Toward More Accurate Scaling Estimates of CMOS Circuits from 180 nm to 22 nm

Aaron Stillmaker, Zhibin Xiao, and Bevan Baas
VLSI Computation Lab
Department of Electrical and Computer Engineering
University of California, Davis
Technical Report ECE-VCL-2011-4
{astillmaker, zxiao, bbaas}@ucdavis.edu

***Abstract*— With deep submicron technology nodes other methods are needed to obtain scaling factors rather than the traditional scaling factors which held for the pre-submicron era. This work presents scaling factors between major technology nodes between 180 nm and 22 nm operating at voltages from 1.8 V to 0.7 V. Common operating data for these technologies were taken from the International Technology Roadmap for Semiconductors (IRTS). HSpice simulations that rely on the Predictive Technology Model (PTM) for transistor characteristics were used to find the scaling factors.**

I. Introduction

It is useful to be able to compare between digital design results of circuits that are made using different technology sizes. It has become apparent that traditional scaling does not hold with deep submicron technology sizes. This work models devices from different technology sizes to get the scaling factors of energy, delay, and area between sizes. One of the large motivations for this project was the ability to compare different chips in a fair fashion, the size and voltage were varied over a large range of values, and the power, energy, and intrinsic delay were all measured. From these numbers, scaling factors were extrapolated.

A. Traditional Scaling Methods

Until the deep submicron era, transistor characteristics scaled very predictably as designers generally focused on simple geometry scaling. This included generalized equations which take short-channel effects into consideration. With this, great gains were to be had following this method, with few outside factors forcing designers to optimize non-traditionally. These scaling factors have been used and taught, so they are easily found in the literature [1], [2], and is shown in Table I where scaling factor S is the ratio of the all transistor geometry between two transistor sizes and U is the ratio between two voltages. With both S and U it is expected that all geometry and voltages scale together.

B. Scaling Equations in Submicron Technologies

As transistors get smaller, short channel effects and other issues such as process variation start playing a larger role, making the traditional scaling equations inaccurate. Leakage current is affected greatly by gate length, oxide thickness, and threshold voltage, so it is becoming a large issue with deep submicron processes. With these new issues now affecting transistor operation, designers started looking to optimize between technology nodes other than simple geometric scaling. Width, length, and oxide thickness are not scaling together as well as V_{DD} and V_T. Which means that scaling factors S and U shown in Table I can not be determined. The above mentioned problems are especially noticeable when the industry switched largely to using high-κ dielectrics and metal gates with technology nodes at 45 nm and smaller.

C. Methods for Accurate Scaling

The physics that affect transistors as they reach the submicron region get far more complicated, with leakage, and other issues becoming a large factor in energy consumption and delay. So the most accurate way to get scaling factors in submicron processes is to use a simulation tool, such as HSpice with a model that specifies the characteristics of the particular technology. Simulating a whole design in Spice with modified technology size and voltages would result in the most accurate comparison, but this is not a practical solution because other issues such as transistor sizing can arise with a large change. Not to mention, one generally does not have the exact specifications of the design they want to compare with. Therefore a good approximation of a device in a certain technology is to use inverters in a chain, with 4 inverters attached to each output, this is known as Fan Out 4, or FO4. A circuit that has a delay and consumption of X number of F04 inverter chains in a certain technology size, should have roughly the same X number of FO4 inverter chains in a different technology size. With this in mind, this work sets out to take measurements of FO4 models in many different sizes and voltages so that roughly accurate scaling factors can be used.

II. International Technology Roadmap for Semiconductors (IRTS)

The International Technology Roadmap for Semiconductors (ITRS) [3], [4], [5], [6], [7], [8] is a report which predicts where semiconductor technology is headed in the next 15 years, and is made on odd years, with updates on even

TABLE I
SCALING SCENARIOS FOR SHORT-CHANNEL DEVICES, PRE-SUBMICRON [1].

Parameter	Relation	Full Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{ox}		$1/S$	$1/S$	$1/S$
V_{DD}, V_T		$1/S$	$1/U$	1
$Area/Device$	WL	$1/S^2$	$1/S$	$1/S^2$
$Power$	$I_{sat}V$	$1/S^2$	$1/U^2$	1
$IntrinsicDelay$	$R_{on}C_{gate}$	$1/S$	$1/S$	$1/S$
$Energy$	Pt	$1/S^3$	$1/SU^2$	$1/S$

years. These reports are formed by a collaboration of many companies and research institutions. In this work, these were used to get industry standard technology sizes, and voltages commonly used, as well as general knowledge about transistor changes over the years. Area is also of interest in digital design. Standard 1/2 metal 1 sizes were taken to make scaling factors. In this report, when technology process sizes are referred to, i.e. 180 nm, 45 nm, etc. we are referring to the minimum feature size. Process sizes were generally identified by their smallest feature size, and for a long time DRAM 1/2 pitch sizes were the smallest, and therefore used to identify technologies. With new fabrications, this has not been the case, and IRTS no longer identifies technologies by their minimum feature size to try and stop confusion, they differentiate using the first year of production [4]. However, as minimum feature technology notes have continued to be the generally accepted term, and because we are comparing with older technologies, in this work we will be identifying technologies by both the production year, and minimum feature size, as shown in Table II.

III. Predictive Technology Model (PTM)

A. PTM Overview

The Predictive Technology Models [9], [10], [11], or PTM, was used to simulate different design characteristics as technology size, and voltage changed. The models were developed for designers who do not have access to proprietary transistor characteristics to test designs with future technology nodes.

B. HSpice Modeling

HSpice was used to model the scaling results. A fan out four, FO4, inverter chain was used. FO4 delay has been shown to be proportional to *CV/I* (intrinsic capacitance, voltage, and drive current of a device) [12]. Intrinsic capacitance and drive current are both proportional to device size, so they scale with *1/S* and as previously mentioned, voltage scales at factor U, thus delay should scale with U/S^2. The chain starts with one minimum size inverter with the output connected to 4 inverters, with that output connected to 16 inverters, and so on until the circuit ends with 64 inverters, creating a total of 4 FO4 stages. A square wave input was sent into the chain. The set of 16 inverters, directly in the middle of the chain, was used for the sampling. The delay between when the input signal to the set of 16 inverters crossed the mid point and the output crossed the midpoint was measured. The voltage was measured along with the current, and calculations were made by the simple equations 1-4 where t_0 to t_1 is the transition time as the signal goes high, and t_2 to t_3 is the transition time as the signal goes low.

$$P_{ave} = \frac{1}{T}\int_0^T I(t)\cdot V dt \tag{1}$$

$$E_{up} = \int_{t_0}^{t_1} I(t)\cdot V dt \tag{2}$$

$$E_{down} = \int_{t_2}^{t_3} I(t)\cdot V dt \tag{3}$$

$$E_{ave} = \frac{E_{up} + E_{down}}{2} \tag{4}$$

All these calculations are done by HSpice and provided as a simulation output.

1) Simulation Parameters: The simulations were run on the technology sizes: 180 nm, 90 nm, 65 nm, 45 nm, 32 nm, and 22 nm with supply voltages varying from 1.3 V to 0.7 V in 0.1 V increments and 1.8 V. The 180 nm stopped working at 0.8 V and 0.7 V, as this is a lower voltage than standard 180 nm was expected to run at. Similarly technology nodes are not designed to handle voltages much higher then their target voltages, so even though HSpice gave results for the smaller technology nodes operating at higher voltages, they were removed from the results as the PTM characteristics should not hold for these values. These spots are marked as N/A in the results tables.

With the industry standard of using high-κ dielectric transistors at 45nm and below, high-κ PTM models were used as the most accurate models available because semiconductor companies do not readily provide characteristics of their specific technologies. This level of generality can actually help when being used to compare two chips of different fabrications.

IV. Simulation Results and Scaling Factors

As previously mentioned, industry standard numbers were taken from IRTS reports. Table II shows the standard values labeled by IRTS as a high performance circuit with a heat sink, at each technology node that was investigated, as well as the delay and power simulated using the inverter chain in HSpice as described in Section III-B. The V_{DD} is taken from the IRTS tables for high-performance.

A. Area

To determine a factor for scaling area between technologies, minimum feature sizes and Metal 1 half pitches were taken from the IRTS reports. The Logic Gate (4 Transistors) size was taken from the IRTS tables of MPU (High-volume Microprocessor). An exact scaling would be dependent on the design, but using either of the aforementioned values should give a good rough estimate. The relative scaling of area is shown in Figure 1 and detailed in Table III. Tables IV-VII can be used to scale areas depending on preference. To scale find the

TABLE II

CHARACTERISTICS OF DIFFERENT TECHNOLOGIES NODES FOR HIGH PERFORMANCE PROCESSOR WITH A HEATSINK.

Production Year	Tech. Node (nm)	Min. Vdd (V)	Max. Power (W)	Energy (fJ)	Delay (ps)
1999	180	1.8	90	19	80.0
2001	130	1.2	130	3.4	35.5
2004	90	1.1	149	1.7	27.2
2007	65	1.1	189	1.1	20.4
2008	45	1.1	146	0.64	7.82
2010	32	0.97	146	0.30	6.33
2012	22	0.9	158	0.17	5.12

TABLE III

AREAS OF DIFFERENT ASPECTS OF DIFFERENT TECHNOLOGY NODES [3], [4], [5], [6], [7], [8].

Minimum Feature Size (nm)	Metal I Half Pitch (nm)	(4T) Logic Gate Size (μm^2)	Relative Area from Feature Size	Relative Area from M1 $\frac{1}{2}$ Pitch	Relative Area from Logic Gate
180	230	57	7.7	11	22
130	150	10.4	4	4.9	4
90	90	5.2	1.9	1.8	2
65	68	2.6	1	1	1
45	59	2.1	0.48	0.75	0.81
32	45	0.71	0.25	0.44	0.27
22	32	0.35	0.11	0.22	0.14

TABLE IV

AREA SCALING USING MINIMUM FEATURE SIZE.

Technology Nodes (nm)	180	130	90	65	45	32	22
180	1	1.9	4.0	7.7	16	32	67
130	0.52	1	2.1	4.0	8.4	17	35
90	0.25	0.48	1	1.9	4.0	7.9	17
65	0.13	0.25	0.52	1	2.1	4.1	8.7
45	0.063	0.12	0.25	0.48	1	2.0	4.2
32	0.032	0.061	0.13	0.24	0.51	1	2.1
22	0.015	0.029	0.060	0.11	0.24	0.47	1

starting technology node from the top column, then go down until the desired technology is reached, then that number can be multiplied by the chip's area to determine an equivalent area in the different technology node.

B. Delay

Table VIII contains the results from the simulations, showing the time for a signal to propagate through one minimum sized inverter in the middle of the FO4 inverter chain. Equation 5 can be used to scale delay where the factor Delay Factor (*DF*) is found in table VIII.

$$D_x = \frac{DF_x}{DF_y} \cdot D_y \tag{5}$$

$$(DF \text{ from Table VIII})$$

TABLE V

AREA SCALING USING METAL I HALF PITCH SIZE.

Technology Nodes (nm)	180	130	90	65	45	32	22
180	1	2.4	6.5	11	15	26	52
130	0.43	1	2.8	4.9	6.5	11	22
90	0.15	0.36	1	1.8	2.3	4.0	7.9
65	0.087	0.21	0.57	1	1.3	2.3	4.5
45	0.066	0.15	0.43	0.75	1	1.7	3.4
32	0.038	0.090	0.25	0.44	0.58	1	2.0
22	0.019	0.046	0.13	0.22	0.29	0.51	1

TABLE VI

AREA SCALING USING 4 TRANSISTOR LOGIC SIZE.

Technology Nodes (nm)	180	130	90	65	45	32	22
180	1	5.5	11	22	27	80	160
130	0.18	1	2	4	5	15	30
90	0.091	0.5	1	2	2.5	7.3	15
65	0.045	0.25	0.5	1	1.2	3.7	7.4
45	0.037	0.2	0.4	0.81	1	3	6
32	0.013	0.068	0.15	0.27	0.35	1	2
22	0.0061	0.033	0.068	0.13	0.17	0.49	1

TABLE VII

AREA SCALING FACTORS USING GEOMETRIC MEAN OF THREE ASPECTS FROM TABLE III.

Technology Nodes (nm)	180	130	90	65	45	32	22
180	1	2.9	6.6	12	19	40	83
130	0.34	1	2.3	4.3	6.4	14	28
90	0.15	0.44	1	1.9	2.8	6.1	13
65	0.080	0.23	0.53	1	1.5	3.3	6.6
45	0.053	0.16	0.35	0.66	1	2.2	4.4
32	0.025	0.072	0.16	0.31	0.46	1	2.1
22	0.012	0.035	0.080	0.15	0.23	0.49	1

Figure 4 shows the results of simulations using the values in Table II compared to what we would expect the results to be if we were using the equations from Table I. Figures 5 shows the scaling of delay with varied technology sizes and operating voltages.

C. Energy and Power

Table IX contains the results from the simulations. Equation 6 can be used to scale energy. Equation 7 combines Equations 5 & 6 using the power equation $Power = Energy/Delay$. The factors Delay Factor (*DF*) and Energy Factor (*EF*) are found in table VIII & IX.

$$E_x = \frac{EF_x}{EF_y} \cdot E_y \tag{6}$$

$$P_x = \frac{EF_x \cdot DF_y}{EF_y \cdot DF_x} \cdot P_y \tag{7}$$

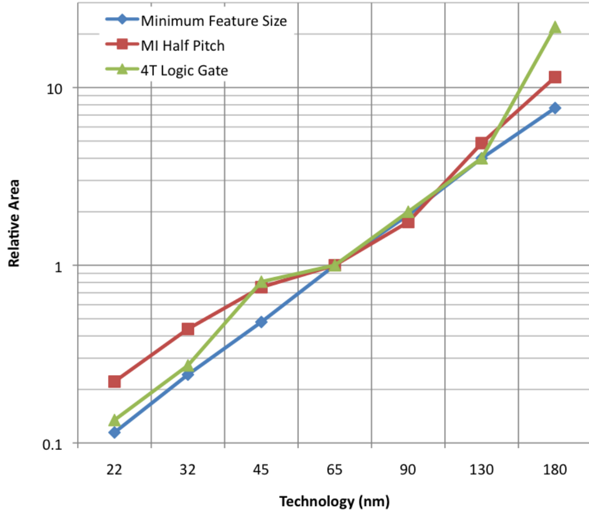

Fig. 1. Relative area scaling of different area aspects over different technology nodes.

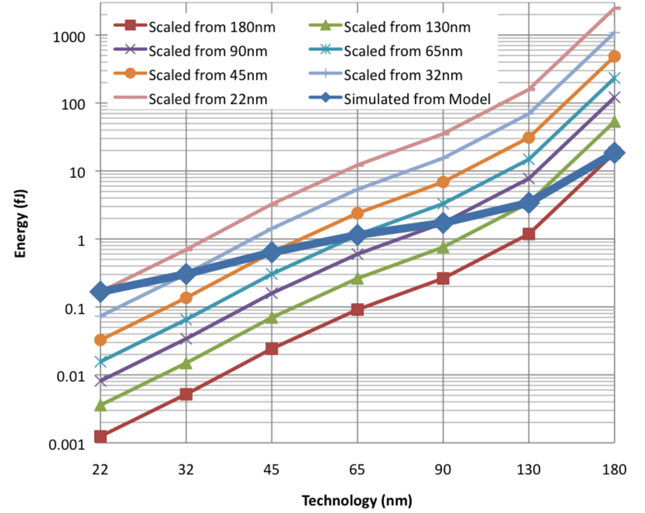

Fig. 2. Energy used to toggle one minimum sized inverter in the middle of the FO4 inverter chain simulated using Table II values and scaled using Table I equations.

TABLE VIII
TIME FOR A SIGNAL TO PROPAGATE THROUGH ONE MINIMUM SIZED INVERTER IN THE MIDDLE OF THE FO4 INVERTER CHAIN IN ps.

	1.8 V	1.3 V	1.2 V	1.1 V	1 V	0.9 V	0.8 V	0.7 V
180 nm	80.03	110.3	121.2	134.8	152.2	176.1	N/A	N/A
130 nm	N/A	33.44	35.52	38.29	42.07	47.53	55.84	69.81
90 nm	N/A	N/A	25.2	27.22	30.02	34.1	40.39	51.19
65 nm	N/A	N/A	18.84	20.45	22.7	25.99	31.23	40.49
45 nm	N/A	N/A	7.656	7.824	8.092	8.502	9.166	10.24
32 nm	N/A	N/A	N/A	6.042	6.255	6.576	7.093	7.967
22 nm	N/A	N/A	N/A	N/A	4.878	5.17	5.619	6.359

TABLE IX
ENERGY USED TO TOGGLE A SINGLE MINIMUM SIZED INVERTER IN THE MIDDLE OF THE FO4 INVERTER CHAIN IN fJ.

	1.8 V	1.3 V	1.2 V	1.1 V	1 V	0.9 V	0.8 V	0.7 V
180 nm	19	9.2	7.7	6.3	5.0	3.98	N/A	N/A
130 nm	N/A	4.1	3.4	2.8	2.2	1.8	1.4	1.1
90 nm	N/A	N/A	2.1	1.7	1.4	1.1	0.88	0.69
65 nm	N/A	N/A	1.3	1.1	0.92	0.74	0.59	0.45
45 nm	N/A	N/A	0.80	0.64	0.48	0.36	0.27	0.202
32 nm	N/A	N/A	N/A	0.45	0.33	0.27	0.18	0.13
22 nm	N/A	N/A	N/A	N/A	0.23	0.17	0.12	0.084

$$(EF \text{ from Table IX } DF \text{ from Table VIII})$$

Figure 2 shows the results of simulations using the values in Table II compared to what we would expect the results to be if we were using the equations from Table I. Figure 3 show the scaling of energy with varied technology sizes and operating voltages.

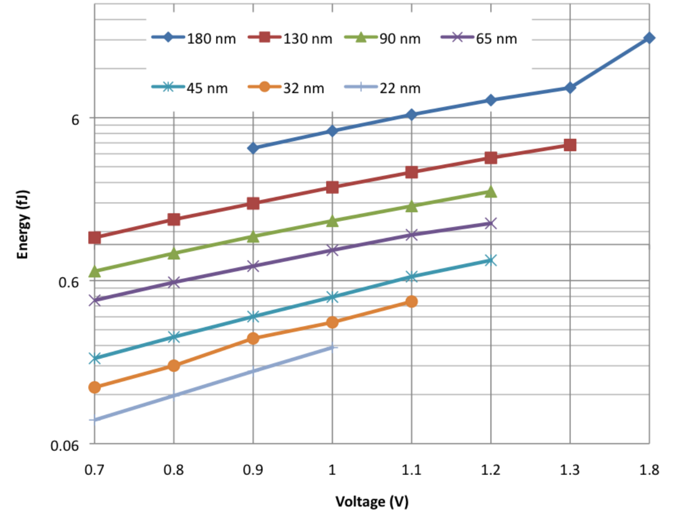

Fig. 3. Energy required for one minimum sized inverter in the middle of the FO4 inverter chain to toggle for different technology nodes with scaling voltage.

V. CONCLUSION

Comparing the data presented in this report with the traditional scaling methods, indeed does not hold into these submicron transistors. The general trend is similar, but would not make an accurate comparison. Thus the factors gleaned from simulations presented in this work is a more accurate estimation that can be used to compare two devices from different technologies.

VI. ACKNOWLEDGEMENT

The authors gratefully acknowledge support from Fudan University ASIC and System National Lab open project funding, ST Microelectronics, C2S2 Grant 2047.002.014, NSF

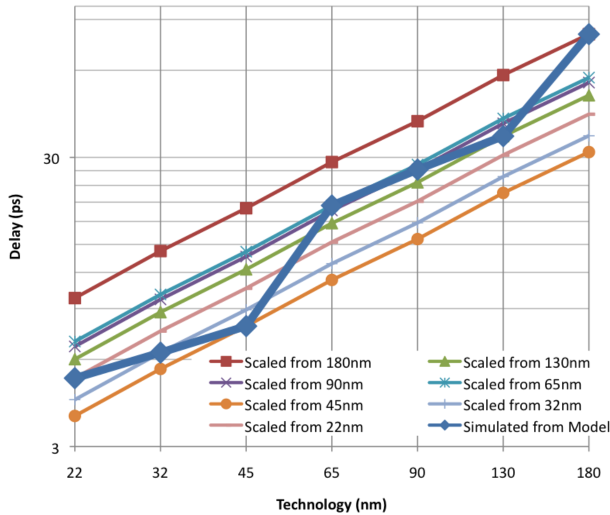

Fig. 4. Delay Simulated for a signal to propagate through one minimum sized inverter in the middle of the FO4 inverter chain using Table II values and scaled using Table I equations.

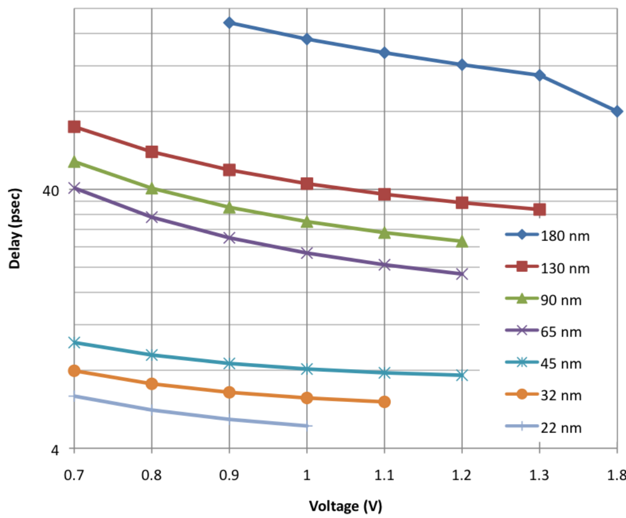

Fig. 5. Delay for a signal to propagate through one minimum sized inverter in the middle of the FO4 inverter chain for different technology nodes with scaling voltage.

Grant 0430090 and CAREER Award 0546907, SRC GRC Grant 1598 and CSR Grant 1659, Intel, UC Micro, Intellasys, SEM, and a UCD Faculty Research Grant.

References

[1] Jan M. Rabaey, Anantha P. Chandrakasan, and Borivoje Nikolic, *Digital integrated circuits : a design perspective*, Pearson Education, Upper Saddle River, NJ, second edition, 2003.

[2] John P. Uyemura, *Introduction to VLSI circuits and systems*, John Wiley & Sons, Inc., Hoboken, NJ, first edition, 2002.

[3] "International technology roadmap for semiconductors 2010 update," Tech. Rep., ITRS, 2010.

[4] "International technology roadmap for semiconductors 2009 edition," Tech. Rep., ITRS, 2009.

[5] "International technology roadmap for semiconductors 2007 edition," Tech. Rep., ITRS, 2007.

[6] "International technology roadmap for semiconductors 2004 update," Tech. Rep., ITRS, 2004.

[7] "International technology roadmap for semiconductors 2002 update," Tech. Rep., ITRS, 2002.

[8] "International technology roadmap for semiconductors 2000 update," Tech. Rep., ITRS, 2000.

[9] W. Zhao and Y. Cao, "New generation of predictive technology model for sub-45nm early design exploration," *IEEE Transactions on Electron Devices*, vol. 53, no. 11, pp. 2816–2823, Nov. 2006.

[10] Y. Cao, T. Sato, D. Sylvester, M. Orshansky, and C. Hu, "New paradigm of predictive mosfet and interconnect modeling for early circuit design," in *CICC*, 2000, pp. 201–204.

[11] "Predictive technology model," `http://ptm.asu.edu/`.

[12] "Fo4 writeup: International technology roadmap for semiconductors 2003 edition," Tech. Rep., ITRS, 2002.