Improved Device Technology Evaluation and Optimization

Daniel Connelly, Mark Foisy
Motorola, DigitalDNA Labs, Austin TX 78721

Abstract – The conventional I_{Dsat}-I_{DL} curve falls short in predicting which of two technology options will result in the best circuit performance. Here, for the first time, we demonstrate an improved evaluation method which accounts for process variation and leakage current budgeting for a target gate length. By using iteration or interpolation to compare tuned technologies, and by evaluating leakage and drive currents from the appropriate portions of their distribution curves, more effective optimization is achieved, giving stronger weight to robust device design.

I. Introduction

Fig. 1 shows a typical device performance comparison, as presented by Intel at the 1999 International Electron Device Conference [1]. I_{Dsat} (drain current for $V_{DS} = V_{GS} = V_{DD}$, $V_{BS} = 0$) is plotted *vs* I_{DL} (drain current for $V_{DS} = V_{DD}$, $V_{GS} = V_{BS} = 0$) over a range of gate lengths. For a target I_{DL}, the process which yields the greater I_{Dsat} is considered to have superior performance.

There are several problems with this sort of comparison:

- Devices between technologies at a given level of I_{DL} may have different L_G values. Since it is typical in a technology to establish a target gate length, then tune the threshold voltage to yield the desired I_{DL}, this comparison is misleading. Devices with smaller L_G at a given I_{DL} will tend to have greater I_{Dsat} at that value of I_{DL} due to the improved superthreshold transconductance which tends to come with shorter gate lengths.
- The metric fails to consider the variation in transistors at a given design point across a typical circuit. In multi-transistor circuits, the characteristic leakage and/or drive may not be the leakage and/or drive of an "average" device. Thus, simply constraining the leakage while optimizing the drive of a nominal device can lead to suboptimal real-world circuit performance.

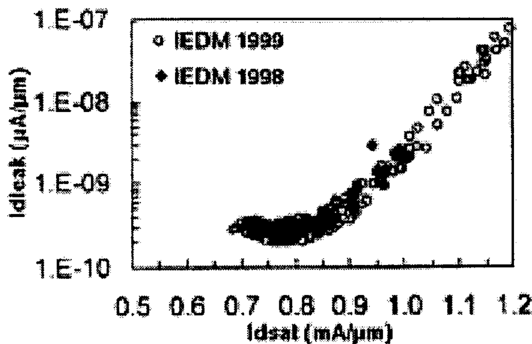

Fig. 1. Intel NMOS device comparison (1999 IEDM paper 17.1)

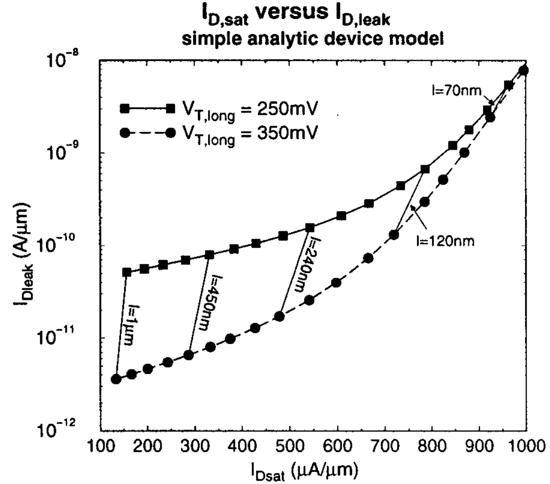

Fig. 2. Comparison of I_{Dsat} *vs* I_{DL} for a hypothetical technology differing only in long channel threshold, calculated using a simple analytic model.

II. Channel Length Effect

Consider the following simple model for the width-normalized drive and leakage currents in a n-channel field effect transistor (nFET) :

$$I_{Dsat} = K_{sat}(1/L_{eff})\,(V_{GS} - V_T)^2 \tag{1}$$

$$I_{DL} = (1/L_{eff})I_{Vt}\, e^{-K_{sub}Vt} \tag{2}$$

where V_T is a function of L_G, perhaps according to a relation such as :

$$V_T = V_{T\infty}\,(1 - L_{Vt}/L) \tag{3}$$

And, to crudely account for velocity saturation, L_{eff} is a function of length :

$$L_{eff} = \left(L^2 + L_0^2\right)^{1/2} \tag{4}$$

No claim is made that this model is an accurate representation of device currents, but it has the correct general behavior, and is sufficient for the purposes of this example.

A plot of I_{Dsat} *vs* I_{DL} with L_G as a parameter is shown in Fig. 2 for this model using typical values for the primary constants and two values for $V_{T\infty}$. Note these represent basically the same technology with, for example, two different threshold-adjust implant doses.[1] However, from the plot, the version with the greater V_T appears superior. This is a false conclusion, since when each version is "tuned" to achieve the target leakage at the target length, the two processes will be identical. For a proper comparison to be done, untuned processes should not be used.

[1]Neglected are such things as the influence of doping on the subthreshold slope and short-channel V_T-rolloff.

0-7803-6279-9/00/$10.00 © 2000 IEEE

III. Process Variation : Leakage

The leakage-drive tradeoff is a clear factor in the choice of the V_T for the target L_G. However, what is of interest is the performance and leakage current of the circuit, not of individual devices.

First consider leakage current. The sum of the leakage currents over all leakage-critical devices will be proportional to the average leakage of a device, not the leakage of an "average" device (*ie* the median leakage current).

Leakage current, unlike drive current, typically varies more smoothly on a logarithmic than on a linear scale versus device parameters. In other words, a statistical distribution of leakage current will be less normal than the distribution function of the logarithm of leakage current.

Consider a model in which process variation is abstracted into a normally distributed variable of unity variance, ϵ. Then, the probability distribution of ϵ is

$$P_\epsilon = \frac{1}{\sqrt{2\pi}} \exp[-\epsilon^2/2]. \tag{5}$$

As an example, the logarithm of leakage will be assumed to vary in a second-order fashion with ϵ,

$$I_{DL}\big|_\epsilon = I_0 \exp[\alpha_L \epsilon + \beta_L \epsilon^2]. \tag{6}$$

allowing the calculation of the device-averaged leakage,

$$\begin{aligned} \langle I_{DL} \rangle &= \int_{-\infty}^{\infty} P_\epsilon I_{DL}\big|_\epsilon d\epsilon \\ &= \frac{I_0}{\sqrt{1-2\beta_L}} \exp\left[\frac{\alpha_L^2/2}{1-2\beta_L}\right]. \end{aligned} \tag{7}$$

Note convergence is achieved only for $\beta_L < 1/2$. Values of ϵ corresponding to the average leakage current are plotted in Fig. 3. Note they tend to be in the range 0.2 to 2, depending on the values of α_L and β_L. Thus optimizing using the median of the leakage distribution, if doing so trades off on sensitivity to process parameters (*ie* greater α_L and/or β_L), may result in greater than expected circuit leakage.

An example of the use of this model as applied to realistic simulated devices is shown in Fig. 4. Simulations were done using ISE's DESSIS version 6.0.5 on structures generated using MDRAW/ISE version 6.0.6 [2]. The mobility model of Darwish *et al* [3] was used. Assuming the dominant contributor to variation in device performance is variation in gate length, and a "3σ" variation in gate length is 20 nm about a nominal 100 nm, then the resulting parameters from the second-order fit to the data are $\alpha_L = 1.04$, $\beta_L = 0.133$. These yield $\langle I_{DL} \rangle = 2.44\, I_0$, corresponding to a gate length shortfall of approximately 5.2 nm (0.78 standard deviations).

The above analysis addresses the difference between targeting the device with the average leakage and targeting the device with median leakage. A reason to target a device with still greater leakage is yield, accommodating chip-to-chip variation.

IV. Process Variation : Delay

Process variation also affects gate delay.

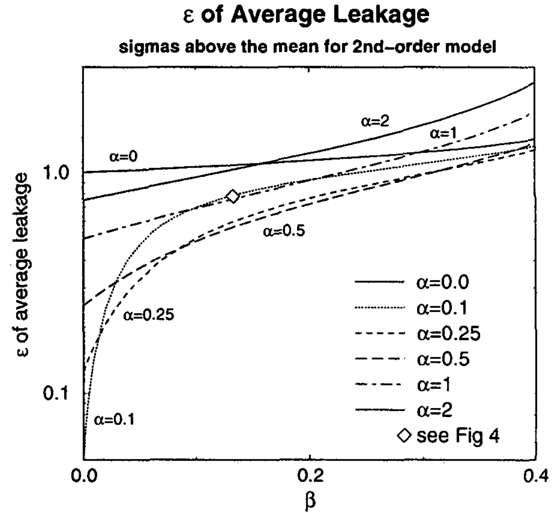

Fig. 3. Equation 7 evaluated for different values of α_L and β_L.

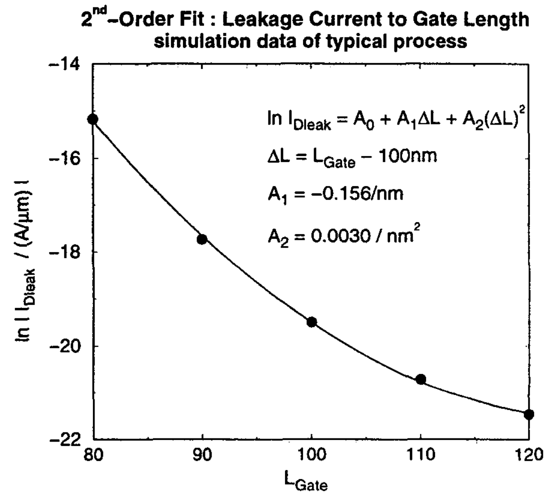

Fig. 4. Second-order fit of natural log of leakage current to gate length for simulated device data.

For a chain of gates, the net delay can be represented as the sum of the delays of the constituent gates. The net delay is thus proportional to the average per-gate delay. But for a well-designed large-scale synchronous digital circuit, there will likely be more than one candidate rate-limiting path. Any of a number of paths, given the right variation in individual transistor parameters, could be clock-limiting. Thus, performance will be limited not by the average gate delay, but by a slower-than-average portion of the population.

Clearly the distribution of critical paths in a circuit is design-dependent. However, a test case was considered in which there were 256 independent critical gate chain candidates (for example, two 64-bit bus lines, each with a rising and a falling edge of comparable speed), each dominated by four statistically-independent gate delays. The slowest of these chains was determined, and the average "σ" associated with the gate delays in the chain was calculated.

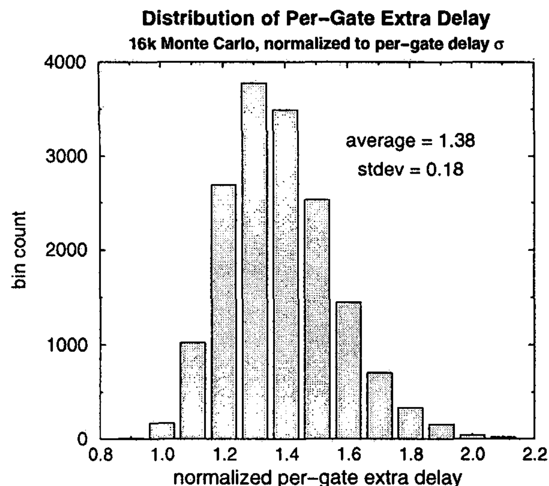

Fig. 5. Distribution of mean gate delays of clock-limiting gate chain for 256 paths each consisting of 4 statistically-independent gates. The average is 1.38 with $\sigma = 0.18$.

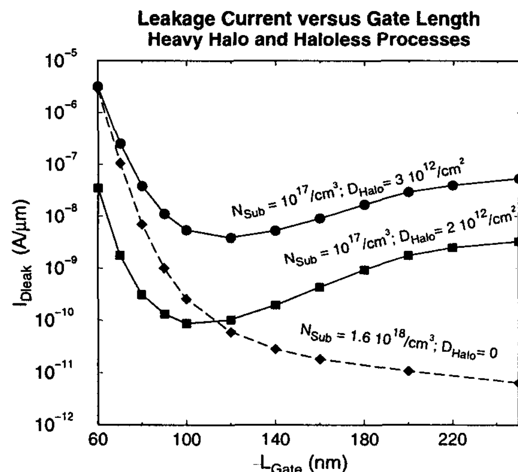

Fig. 6. Leakage current plotted *vs* gate length for two simulated technologies, a haloless process and a heavy halo process. For the latter, two halo doses are shown. Note the strong reverse short channel effect exhibited by the heavy halo process.

The binned result of 2^{14} such simulations is shown in Fig. 5.

The clear implication is that, for the simulated circuit, it isn't the delay of an average gate which determines performance limits, but rather the delay of a slower-than-average gate – in this case approximately 1.4 standard deviations slower. Note this is just the average result – in a substantial fraction of the cases, the rate-limiting path had gates on average even slower than this (standard deviation in average = 0.18 gate delay standard deviations). If gate delay variation is dominated by variation in drive current, then rather than gauge performance by the average drive current, this result implies the drive current of devices perhaps 1.6 standard deviations below the mean should be considered. For consideration of gate-length variation alone, this would be the "1.6σ supernominal" device.

V. Gate Length Effect

It is clear from the preceding analysis that simple evaluation of I_{Dsat} *vs* I_{DL} for typical devices over a range of gate lengths is not the best approach to evaluate tunable technologies.

The length issue can be avoided by comparing only tuned versions of the technologies. For example, if a technology generation is targeted at $L_G = 100\,\text{nm}$ with a target $I_{DL} = 1\,\text{nA}/\mu\text{m}$, then only technology options meeting that restriction should be compared. Usually, a few process parameters are tuned to hit the target, such as a threshold adjust implant dose or perhaps a halo implant dose.

Two approaches can be used – full process tuning, or response surface modeling and interpolation. The former approach is perhaps best suited to the simulation domain, where iteration can be used to find a value of the tuning parameter which results in an approach to within acceptable limits of the technology constraint. For silicon runs, however, due to the expense associated with experimental iterations, modeling and interpolation may be more appropriate. An example would be to interpolate the logarithm of the leakage current and the value of the drive current to the target device.

VI. Example : Doping Strategy Study

Consider an example of studying two different doping strategies for a technology, a heavy-halo approach and a uniform doping approach. Characteristic leakage *vs* gate-length curves for these technologies are shown in Fig. 6, with two values of the halo dose shown for the heavy-halo process. The heavy halo process used a $10^{17}/\text{cm}^3$ doping level in the body, sufficient to prevent punchthrough.

The target gate length was 100 nm. Target average static leakage was 1 nA/μm. Subject to this constraint, the goal was to maximize I_{Dsat}.[2] Only the effect of variation in gate length was considered, which was assumed to be normal with a standard deviation of 10 nm.

Two approaches can be considered. One is to compare nominal devices, tuned to the target leakage. The other is to take the approach recommended here, and look at leakage from subnominal and drive from supernominal devices.

To account for some degree of interdie process variation, an extra spread was applied to the values determined in the previous analysis – a 1σ subnominal device was used for leakage current, and a 2σ supernominal device was used for drive assessment.

The heavy halo process used halo dose to control the leakage current. The uniformly doped process used the well doping concentration. The key doping levels were iterated using the logarithm of leakage current until the target leakage at the target gate length was achieved to within 0.1%. For real devices, interpolation could have been done to somewhat lower accuracy.

A summary of the relevant simulations and results for both the nominal device approach and the multi-channel-length approach are shown in Table I. For the nominal device approach, a 100 nm gate length was used for leakage and drive currents. For the multi-length approach, leakage was specified at the 1-σ subnominal length while performance is evaluated for the 2-σ supernominal device.

In the table, L_G is in μm, current in per μm width. "(s)"

[2] Whether this is the best parameter to optimize is left as an open question.

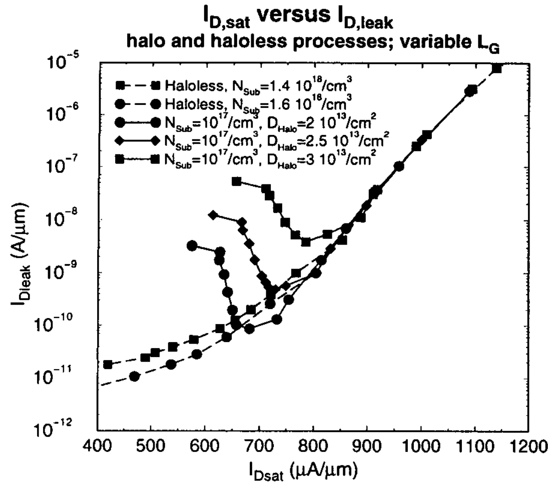

Fig. 7. Conventional I_{Dsat} - I_{DL}, where gate length is he variable parameter, comparing heavy halo and haloless processes. The comparison is somewhat inconclusive.

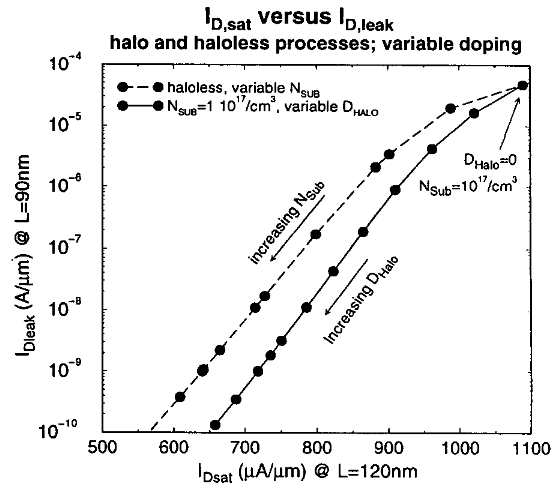

Fig. 8. Improved I_{Dsat} - I_{DL}, where doping (substrate or halo dose) is the variable parameter, comparing heavy halo and haloless processes. Leakage is evaluated at 1σ subnominal gate length while drive is evaluated at 2σ supernominal, for a nominal gate length of 100 nm and $\sigma = 10$ nm. The heavy halo process, with its improved short channel margin, is seen to be superior.

specifies substrate doping with no halo. "(h)" implies halo dose with $10^{17}/\text{cm}^{-3}$ substrate doping.

L_G	key doping	I_{DL}	I_{Dsat}
100	$2.36\,10^{13}/\text{cm}^2$ (h)	1.0 nA	768 μA
100	$1.42\,10^{18}/\text{cm}^3$ (s)	1.0 nA	767 μA
90	$2.50\,10^{13}/\text{cm}^2$ (h)	1.0 nA	
120	$2.50\,10^{13}/\text{cm}^2$ (h)		718 μA
90	$1.61\,10^{18}/\text{cm}^3$ (s)	1.0 nA	
120	$1.61\,10^{18}/\text{cm}^3$ (s)		640 μA

TABLE I
DEVICE EVALUATION SPLITS

The conclusion of which process is superior differs using the two approaches. Using the conventional approach, the two technologies are of virtually indistinguishable merit. However, using the multi-channel-length approach, the superior short-channel-immunity of the heavy halo process make it the clear preference. The "nominal" approach fails to optimize circuit performance.

VII. EXAMPLE : I_{Dsat} - I_{DL} CURVE

The conventional I_{Dsat} - I_{DL} curve is done for fixed doping levels, gate length parametrically varied for each technology, and I_{DL} plotted logarithmically *vs* I_{Dsat}. An example for the simulated devices in the previous section is shown in Fig. 7. The curves are shown for various halo doses along with haloless designs.

An improved version of the I_{Dsat} - I_{DL} uses the principles discussed in this work. Instead of parametrically varying length, doping is parametrically varied, as this is to be the parameter tuned to match the leakage target for the technology. And instead of evaluating leakage and drive current on the same device, a 1σ subnominal device is used to evaluate the leakage current while a 2σ supernominal device is used to evaluate drive (again, here, it is assumed gate length is the principle contributor to variation in device performance). A comparison of the heavy halo and haloless processes done in this fashion is shown in Fig. 8. The superiority of the heavy halo process is clear.

VIII. CONCLUSION

We have shown, for the first time, how improvements to the conventional I_{Dsat} - I_{DL} approach to device optimization can be improved to better predict the performance of complex circuits. Two principle improvements were suggested. One is that, for technologies in which gate length is established for reasons other than leakage current control, comparisons be made between devices at comparable gate lengths. Another is to properly consider the effects of variation in device characteristics. On the leakage side, it must be recognized that leakage tends to vary exponentially with variations in physical characteristics of the device, and therefore leakage is not normally distributed. On the drive side, it must be recognized if there are multiple candidates for performance-limiting circuit path status, performance will likely be limited by below-average performing devices. These considerations result in different devices (or portions of the associated distribution curves) being used to gauge leakage and drive. The result is stronger emphasis given to processes which more robustly handle process variation, such as heavy halo processes.

ACKNOWLEDGMENT

Thanks to Srinivas Jallepalli, Scott Luning, Surya Veeraraghavan, Michael Duane for their contributions to this work.

REFERENCES

[1] T. Ghani *et al.*, IEDM Tech. Digest, 1999

[2] Integrated Systems Engineering Inc.; San Jose, CA

[3] M. Darwish *et al.*, IEEE Trans Elec Dev, v44 n 9, p 1529, Sep 1997