigma$ denote functions calculating the mean and standard deviation of the maximum of a number of normal distributed stochastic variables. Note that the constraints in equation 17 are either all equality constraints or simple constraints on the range of individual variables, and that, while some are linear, others are highly nonlinear. We will solve the constrained nonlinear programming formulation of the form described in equation 17 using the large scale nonlinear programming package LANCELOT [5].

$$\text{minimize} \quad \mu_{T_{max}} \tag{17}$$

$$\mu_{T_{max}} = \max_\mu(\mu_{T_{o,1}}\sigma_{T_{o,1}}, \ldots, \mu_{T_{o,n}}, \sigma_{T_{o,n}})$$

$$\sigma_{T_{max}} = \max_\sigma(\mu_{T_{o,1}}\sigma_{T_{o,1}}, \ldots, \mu_{T_{o,n}}, \sigma_{T_{o,n}})$$

and for each gate:

$$\mu_{t_{cell}}S_{cell} = t_{int}S_{cell} + c\left(C_{load}+\sum_i C_{in,i}S_i\right)$$

$$\mu_{T_o} = \mu_{U_{max}} + \mu_{t_{cell}}$$

$$\sigma_{T_o}^2 = \sigma_{U_{max}}^2 + \sigma_{t_{cell}}^2$$

$$\mu_{U_{max}} = \max_\mu(\mu_{T_{i,1}}\sigma_{T_{i,1}}, \ldots, \mu_{T_{i,n}}, \sigma_{T_{i,n}})$$

$$\sigma_{U_{max}} = \max_\sigma(\mu_{T_{i,1}}\sigma_{T_{i,1}}, \ldots, \mu_{T_{i,n}}, \sigma_{T_{i,n}})$$

$$\sigma_{t_{cell}}^2 = (f(\mu_{t_{cell}}))^2$$

$$1 \le S_{cell} \le \text{limit}$$

In order for LANCELOT to solve a nonlinear gate sizing formulation as in equation 17 we have to calculate the first and second order derivatives of all terms in the problem to every problem variable. Only when first and second order derivative information is available will LANCELOT be able to deal with highly nonlinear problems efficiently. We find it advantageous to have as many linear terms, as opposed to nonlinear terms, as possible in each constraint, because this increases the efficiency of LANCELOT. This is the reason behind the reformulation of equation 14 to equation 15. We also use only the squared version of standard deviations in the model. Therefore we introduce a new variable which is equal to the square of the standard deviation. The need to express explicitly in the gate sizing formulation the relation between mean and standard deviation of the result of the maximum–operator and the means and standard deviations of the operandi as a function, as well as to calculate exactly the first and second order derivatives of this function are the reasons behind the exercise of expressing the mean and standard deviation of the maximum–operation as functions of the means and standard deviations of its operandi. These analytical functions enable us to calculate analytical derivatives.

We can also choose other objective functions and add additional constraints to the gate sizing formulations. We can choose a weighted sum of sizing factors in the objective function. This can model area, or, if we take into account capacitances and switching activity under zero delay model in the weights, power. Apart from the change in switching activity due to the change in timing as a result of gate sizing, both area and power scale linear with the sizing factor. This has been shown in [3] and [8].

We can also add delay constraints: on the mean circuit propagation delay or on the mean plus one or several times the standard deviation. Adding the standard deviation in the circuit propagation delay constraint ensures that a larger percentage of the circuits will conform to the delay constraint. In case of just $\mu_{T_{max}}$ 50% of the circuits will conform, in case of $\mu_{T_{max}} + \sigma_{T_{max}}$ 84.1% will conform and in case of $\mu_{T_{max}} + 3\sigma_{T_{max}}$ 99.8% will conform to the delay constraint set.

5. Example

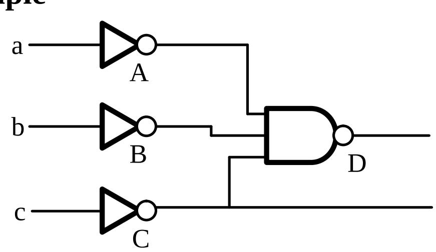

Figure 2. Statistical delay model gate sizing example

We will now look at the example of figure 2. The gates in this network are denoted by the capital letters A to D and the primary inputs by the small letters a to c. We give the corresponding gate sizing formulation for minimal $\mu_{T_{max}} + 3\sigma_{T_{max}}$ delay using the statistical delay method in equation 18. This means that we calculate the sizing factors for each gate in the circuit such that 99.8% of the circuits have the minimum possible propagation delay. The maximum over all outputs is taken in 18a. Note that we can only calculate the statistical maximum for two operandi directly at the same time. For the multiple inputs of gate D, we use the two–operand maximum operation repeatedly (18b). The gate delay is added to the maximum over the inputs in 18c. The equations relating gate size and delay are in 18d. For the function relating mean and standard deviation of a gate we assume the standard deviation to be a quarter of the mean in this example (18e). We assume a maximum speed–up of 3 for each gate (18f).

$$\text{minimize} \quad \mu_{T_{max}} + 3\sigma_{T_{max}} \tag{18}$$

$$\mu_{T_{max}} = \max_{\mu}(\mu_{T_C}, \sigma_{T_C}, \mu_{T_D}, \sigma_{T_D}) \tag{18a}$$

$$\sigma_{T_{max}} = \max_{\sigma}(\mu_{T_C}, \sigma_{T_C}, \mu_{T_D}, \sigma_{T_D})$$

$$\mu_{U_D} = \max_{\mu}(\max_{\mu}(\mu_{T_B}, \sigma_{T_B}, \mu_{T_C}, \sigma_{T_C}), \max_{\sigma}(\mu_{T_B}, \sigma_{T_B}, \mu_{T_C}, \sigma_{T_C}), \mu_{T_A}, \sigma_{T_A}) \tag{18b}$$

$$\sigma_{U_D} = \max_{\sigma}(\max_{\mu}(\mu_{T_B}, \sigma_{T_B}, \mu_{T_C}, \sigma_{T_C}), \max_{\sigma}(\mu_{T_B}, \sigma_{T_B}, \mu_{T_C}, \sigma_{T_C}), \mu_{T_A}, \sigma_{T_A})$$

$$\mu_{T_D} = \mu_{U_D} + \mu_{t_D} \tag{18c}$$

$$\sigma^2_{T_D} = \sigma^2_{U_D} + \sigma^2_{t_D}$$

$$\mu_{t_A} S_A = S_A t_{int_A} + c \cdot \left(C_{load_A} + C_{in} S_D\right) \tag{18d}$$

$$\mu_{t_B} S_B = S_B t_{int_B} + c \cdot \left(C_{load_B} + C_{in} S_D\right)$$

$$\mu_{t_C} S_C = S_C t_{int_C} + c \cdot \left(C_{load_C} + C_{in} S_D\right)$$

$$\mu_{t_D} S_D = S_D t_{int_D} + c \cdot C_{load_D}$$

$$\sigma_{t_A} = 0.25\mu_{t_A} \tag{18e}$$

$$\sigma_{t_B} = 0.25\mu_{t_B}$$

$$\sigma_{t_C} = 0.25\mu_{t_C}$$

$$\sigma_{t_D} = 0.25\mu_{t_D}$$

$$1 \le S_A \le 3 \tag{18f}$$

$$1 \le S_B \le 3$$

$$1 \le S_C \le 3$$

$$1 \le S_D \le 3$$

6. Results

We have done two sets of experiments. The first set of experiments is done to show both the applicability of the statistical delay gate sizing method to circuits of up to a few thousand gates, as well as the additional objectives and constraints we can formulate using our statistical gate sizing method. As can be seen in table 1 the method is able to deal with circuits of up to a few thousand gates. We have done several experiments with the three circuits. The first two entries for each circuit give the range in which the mean propagation delay and sum of speed factors (the measure of area used in our experiments) can vary. The next two entries for each circuit show the results of minimizing the mean propagation delay plus one time the standard deviation and the mean propagation delay plus three times the standard deviation. The last three entries in the table for each of the three circuits minimize the area (sum of speed factors) subject to constraints on the mean propagation delay, the mean propagation delay plus one time and plus three times the standard deviation.

Table 1. Results of statistical sizing for some large benchmark circuits

name	#cells	minimize	constraint	μ_{Tmax}	σ_{Tmax}	ΣS_i	CPU
apex1	982	ΣS_i		173.72	5.867	982	
		μ_{Tmax}		73.21	2.099	1989	41 m 13.5 s
		$\mu_{Tmax}+\sigma_{Tmax}$		73.26	1.972	1949	41 m 10.8 s
		$\mu_{Tmax}+3\sigma_{Tmax}$		73.57	1.701	1843	67 m 54.8 s
		ΣS_i	$\mu_{Tmax}\leq 120$	120.00	2.950	998	67 m 49.9 s
		ΣS_i	$\mu_{Tmax}+\sigma_{Tmax}\leq 120$	117.16	2.842	1001	103 m 19.4 s
		ΣS_i	$\mu_{Tmax}+3\sigma_{Tmax}\leq 120$	112.07	2.645	1007	85 m 43.1 s
apex2	117	ΣS_i		31.50	1.784	117	
		μ_{Tmax}		23.45	1.419	304	18.5 s
		$\mu_{Tmax}+\sigma_{Tmax}$		23.48	1.373	294	10 m 16.5 s
		$\mu_{Tmax}+3\sigma_{Tmax}$		23.79	1.202	279	52.2 s
		ΣS_i	$\mu_{Tmax}\leq 29$	29.00	1.488	123	42.1 s
		ΣS_i	$\mu_{Tmax}+\sigma_{Tmax}\leq 29$	27.64	1.365	131	7.0 s
		ΣS_i	$\mu_{Tmax}+3\sigma_{Tmax}\leq 29$	25.47	1.176	154	38.3 s
k2	1692	ΣS_i		183.98	3.281	1692	
		μ_{Tmax}		75.00	1.293	3750	54 m 26.1 s
		$\mu_{Tmax}+\sigma_{Tmax}$		75.02	1.228	3690	50 m 45.7 s
		$\mu_{Tmax}+3\sigma_{Tmax}$		75.23	1.120	3596	83 m 20.7 s
		ΣS_i	$\mu_{Tmax}\leq 120$	120.00	1.829	1794	221 m 32.0 s
		ΣS_i	$\mu_{Tmax}+\sigma_{Tmax}\leq 120$	118.27	1.744	1801	214 m 38.4 s
		ΣS_i	$\mu_{Tmax}+3\sigma_{Tmax}\leq 120$	115.10	1.637	1814	157 m 50.6 s

Gate sizing combined with the statistical delay calculation method thus enables us to size circuits, in order to get a certain confidence in the circuit realizing its timing constraints, or to optimize the number of circuits that will operate at the required clock frequency given the uncertainties in its propagation delay. All experiments are performed on a Hewlett Packard K260. The CPU–times reported are for solving the various gate sizing formulations under a statistical delay model using the large scale nonlinear optimization package LANCELOT [5].

The second set of experiments is on the tree–circuits of figure 3, which contains seven NAND–gates. These experiments are to show how different constraints and objective functions effect the speed factors for this simple circuit.

The first two entries of table 2 denote the range in which the area and mean propagation delay of the circuit can vary. We have selected three values of the mean propagation delay in this range. One is chosen in the middle and the other two nearer the extremes of the range. Table 2 shows that there is a margin to change the standard deviation given a fixed mean propagation delay, and that the interval is largest for the middle choice in the range of mean propagation delays. It is also clear from table 2 that minimal standard deviation given a fixed mean propagation delay leads to a higher area usage than just minimizing area given the mean propagation delay.

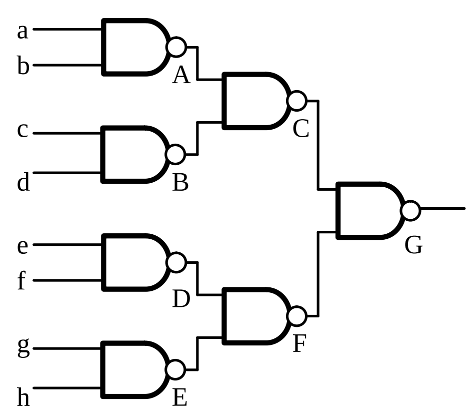

Figure 3. Tree circuit

Table 2. Results for tree–circuit

objective	constraint	μ_{Tmax}	σ_{Tmax}	ΣS_i
min ΣS_i		7.4	0.811	7.00
min μ_{Tmax}		5.4	0.592	21.00
min ΣS_i	μ_{Tmax}=5.8	5.8	0.631	14.73
min σ_{Tmax}	μ_{Tmax}=5.8	5.8	0.622	15.66
max σ_{Tmax}	μ_{Tmax}=5.8	5.8	0.667	19.22
min ΣS_i	μ_{Tmax}=6.5	6.5	0.704	9.54
min σ_{Tmax}	μ_{Tmax}=6.5	6.5	0.689	10.20
max σ_{Tmax}	μ_{Tmax}=6.5	6.5	0.831	15.51
min ΣS_i	μ_{Tmax}=7.2	7.2	0.786	7.21
min σ_{Tmax}	μ_{Tmax}=7.2	7.2	0.689	7.25
max σ_{Tmax}	μ_{Tmax}=7.2	7.2	0.817	9.08

We now look at the speed factors in table 3 corresponding to the sizing experiments for the tree–circuit for minimal area, minimal and maximal standard deviation. Table 3 shows that both sizing for minimal area (sum of speed factors) and for minimal standard deviation treat similar gates (first group: S_A, S_B, S_D and S_E; second group: S_C and S_F) similarly, and gates towards the output of the circuit get larger speed factors. This behavior is more extreme in case of sizing for minimal standard deviation. The standard deviation for gates nearer the input does not need to be as small as for gates nearer the output, because for a balanced mean delay and similar gates the maximum operator results in a slightly higher mean but considerably smaller standard deviation. Sizing for maximal standard deviation clearly differentiates delays on different paths to maximize the standard deviation, as is to be expected. The last gate S_G is then appropriately sized to achieve the required mean propagation delay.

Table 3. Speed factors for tree–circuit for μ_{Tmax}=6.5

objective	S_A	S_B	S_C	S_D	S_E	S_F	S_G
min ΣS_i	1.22	1.22	1.45	1.22	1.22	1.45	1.74
min σ_{Tmax}	1.00	1.00	2.01	1.00	1.00	2.01	3.00
max σ_{Tmax}	3.00	1.00	1.00	3.00	3.00	3.00	1.51

7. Conclusions and future work

We have presented a gate sizing method under a statistical delay model, which we expressed as a nonlinear programming problem. As an essential step in the modeling of the statistical gate sizing method, we have expressed the mean and standard deviation of the result of the maximum–operator as a function of the means and standard deviations of the operandi. We have solved the gate sizing formulation under a statistical delay problem exactly for problems up to a few thousand gates using the large scale nonlinear programming package LANCELOT. We have presented several improvements to the gate sizing formulation and discussed implementation details crucial to solving the gate sizing problem efficiently. We also presented experiments demonstrating the effect of different objective functions and the result of those objective functions on the speed factors.

Future work will look into dealing with correlations between stochastic variables in the circuit, as a result of reconverging paths, which is currently not included in our delay model, as we assume statistical independence. Another interesting challenge could be to express the mean and standard deviation of the maximum of multiple (more than two) operandi explicitly, rather than as the repeated maximum of two operandi.

References

[1] Berkelaar, M.R.C.M., 'Statistical Delay Calculation', Workshop Notes of the International Workshop on Logic Synthesis, May 18–21, 1997

[2] Berkelaar, M.R.C.M., 'Statistical Delay Calculation, a Linear Time Method', Proceedings of TAU'97, Austin, TX, December 4–5, 1997, pp. 15–24.

[3] Berkelaar, M.R.C.M. and J.A.G. Jess, 'Gate Sizing in MOS Digital Circuits with Linear Programming', Proceedings of the European Design Automation Conference 1990, pp. 217–221.

[4] Brayton, R.K., R. Rudell, A.L. Sangiovanni–Vincentelli and A. Wang, 'MIS: A Multiple–Level Logic Optimization System', IEEE Transactions on Computer–Aided Design of Integrated Circuits and Systems, Nov. 1987, Vol. CAD–6, pp. 1062–1081.

[5] Conn, A.R., N.I.M. Gould and Ph.L. Toint, 'LANCELOT A Fortran Package for Large–Scale Nonlinear Optimization (Release A)', Springer Series in Computational Mathematics, Vol. 17, Springer Verlag (Heidelberg, New York), ISBN 3–540–55470–X, 1992.

[6] Hitchcock, R.B., 'Timing Verification and the Timing Analysis Program', Proceedings of the 19th Design Automation Conference, 1982, pp. 594–604.

[7] Hitchcock, R.B., G.L. Smith, D.D. Cheng, 'Timing Analysis of Computer Hardware', IBM Journal of Research and Development, Vol. 26 No. 1, January 1982, pp. 100–105.

[8] Jacobs, E.T.A.F., 'Using Gate Sizing to Reduce Glitch Power', Proceedings of the Prorisc/IEEE Workshop on Circuits, Systems and Signal Processing, Mierlo, The Netherlands, November 27–28, 1996, pp. 183–188.

[9] Jyu, H.H., 'Statistical Delay Model Based Performance Analysis and Optimization in Logic Circuit Design', Ph.D. Thesis, Princeton University, 1994.

[10] Papoulis, A., 'Probability, Random Variables, and Stochastic Processes', Third Edition, McGraw–Hill New York, ISBN 0–07–048477–5, 1991.

Appendix A

We will now derive the mean and standard deviation of a stochastic variable C which is the maximum of two normal distributed statistically independent stochastic variables A and B. In order to derive this mean and standard deviation we will change the bases of the double integration:

$$\int_{-\infty}^{\infty} x f_A(x) \int_{-\infty}^{x} f_B(y)\,dy\,dx \tag{19}$$

which part of the calculation of $\mu_C = Ex_C$ as follows:

$$\frac{x-\mu_A}{\sigma_A} = \frac{u\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} - \frac{v\sigma_A}{\sqrt{\sigma_A^2+\sigma_B^2}} \tag{20}$$

which gives:

$$x = \frac{u\sigma_A\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} - \frac{v\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}} + \mu_A \tag{21}$$

and:

$$\frac{y-\mu_B}{\sigma_B} = \frac{u\sigma_A}{\sqrt{\sigma_A^2+\sigma_B^2}} + \frac{v\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} \tag{22}$$

which gives:

$$y = \frac{u\sigma_A\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} + \frac{v\sigma_B^2}{\sqrt{\sigma_A^2+\sigma_B^2}} + \mu_B \tag{23}$$

For this change of base we calculate:

$$\left|\frac{\delta(x,y)}{\delta(v,u)}\right| = \begin{vmatrix} \frac{\sigma_A\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} & -\frac{\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}} \\ \frac{\sigma_A\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} & \frac{\sigma_B^2}{\sqrt{\sigma_A^2+\sigma_B^2}} \end{vmatrix} \tag{24}$$

$$= \frac{\sigma_A\sigma_B^3 + \sigma_A^3\sigma_B}{\sigma_A^2+\sigma_B^2} = \sigma_A\sigma_B$$

The mean of the stochastic variable C then becomes:

$$\mu_C = \int_{-\infty}^{\infty} x f_C(x)\,dx \tag{25}$$

$$= \int_{-\infty}^{\infty} x f_A(x) F_B(x)\,dx + \int_{-\infty}^{\infty} x F_A(x) f_B(x)\,dx$$

$$= \frac{1}{\sigma_A\sigma_B\sqrt{2\pi}\sqrt{2\pi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}} \left(\frac{u\sigma_A\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} - \frac{v\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}} + \mu_A \right) e^{-\frac{1}{2}(u^2+v^2)} \sigma_A\sigma_B\,dv\,du + \ldots$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}} \left(\frac{u\sigma_A\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} - \frac{v\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}} + \mu_A \right) e^{-\frac{1}{2}(u^2+v^2)}\,dv\,du + \ldots$$

$$= \left[\frac{1}{\sqrt{2\pi}} \frac{\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}} e^{-\frac{1}{2}v^2} \right]_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}} +$$

$$\frac{\mu_A}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}} e^{-\frac{1}{2}v^2}\,dv + \frac{\mu_B}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\mu_B-\mu_A}{\sqrt{\sigma_A^2+\sigma_B^2}}} e^{-\frac{1}{2}v^2}\,dv +$$

$$\left[\frac{1}{\sqrt{2\pi}} \frac{\sigma_B^2}{\sqrt{\sigma_A^2+\sigma_B^2}} e^{-\frac{1}{2}v^2} \right]_{-\infty}^{\frac{\mu_B-\mu_A}{\sqrt{\sigma_A^2+\sigma_B^2}}}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}} \left(-\frac{v\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}} + \mu_A \right) e^{-\frac{1}{2}v^2}\,dv + \ldots$$

$$= \frac{\sqrt{\sigma_A^2+\sigma_B^2}}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)^2} +$$

$$\mu_A\phi\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right) + \mu_B\phi\left(\frac{\mu_B-\mu_A}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)$$

in which $\phi(x)$ is given by:

$$\phi(x) = \int_{-\infty}^{x} e^{-\frac{1}{2}u^2}\,du \tag{26}$$

Note that in some lines of equation 25 we have only given one half of the equation explicitly. The other half is depicted by triple dots, and is similar to the first half of the equation. We will now calculate the standard deviation of stochastic variable C in two steps. The first step is the calculation of Ex_C^2:

$$Ex_C^2 = \int_{-\infty}^{\infty} x^2 f_C(x)\,dx \tag{27}$$

$$= \int_{-\infty}^{\infty} x^2 f_A(x) F_B(x)\,dx + \int_{-\infty}^{\infty} x^2 F_A(x) f_B(x)\,dx$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}}\left(\frac{u\sigma_A\sigma_B}{\sqrt{\sigma_A^2+\sigma_B^2}} - \frac{v\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}}+\mu_A\right)^2 e^{-\frac{1}{2}(u^2+v^2)}dvdu + \ldots$$

$$= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}}\left(\frac{u^2\sigma_A^2\sigma_B^2}{\sigma_A^2+\sigma_B^2}+\frac{v^2\sigma_A^4}{\sigma_A^2+\sigma_B^2} - \frac{2v\mu_A\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}}+\mu_A^2\right)e^{-\frac{1}{2}(u^2+v^2)}dvdu + \ldots$$

$$= \left[\frac{1}{2\pi}\int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}}\frac{-u\sigma_A^2\sigma_B^2}{\sigma_A^2+\sigma_B^2}e^{-\frac{1}{2}(u^2+v^2)}dv\right]_{-\infty}^{\infty} +$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}}\left(\frac{v^2\sigma_A^4}{\sigma_A^2+\sigma_B^2} - \frac{2v\mu_A\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}}+\mu_A^2\right)e^{-\frac{1}{2}v^2}dv +$$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}}\frac{\sigma_A^2\sigma_B^2}{\sigma_A^2+\sigma_B^2}e^{-\frac{1}{2}(u^2+v^2)}dvdu + \ldots$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}}\left(\frac{v^2\sigma_A^4}{\sigma_A^2+\sigma_B^2} - \frac{2v\mu_A\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}} + \mu_A^2 + \frac{\sigma_A^2\sigma_B^2}{\sigma_A^2+\sigma_B^2}\right)e^{-\frac{1}{2}v^2}dv + \ldots$$

$$= \left[-\frac{1}{\sqrt{2\pi}}\frac{v\sigma_A^4}{\sigma_A^2+\sigma_B^2}e^{-\frac{1}{2}v^2}\right]_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}} + \left[\frac{2\mu_A\sigma_A^2}{\sqrt{\sigma_A^2+\sigma_B^2}}e^{-\frac{1}{2}v^2}\right]_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}} +$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}}\left(\frac{\sigma_A^4+\sigma_A^2\sigma_B^2}{\sigma_A^2+\sigma_B^2}+\mu_A^2\right)e^{-\frac{1}{2}v^2}dv + \ldots$$

$$= (\sigma_A^2+\mu_A^2)\phi\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right) +$$

$$\frac{e^{-\frac{1}{2}\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)^2}}{\sqrt{2\pi}\sqrt{\sigma_A^2+\sigma_B^2}}\left(2\mu_A\sigma_A^2 - \frac{\sigma_A^4(\mu_A-\mu_B)}{\sigma_A^2+\sigma_B^2}\right) + \ldots$$

$$= \frac{e^{-\frac{1}{2}\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)^2}}{\sqrt{2\pi}\sqrt{\sigma_A^2+\sigma_B^2}}\left(\frac{\mu_A\sigma_A^4+2\mu_A\sigma_A^2\sigma_B^2+\mu_B\sigma_A^4}{\sigma_A^2+\sigma_B^2}\right) +$$

$$(\sigma_A^2+\mu_A^2)\phi\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right) + \ldots$$

$$= (\sigma_A^2+\mu_A^2)\phi\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right) +$$

$$(\sigma_B^2+\mu_B^2)\phi\left(\frac{\mu_B-\mu_A}{\sqrt{\sigma_A^2+\sigma_B^2}}\right) +$$

$$(\mu_A+\mu_B)\frac{\sqrt{\sigma_A^2+\sigma_B^2}}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{\mu_A-\mu_B}{\sqrt{\sigma_A^2+\sigma_B^2}}\right)^2}$$

Note that in equation 27 we also have given only one half of the equation explicitly, with the other half which is similar to the first, depicted by triple dots. We can now calculate the standard deviation of stochastic variable C with the following equation:

$$\sigma_C^2 = Ex_C^2 - \mu_C^2 \tag{28}$$

We have now expressed μ_C and σ_C as functions of just μ_A, μ_B, σ_A and σ_B.