Impact of Technology Scaling on Metastability Performance of CMOS Synchronizing Latches

Maryam Shojaei Baghini, Madhav P. Desai
MicroElectronics Group, E.E. Department of IITB, Mumbai-400076, India
maryshojaei@ieee.org, madhav@ee.iitb.ernet.in

Abstract

In this paper, we use circuit simulations to characterize the effects of technology scaling on the metastability parameters of CMOS latches used as synchronizers. We perform this characterization by obtaining a synchronization error probability curve from a histogram of the latch delay. The main metastability parameters of CMOS latches are τ_m and T_w. τ_m is the exponential time constant of the rate of decay of metastability and T_w is effective size of metastability window at a normal propagation delay. Both parameters can be extracted from a histogram of the latch delay. This paper also explains a way to calibrate simulator for enough accuracy. Our simulations indicate that τ_m scales better than the technology scale factor. T_w also scales down but its factor cannot be well estimated as that of τ_m. This is because T_w is a complex function of signal and clock edge rate and logic threshold level.

1. Introduction

The main behavioral characteristic curve of a synchronizing latch is the relationship between the synchronization error probability curve and the latch delay. A general view of this curve is shown in Fig.1. Note that in this figure we have not distinguished the cases when the data is latched or not because our concern is the actual fault of a synchronizer from latch delay point of view. Thus our aim in this paper is to survey the probability that the latch decision takes time larger than a predefined threshold value. The curve of Fig.1 has three main parts. The first region ranges from zero delay to the normal delay of the latch. In this region, if the data is latched, the required setup/hold time of data relative to clock edge has been satisfied and for the case data is not latched the circuit delay is less than or equal to normal delay of latch. The second part is the deterministic part or quasi-metastable region. In this region the latch output transition is delayed from a normal propagation but its delay is determined by the setup time [1]. The third region, which extends to infinity is the region of true

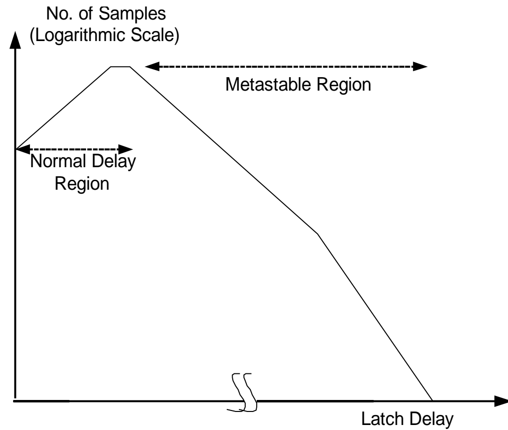


Figure 1. Latch delay histogram

metastability. In this region the delay is much larger than the normal latch delay and it is not determinable. Here, the internal nodes of the latch are balanced within the latch noise range, for example thermal noise. The latch must relay on thermal noise in order to resolve into a stable state. The slope of the curve in the second and the third region enables us to determine τ_m and the intercept point of the vertical axis (when it is scaled by clock/data separation time), determines T_w of the latch. τ_m and T_w determine the mean time between failures (MTBF) of synchronizer as [1]

$$MTBF = \frac{e^{t/\tau_m}}{T_w.f_c.f_d} \qquad (1)$$

We assume that the structure of a latch in the metastable region can be summarized as two cascaded inverters in a positive feedback loop independent of latch architecture. Thus as a rough approximation τ_m is given by [2]

$$\tau_m = (C_Q+2C_F)/(Gm-G_o) \qquad (2)$$

where C_Q is the equivalent capacitance from the output nodes of latch inverters to ground and C_F is the feedback capacitance of each inverter. Gm and Go are transconductance and output conductance of each inverter, respectively. Relation (2) shows the basic parameters from which the impact of scaling on τ_m can be theoretically surveyed. For example a first approximation obviously results to the scaling of τ_m by a factor of larger

than one and less than 1/s, where s is the scaling factor about 0.7V.

Our strategy and considerations for simulation of metastability behavior are explained in the next section. After that the effect of technology scaling on the metastability of two different latch architectures with different sizing schemes will be surveyed.

2. Metastability simulation considerations

To characterize metastabe behavior versus technology scaling we use the basic characteristic curve of a latch's metastability behavior, i.e. probability density function of latch delay versus latch time delay ($\rho_{e(td)}$). This is because this curve implicitly covers all metastability parameters. $\rho_{e(td)}$ is obtained from histogram of sampled points versus latch delay.

Sizing strategies of synchronization circuits differ based on the application. For example in ASIC design, sizing is primarily driven by setup and hold time considerations. On the other hand device size optimization with respect of metastable parameters leads to different aspect ratios [4]. Both sizing schemes are considered in this paper. Fig. 2.a shows schematics for the first considered latch, which is one of the most common used latches in cell libraries. The second latch, shown in Fig. 2.b, is selected based on the configuration and device size optimization concerned with the minimization of metastability resolving time constant [1]. In our simulations, we were careful to include all parasitic effects such as source/drain diodes with appropriate area and periphery values.

.

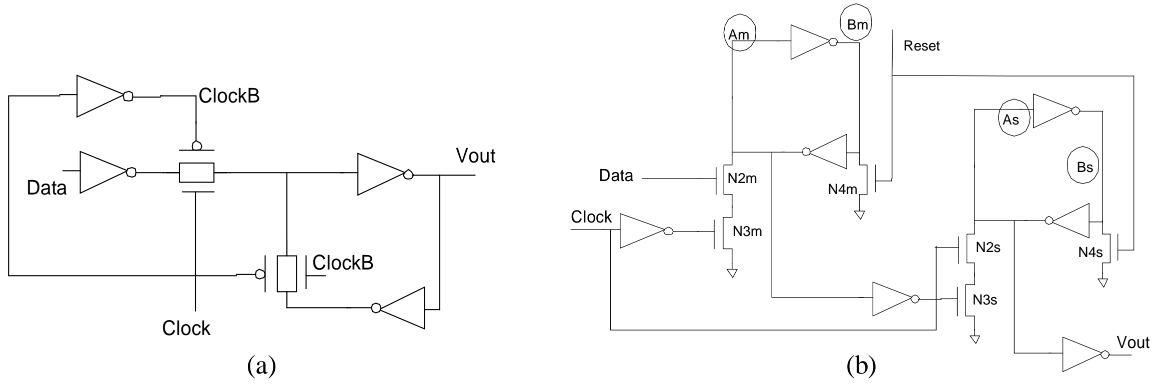


Figure 2. Latch schematics. (a) Conventional cell CMOS D-latch (b) Synchronously set-asynchronously reset flip-flop

To survey the impact of technology scaling on the metastability performance of the two latches considered in above transient simulations were performed with SPECTRE by parametric shifting of clock edge relative to the data transition edge to trigger the latch into metastable region. For all technologies, a level 11 CMOS transistor model of SPECTRE, which is actually BSIM3v3, is used. To do simulations and data processing in an automatic manner, a CSHELL script was written to dynamically change clock and data separation time, simulator accuracy options and calculate rise or fall time of output signal in each iteration of clock delay sweep. Sweep iteration is in fact a set of simulations by which a determined value of metastability window is swept. This metastability window, in which the clock delay time is changed in consecutive steps, is calculated from the previous sweep iteration such that the resulting window is narrower than the previous window. There is a point worth mentioning here regarding the calibration of the simulator. Simulator accuracy options are set dynamically to provide the required accuracy for evaluating vd(0) (initial differential output voltage of latch in metastable region). "vd(0)" is obtained from the following relation [5]

$$vd(0) = s \times \delta \quad (3)$$

where s is metastability slope of latch and it is a constant, which depends on the latch architecture and transistor sizes. δ is clock/data separation time. As the latch goes more deeply into the metastable region, more accurate simulator options are required. Thus, for the first run, accuracy options are set to appropriate values and then they are changed for each simulation iteration such that the accuracy increases as the simulation iterations move more deeply towards metastable region. For this paper the main accuracy options for the first run are obtained to be iabstol=1.0e-13, vabstol=1.0e-12, reltol=1.0e-8 and gmin=1e-19. Scale factor of accuracy options should be low enough to provide sufficient accuracy for the next run and high enough to ensure that the simulator does not have convergence problems.

The clock signal is varied relative to the data signal within a range focusing on the metastable region. In each run, the variation of the clock signal edge is determined from the previous runs to move latch more deeply into metastable region. The output voltage vectors from each simulation are searched from the end of the simulation back in time to measure rise or fall time of the output signal. In this way any probable ringing does not affect on the calculation of rise or fall time of the latch.

3. Simulation of metastability with respect of technology scaling

Three processes were selected for our purpose, namely the SCN035 (0.35u/3.3V, lambda=0.2u), SCN025 (0.25u/2.5V, lambda=0.12u) and SCN018 (0.18u/1.8V, lambda=0.09u) TSMC processes. We used typical process parameters for our simulations.

For a particular technology and latch architecture, the transistors were sized according to a constant rule (all device lengths were kept to the minimum device value for each technology). To be more conservative in our survey two sizing schemes were considered for the latches. For the conventional architecture of Fig. 2.a aspect ratios 4 and 2 were considered for n-channel transistors of inverters and pass transistors, respectively. The sizing scheme for p-channel transistors of this latch was considered based on optimizing inverter delay. For the architecture of Fig. 2.b aspect ratios are similar to the first latch except that for the internal inverters of latch (connected to the nodes Am, Bm, As and Bs), of which the n-channel and p-channel devices had the same size. This has been demonstrated to be an approach for device sizing to make τ_m minimum [1].

We measured signals from the data switching point [3]. We considered the delay of the latch to be the completion time of latch outputs, i.e. the time it takes for the output of latch to reach to the 90% of supply voltage for rising edge of output data and 10% of supply voltage for falling edge of output data. This consideration is because of the following reasons.

- An arbitrary circuit may follow the synchronizer, and the logic thresholds of this circuit can vary significantly..
- Synchronizers have a long latency along with swing when they are at the vicinity of the metastable region. For a decision circuit to be able to make correct decisions it is better to set the logic threshold levels more strictly than for the conventional digital circuits.

During our simulations, we modified the rise and fall time of the data and clock signals as the technology is scaled. For each process the circuit of Fig.3 was simulated to obtain a nominal data rise and fall time. To simulate metastability behavior, the rise and fall time of input data pulses were set equal to this obtained value and that of clock pulses were set equal to the half of this value. Table 1 shows the main characteristics of the p-channel transistors of the concerned processes in this paper and also the obtained characteristics from the circuit of Fig. 3.

As depicted in table 1, our simulations do not show a reduction of rise times consistent with the scaling factor when going from 0.35u to 0.25u. However the rise time scales as expected when going from 0.25u to 0.18u. This is because of poor p-channel transistors of 0.25u process as can be seen from table 1 by comparing parameters of K'_p and μ_p of two 0.35u and 0.25u processes. In our survey we pay attention to the unequal scaling factor between consecutive processes in the selected set of technologies.

To obtain, compare and correlate the resulting data points we focus on the range of clock/data separation time window, which produces longer latch delays than the normal delay of latch. First the obtained points from simulation were interpolated (using the Matlab v5.4 software package) to obtain 6×10^4 samples from about 110 to 120 simulated points in the metastable region for each process technology. Then the related histogram was drawn by considering 25 equal intervals in the range of minimum to maximum delay of the latch. It is well known that probability density function of producing a latch delay of "td" is proportional to e^{-td/τ_m} [3]. Thus τ_m is obtained from the histogram by the use of relation (4).

$$\tau_m = \frac{td2 - td1}{\ln N1 - \ln N2} \qquad (4)$$

In relation (4) N1 and N2 are the number of samples corresponding to the latch delays td1 and td2, respectively. τ_m is the negative inverse of slope of histogram with logarithmic scaled vertical axes. It must be noted that the mentioned proportionality is related to the linear region of the latch behavior, i.e. close to metastable point. For cases such as ours, in which the completion time is of interest, nonlinear characteristics of latch behavior as well as τ_m may affect the error probability. We discuss this further when we consider the detailed simulation results in the next few sections.

T_w is defined as the asymptotic width of clock/data separation time window when the delay of latch ideally goes towards zero. Relation (5) gives the width of metastability window for a given delay time [3].

$$\delta = T_w e^{-td/\tau_m} \qquad (5)$$

where td is the circuit delay time when clock and data separation time window is δ. From relation (5) to obtain T_w it is enough to obtain the intercept point of logarithmic curve δ versus td in the metastable region.

Table 1. Characteristics of the processes

Process / Parameter	0.35u/ 3.3V	0.25u/ 2.5V	0.18u/ 1.8V
$K'_p=\mu_p Cox_p/2$ ($\mu A/V^2$)	31	24.6	35.5
μ_p (Low Field) ($cm^2/V.s$)	136.46	81.21	86.36
$K'_n=\mu_n Cox_n/2$ ($\mu A/V^2$)	93.4	120.9	164.7
μ_n (Low Field) ($cm^2/V.s$)	411.14	399.14	400.65
Simulated rise time (ns)	0.323	0.29	0.146

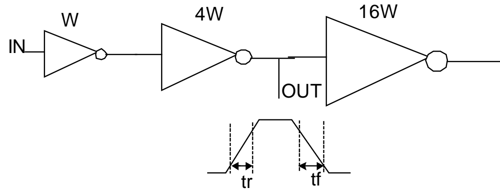


Figure 3. The measurement circuit for determination of rise and fall time of data pulses

Simulation Results of Conventional CMOS D-latch: In Fig. 4, we show the histogram obtained for the three technologies described above. These histograms have been drawn from generated samples in the manner, which was described previously. Roughness of curves is related to the simulator accuracy, which actually operates like noise. As we consider completion time of the latch it is better to consider an effective τ_m denoted as τ_{meff}.

From the three asymptotic lines in Fig. 4, we obtain the values of τ_{meff} shown in table 2. A useful measurement is to know how much τ_{meff} will be scaled as the latch delay time is scaled. This information is provided in table 2. The resulting value of τ_{meff} during scaling from 0.35u to 0.25u is the same as value resulted from theoretical calculations for τ_{m2}/τ_{m1} as the continuing calculations show. From relation (2) we have

$$\frac{\tau_{m2}}{\tau_{m1}} = \frac{G_{m1}(Area_2.C_{ox2}+2C_{F2})}{G_{m2}(Area_1.C_{ox1}+2C_{F1})} = \frac{C_{ox1}(V_{dd1}-V_{thn1}+V_{thp1})(Area_2.C_{ox2}+2C_{F2})}{C_{ox2}(V_{dd2}-V_{thn2}+V_{thp2})(Area_1.C_{ox1}+2C_{F1})} \quad (6)$$

Table 3 shows the required parameters of the processes to calculate the relation (6) for scaling from 0.35u to 0.25u. By the use of these parameters we obtain a value of 0.7 for τ_{m2}/τ_{m1}, which is exactly the value obtained from our simulations. This result also shows that the main delay is essentially related to τ_m, i.e. latch behavior in the non-linear region has a negligible contribution to the completion time. Thus $\tau_{meff2}/\tau_{meff1}$ (from simulation)=τ_{m2}/τ_{m1} (from hand calculations) when going from 0.35u to 0.25u.

The secondary effects such as velocity overshoot cause the "gm" of inverter transistors of the latch to be increased much more when going from 0.25u to 0.18u as we observed from simulation (this is also reflected in the values of low field mobility given in table 1). Therefore τ_{meff} and data/clock rise and fall times are scaled down by a factor more than the usual scaling factor of technology when going from 0.25u to 0.18u.

Fig. 5 shows the plot of logarithm of clock/data separation time window versus circuit delay in metastable region, which are well approximated by lines. The resulting T_w for each process is also given in table 2. T_w is a complex function of signal and clock edge rate as well as logic threshold level. Thus its scaling factor cannot be approximated. However simulation results show a considerable decrease of T_w when process is scaled down for conventional CMOS-D latch.

Simulation Results of Synchronously set-asynchronously reset flip-flop: Curves similar to that of figure 4 were also drawn for the synchronously set-asynchronously reset flip-flop. Table 4 shows the simulation results for the synchronously set-asynchronously reset flip-flop. By comparing tables 2 and 4 it can be seen that the simulation results for τ_{meff} and its scaling factor are very similar for both the latches considered in this paper. T_w is much higher for synchronously set-asynchronously reset flip-flop than that of conventional D-MOS latch. This is so especially because of its sizing, which was concerned with τ_m and not the speed of latch (the setup/hold time and latch delay is higher for this synchronously set-asynchronously reset flip-flop). It must be noted that setup and hold times are two of the factors affecting on the T_w.

4. Conclusions

In this paper, we have used simulation to study the impact of technology scaling on the metastability behavior of CMOS latches. It was shown that τ_m scales better than technology. It was also shown that considering completion time to account logic threshold mismatch effects does not change the results. T_w also scales down as technology scales but this scaling is a complex function of data and clock edge rate and setup/hold time scaling. In the future, we plan to include noise effects in our survey as technology scales down.

Acknowledgments

This work is carried out under financial support from INTEL Corporation, USA.

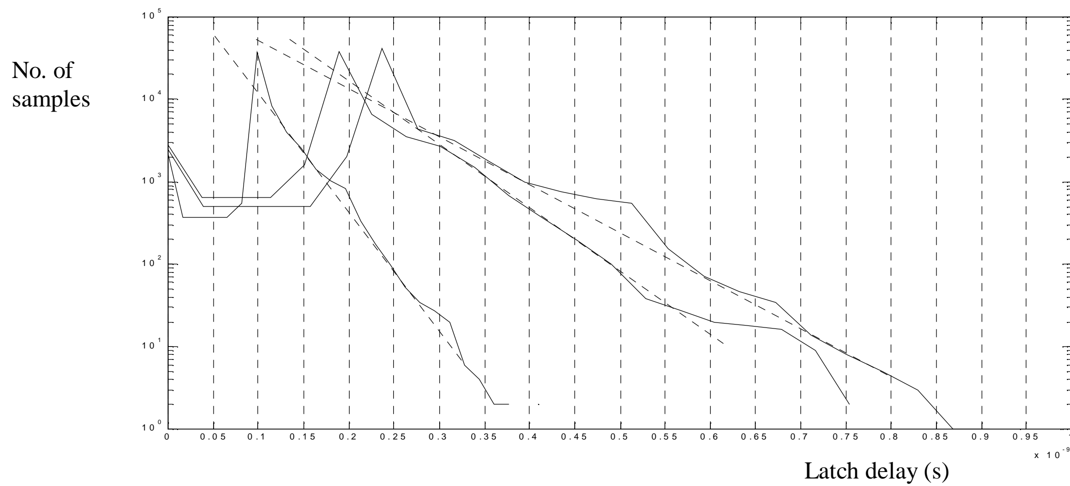
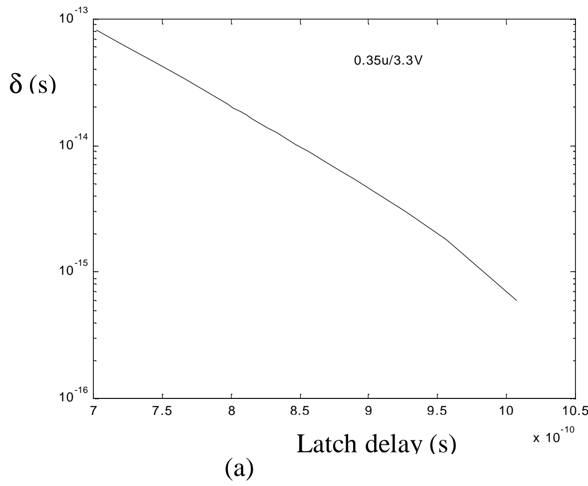
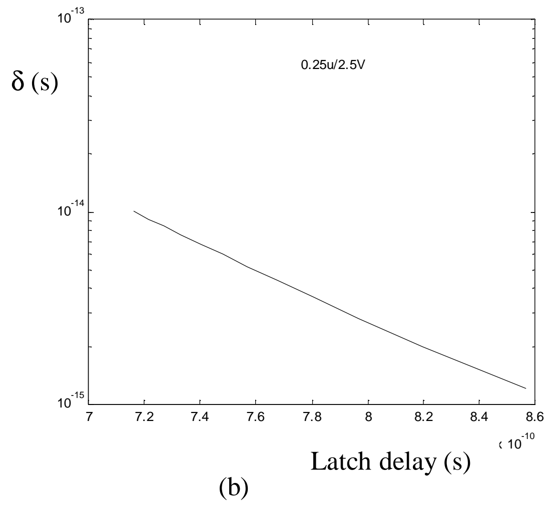


Figure 4. Latch delay histograms for three considered technologies related to conventional CMOS D-latch, focused in the vicinity of metastability region





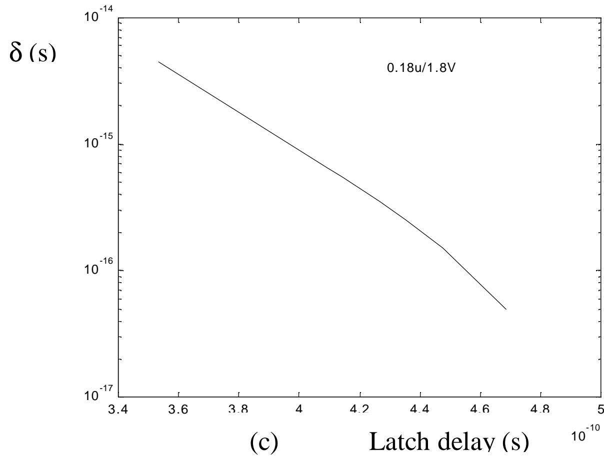


Figure 5. Plot of logarithm of clock data separation time window (δ) versus circuit delay for conventional CMOS D-latch. (a) 0.35u process (b) 0.25u process (c) 0.18u process

Table 2. Observed metastabilty characteristics for conventional CMOS D-latch

Specification---------------------------Process	0.35u/3.3V	0.25u/2.5V	0.18u/1.8V
Normal delay of latch (ns)	0.23	0.19	0.1
Simulated reference delay of the process from inverter chain of Fig. 3 (ns)	0.32	0.29	0.15
Swept metastability time window (for td > normal td) (ns)	0.25	0.2	0.09
τ_{meff} (ps)	78.2	54.3	21.7
S_λ=Scaling factor of lambda		0.6 (0.35u → 0.25u)	0.75(0.25u → 0.18u)
Sr=Scaling factor of clock/data edge rate (from inverter chain of Fig. 3)	-	0.9 (0.35u → 0.25u)	0.5 (0.25u → 0.18u)
Sm=Scaling factor of τ_{meff}	-	0.7 (0.35u → 0.25u)	0.4 (0.25u → 0.18u)
Sm/Sr	-	0.78	0.8
T_w (ns)	3.3	2.25	0.77

Table 3. Processes parameters

Process / Parameter	0.35u/ 3.3V	0.25u/ 2.5V
Lambda (um)	0.2	0.12
t_{ox} (m)	7.6e-9	5.7e-9
V_{th0n} (V)	0.486	0.39
V_{th0p} (V)	-0.735	-0.56
$CGDO_n$(F/m)	2.79e-10	6.2e-10
$CGDO_p$(F/m)	2.9e-10	6.66e-10

Table 4. Observed metastabilty characteristics for synchronously set-asynchronously reset flip-flop

Specification -------------------------Process	0.35u/3.3V	0.25u/2.5V	0.18u/1.8V
Normal delay of latch (ns)	0.45	0.36	0.19
Simulated reference delay of the process from inverter chain of Fig. 3 (ns)	0.32	0.29	0.15
Swept metastability time window (for td > normal td (ns))	0.25	0.23	0.1
τ_{meff} (ps)	72.8	51.4	21.7
S_λ=Scaling factor of lambda		0.6 (0.35u → 0.25u)	0.75(0.25u → 0.18u)
Sr=Scaling factor of clok/data edge rate (from inverter chain of Fig. 3)	-	0.9 (0.35u → 0.25u)	0.5 (0.25u → 0.18u)
Sm=Scaling factor of τ_{meff}	-	0.7 (0.35u → 0.25u)	0.4 (0.25u → 0.18u)
Sm/Sr	-	0.78	0.8
T_w (ns)	565	470	440

5. References

[1] Charles Dike and Edward (Ted) Burton, “Miller and noise effects in a synchronizing flip-flop,” *IEEE JSSC*, VOL. 34, NO. 6, pp. 849-855, June 1999.

[2] Tomasz Kacprzak and Alexander Albicki, “Analysis of metastable operation in RS CMOS flip-flops,” *IEEE JSSC*, VOL. SC-22, NO. 1, pp. 57-64, Feb. 1987.

[3] Clemenz L. Portmann and Tresa H. Y. Meng, “Metastability in CMOS library elements in reduced supply and technology scaled applications,” *IEEE JSSC*, VOL. 30, NO. 1, pp. 39-46, January 1995.

[4] S. T. Flanagan, “Synchronization reliability in CMOS technology,” *IEEE JSSC*, VOL. SC-20, NO. 4, pp. 880-882, Aug. 1985.

[5] Jackob H. Hohl, Wendell R. Larsen and Larry C. Schooley, “Prediction of error probabilities for integrated digital synchronizers”, *IEEE JSSC*, VOL. SC-19, NO. 2, pp. 236-244, April 1984.