Introducing 20 nm technology in Microwind

Etienne Sicard

Introducing 20 nm technology in Microwind

Etienne SICARD
Professor
INSA-Dgei, 135 Av de Rangueil
31077 Toulouse – France
www.microwind.org
email: Etienne.sicard@insa-toulouse.fr

This paper describes the implementation of the CMOS 20 nm technology of a High Performance Bulk Planar 20nm CMOS Technology proposed by the Joint Development Alliance (JDA) in Microwind38. Power, performance and area (PPA) gains related to the 20 nm technology are illustrated, and new concepts such as design for manufacturing, double-patterning, replacement metal gate process are described. The performances of a ring oscillator layout and a 6-transistor RAM memory layout are also analyzed.

1. The merge of giants

The Joint Development Alliance (JDA) (http://www.commonplatform.com) gathers several IC companies including IBM, Renesas, STMicroelectronics, Samsung Electronics, Toshiba and GLOBALFOUNDRIES for combining the expertise and research resources of all of its partners, share the enormous fab development costs (estimated to more than 5 B$), and propose joint nano-CMOS technologies faster than by operating as individual companies. The JDA has recently released a High Performance Bulk Planar 20nm CMOS technology [Huiling2012], as a result of a joint R&D collaboration between partners. We recall in table 1 the main innovations over the past 15 years, with an illustration of some key novelties in Fig. 1.

Technology node	Year of introduction	Key Innovations	Application note
180nm	2000	Cu interconnect, MOS options, 6 metal layers	
130nm	2002	Low-k dielectric, 8 metal layers	
90nm	2003	SOI substrate	[Sicard2005]
65nm	2004	Strain silicon	[Sicard2006]
45nm	2008	2nd generation strain, 10 metal layers	[Sicard2008]
32/28nm	2010	High-K metal gate	[Sicard2010]
20nm	2013	Replacement metal gate, Double patterning, 12 metal layers	This application note
14nm	2015	FinFET	To appear

Table 1: most significant technology improvements over the past 15 years

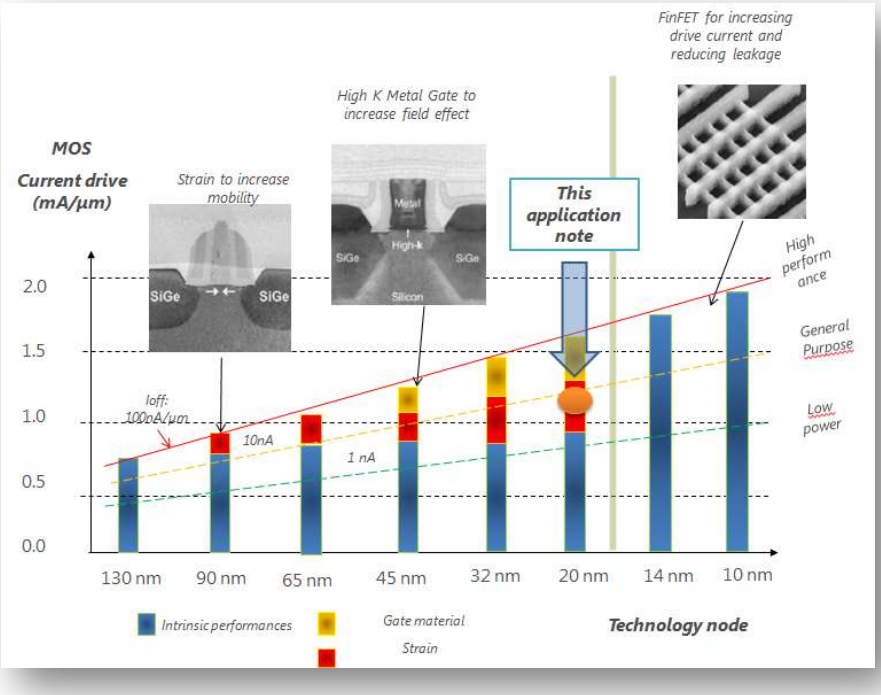


Figure 1: Increased switch performances thanks to continuous innovations

The naming of the technology nodes (130, 90.. 14nm) comes from the International Technology Roadmap for Semiconductors (http://www.itrs.net). The trend of CMOS technology improvement continues to be driven by the need to

- Integrate more functions within a given silicon area
- Reduce the size of a given function
- Dissipate less power
- Reduce the IC fabrication cost
- Increase operating speed

At each technology node corresponds a wide variety of performances, depending whether the foundry is targeted to “high performance” devices (speed whatever the power consumption), “general purpose” or “Low Power” (lower speed but power-efficient). As we may see, the 20-nm technology proposed in Microwind is close to “General Purpose” characteristics.

2. Towards 100 Billion devices on a chip: what for?

The compilation of more than 100 device characteristics (Intel, AMD, ARM..) with associated technology, year of introduction and number of transistors (expressed in million) is reported in Fig. 1. It can be seen that the 20-nm technology node enables chip designs with around 10 Giga-transistors on a die.

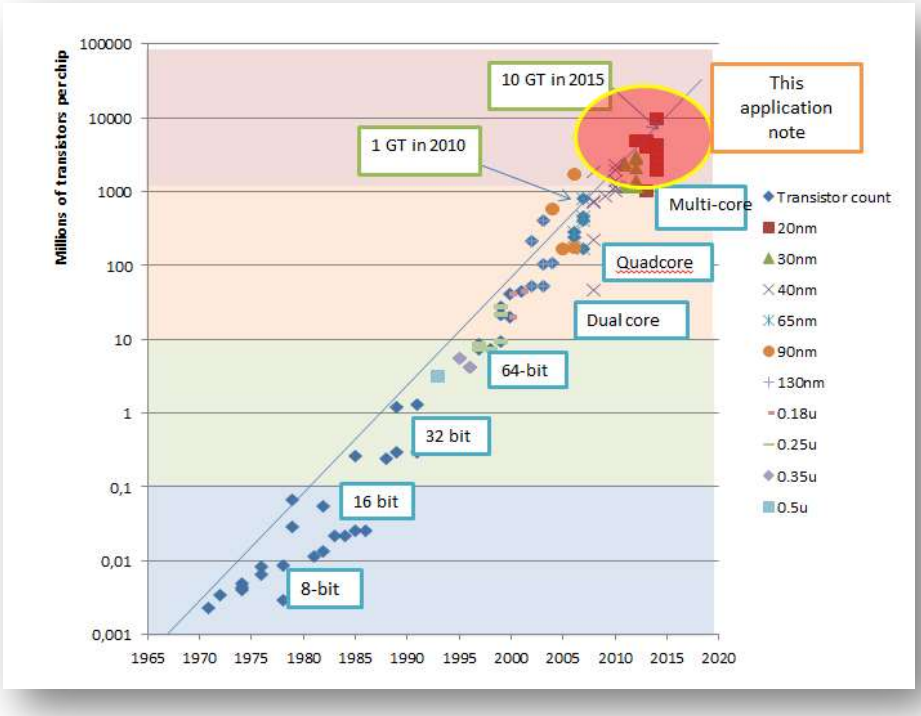


Figure 2: The 20-nm technology node offers design complexities approaching 10 billion devices on a same silicon die

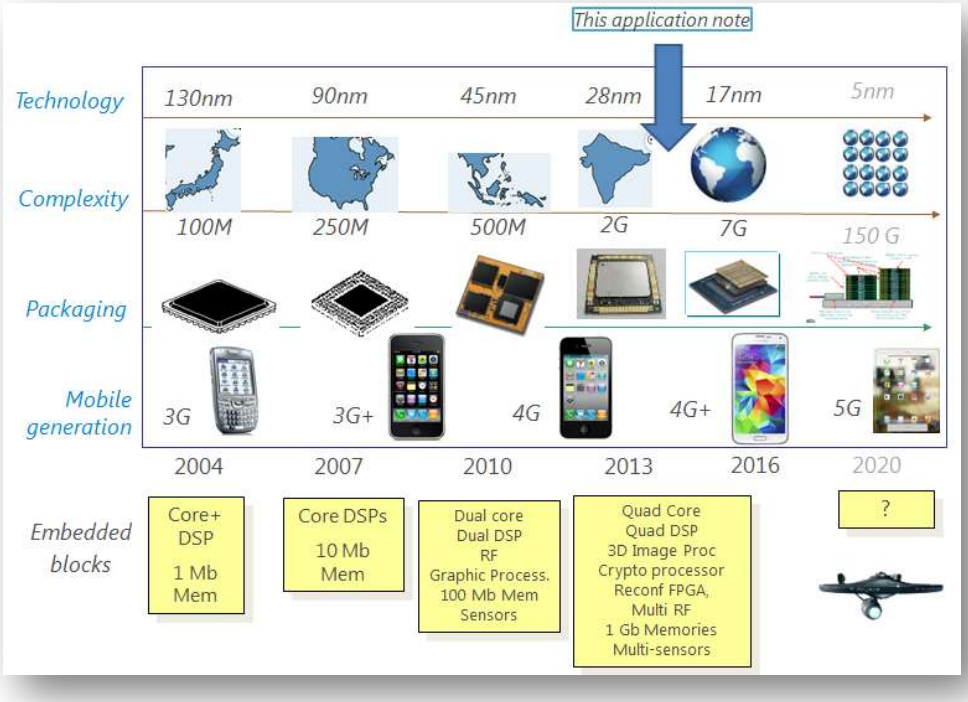


Figure 3: The 20-nm technology is targeted to mobile applications

Samsung and Apple have been the first major players in the mobile phone business to supply 20nm-technology based platforms: Exynos 5430 for Samsung, and A8 processor from Apple. The 20nm shrink reduces power consumption and extend battery life, in part to compensate for the additional power required to support the 4G+ novel standard. The Apple A8 contains 2 billion transistors, that is double the number in the A7 chip, with a die area less than 100 mm2, to improving CPU performance (basic computing) and speeding up GPU performance (enhanced graphics, higher resolution). In 2020, the 14nm/7nm generation with extensive use of FinFET technology should be used to develop mobile applications based on the 5G standard.

3. Power, performance and area gains

Expected benefits

The Power, performance and area (PPA) gains are an important metric for justifying a shift from older technology nodes to new ones. Some of the key features of the 20-nm technology from the JDA [Huiling2012][Goldberg2013] are summarized in Table 2. Compared to 32/28-nm technology node [Sicard2010], the 20-nm technology offers:

- 30 % less power consumption
- 30 % increase in switching performance
- 2 times higher density

A comparison between 90nm, 45nm and 20nm technologies in terms of density and power savings is proposed in Fig. 4. The IC surface is shrinked by a factor of 20 between 90nm and 20nm nodes, while the power consumption is reduced by a factor of 5.

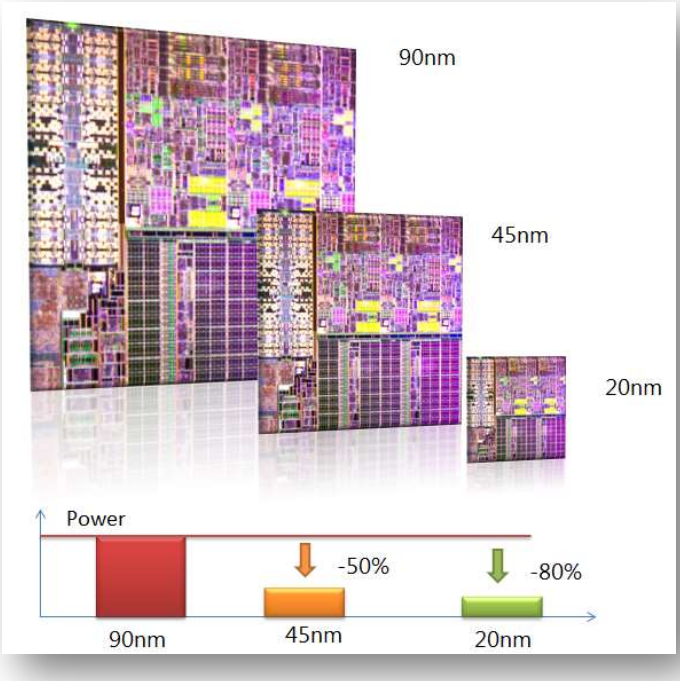


Figure 4: The 20-nm technology node as compared to 90-nm and 45-nm

Put more cores in a same area

Knowing that the silicon die size is limited by yield to approximatively 600 mm2 (3 x 2 cm) multi-core processors benefit from the downsizing of cores, as illustrated in Fig. 5. While a 2-core die approaches the maximum die size in 65nm, an octa-core processor fits in the same space with some margin in 22nm.

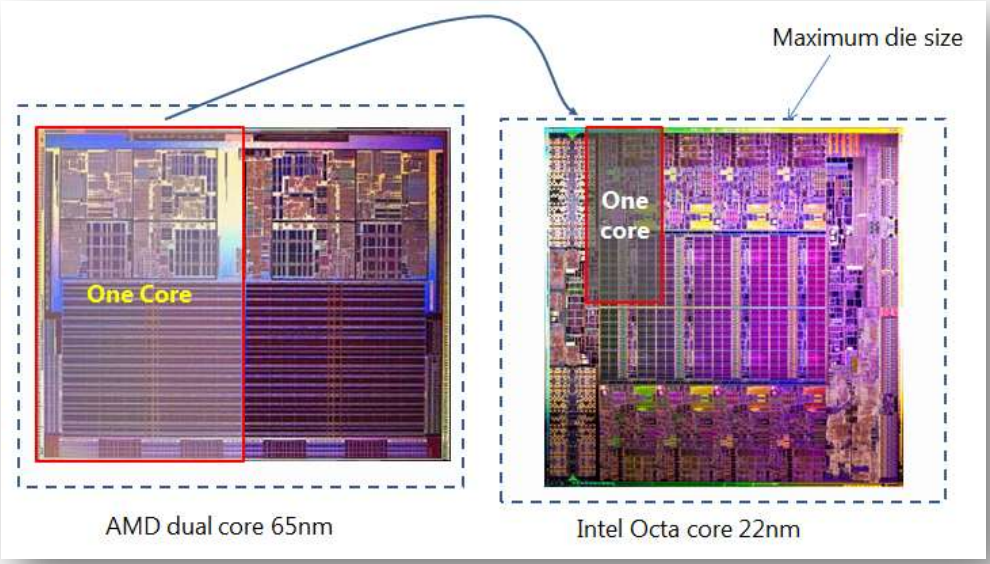


Figure 5: Multi-core architectures benefit from lithography improvements

Pass the 1.0V barrier

The supply voltage, both internal to the cores and the external I/O supply have been continuously decreased due to the thinning of the gate oxide and faster switching rates thanks to reduced voltage swings. The 20-nm technology passes the 1.0V barrier and is fixed to 0.9V, while I/Os are supplied at 1.5V (Fig. 6).

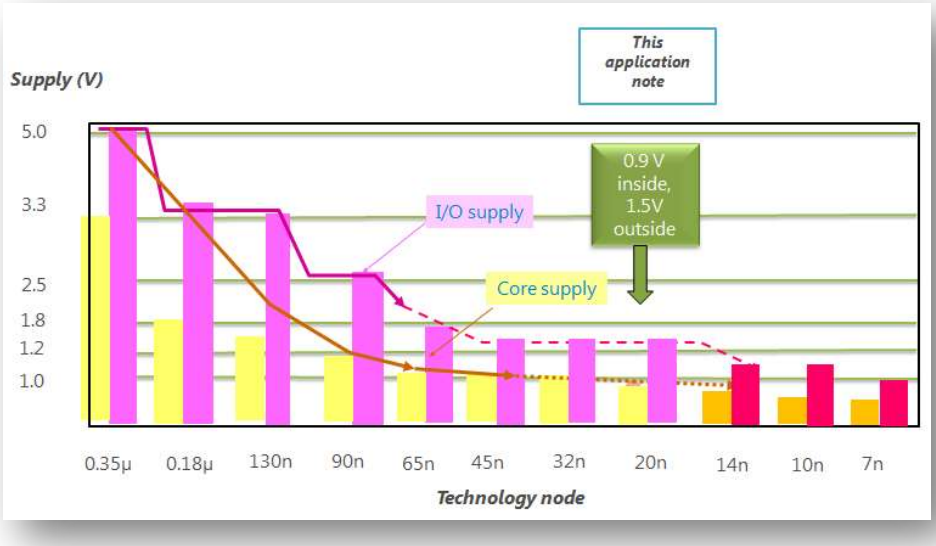


Figure 6: Passing the mythic barrier of 1V as supply voltage: 0.9V in 20-nm

4. Key features of the 20-nm technology

Lambda

In Microwind, we use an integer unit for drawing, which is fixed to 11nm. The drawn gate length is 22nm. The lower metal pitch is 64nm in [Goldberg2013], very close to 66 nm in Microwind using λ=11nm and historical 6 λ metal pitch. Fig. 7 presents the N-Channel MOS device using conservative design rules. Note that nearly 5,000 design rules exist for the 20-nm technology, of which only 100 basic rules are checked by Microwind.

Parameter	20-nm technology	In Microwind
Lambda		11 nm
Minimum gate length	20 nm	2 λ (22 nm)
Minimum gate width	60 nm	6 λ (66 nm)
Metal pitch	64	6 λ (66 nm)

Table 2: Using lambda-based design in Microwind

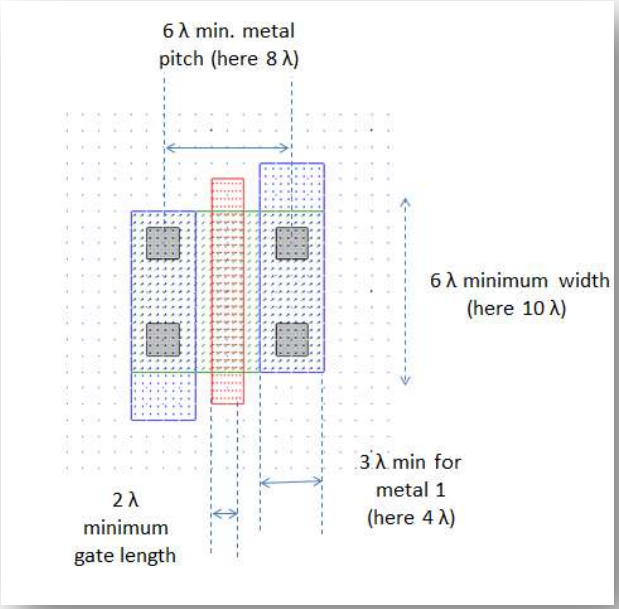


Figure 7: Example of basic n-channel MOS device in 20-nm technology. Microwind do not use minimum dimensions to ease design and routing

Core MOS devices

Table 2 gives an overview of the key parameters for the 20-nm technological node according to [Huiling2012], concerning the internal MOS devices and layers. In Microwind (Fig. 8), we select

- the "RVT" (reduced Vt) as Low leakage MOS
- the "SLVT" (super low Vt) devices as High Speed MOS.

Other MOS options are not supported in Microwind for shake of simplicity.

Parameter	Value	In Microwind
V_{DD} core (V)	0.9	0.9
Effective gate length (nm)	20	20
MOS variants	5	2
Ion N (mA/µm) at VDD	0.7-1.2	0.9 (LL) 1.1 (HS)
Ion P (mA/µm) at VDD	0.7-1.4	0.8 (LL) 1.0 (HS)
Ioff N (nA/µm)	0.06-200	1 (LL) 10 (HS)
Ioff P (nA/µm)	0.06-200	1 (LL) 10 (HS)
Gate dielectric	HfO_2	HfO_2
Gate stack	Al/TiN	Al/TiN
Equivalent oxide thickness (nm)	1	1

Table 2: Key features of the core devices proposed in the 20 nm technology

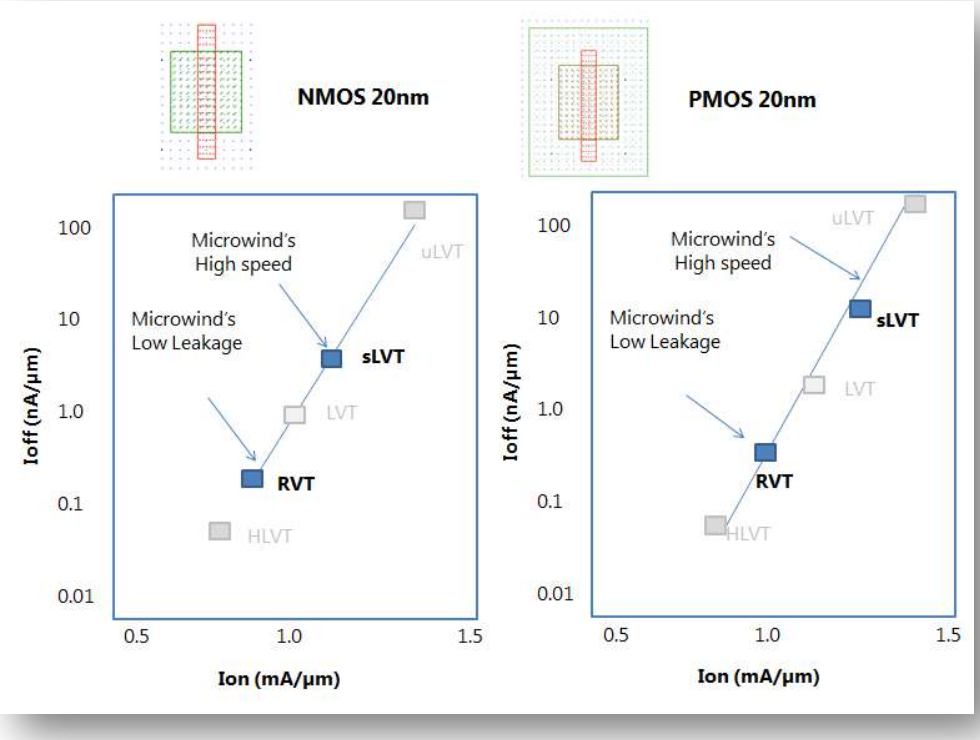


Figure 8 MOS options of the Joint Development Alliance 20-nm and Microwind targets

IO MOS devices

Table 3 gives an overview of the key parameters for the 20-nm technological node according to [Huiling2012], concerning the Input/output MOS devices and associated supply voltage. In Microwind, we only consider 1.5V I/O supply and tune the "High-Voltage" (HV) MOS device on the median performances.

Parameter	Value	High Voltage (HV) MOS in Microwind
VDD IOs (V)	1.2, 1.5 or 1.8	1.5
Effective gate length (nm)	70-150	100
Ion N (mA/µm)		0.3
IonP (mA/µm)		0.22
Ioff N (nA/µm)	0.013-6	0.1
Ioff P (nA/µm)	**0.003-2**	**0.1**

Table 3: Key features of the I/O devices proposed in the 20 nm technology and corresponding values in Microwind

Figure 9: Vertical cross-section of N-P MOS devices with associated metal layers

5. Transistor performances in 20-nm technology

Key features

In Microwind, only three types of MOS devices (3 nMOS, 3 pMOS, 6 MOS devices in total) exist in the 20-nm technology :

- the “low-leakage - Reduced Vt ” (RVT) MOS is the default MOS device, with reasonable leakage (6nA/µm). The main objective of this MOS device is to reduce the *Ioff* current, that is the parasitic current that flows between drain and source with a zero gate voltage.
- the “high-speed - Super Low Vt” (SLVT) MOS has higher switching performance, thanks to a shorter effective channel length, at the price of a leakage multiplied by 10 (60nA). The designer may choose this high-speed MOS device for cells for which speed is the critical point, at the price of an important leakage current.
- the high voltage MOS used for input/output interfacing. In Microwind’s cmos20nm rule file, the I/O supply is 1.5 V.

N-channel MOS device characteristics

The 20-nm technology uses a stack of high-k dielectric and metal gate (HKMG). As mentioned in [Huiling2012], it consists of a HfO2 oxide combined with TiN-based “WorkFunction Metal”, with a central part filled with Aluminum (Al). The process is called Replacement Metal Gate (RMG), as a first gate is constructed at early stages of MOS lithography, and later removed and replaced by the final gate (Figs. 10 & 11).

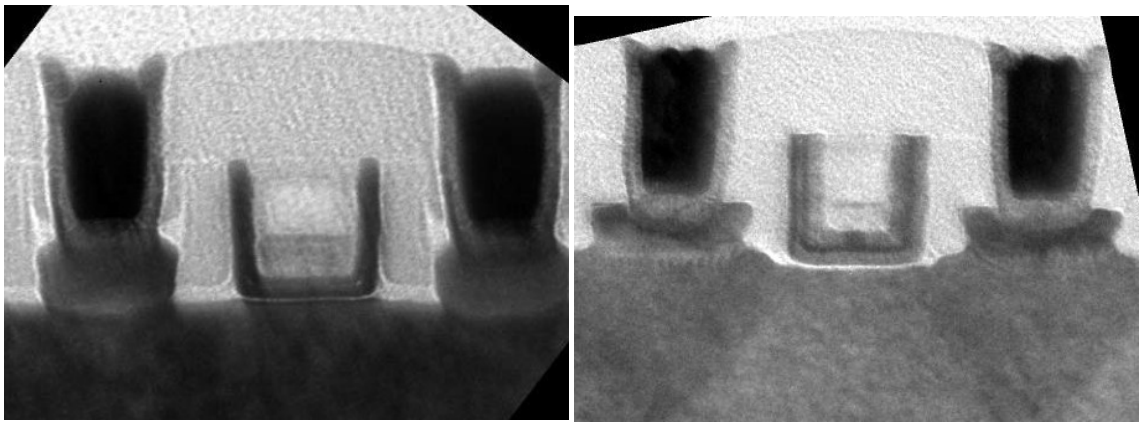


Figure 10: Representative cross-sectional of the n-channel MOS (left) and p-channel MOS (right) [Huiling2012, Fig. 3]

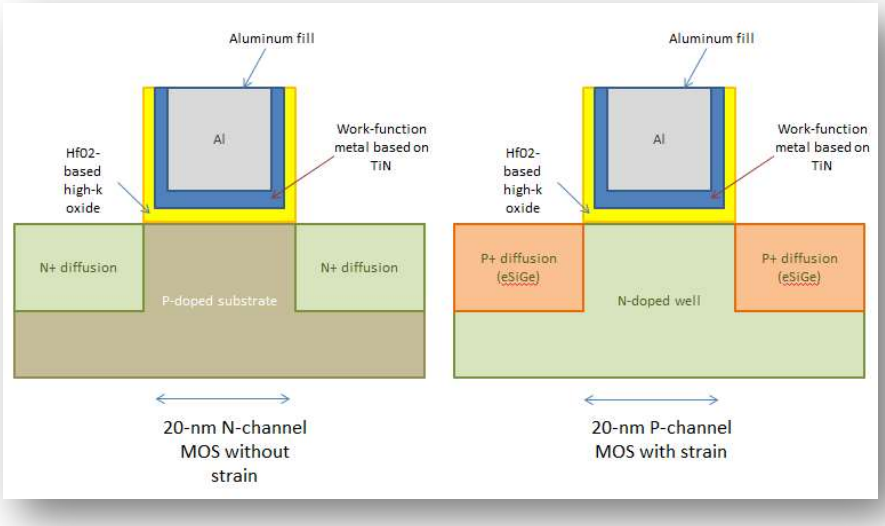
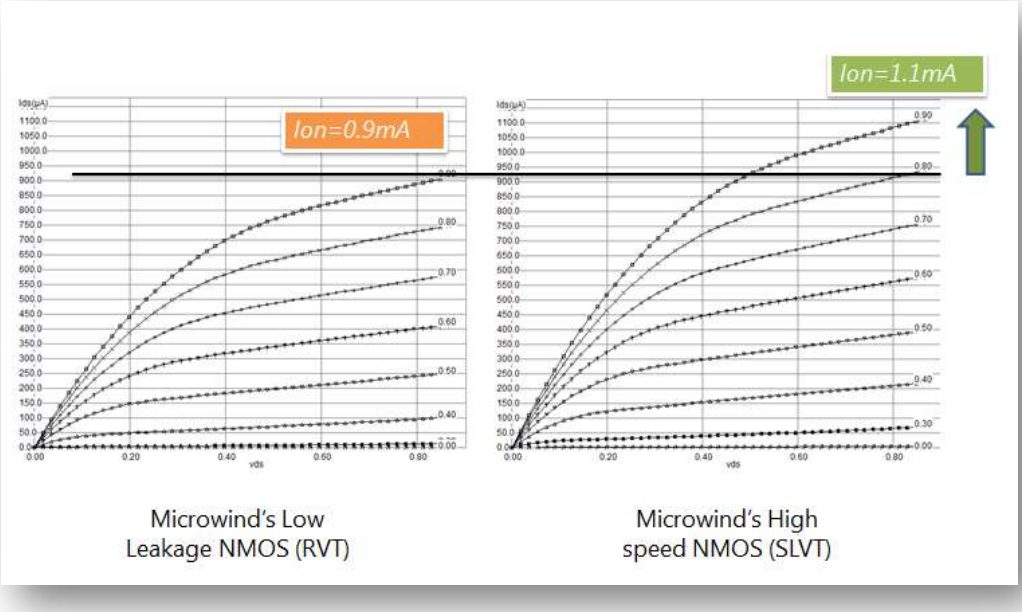
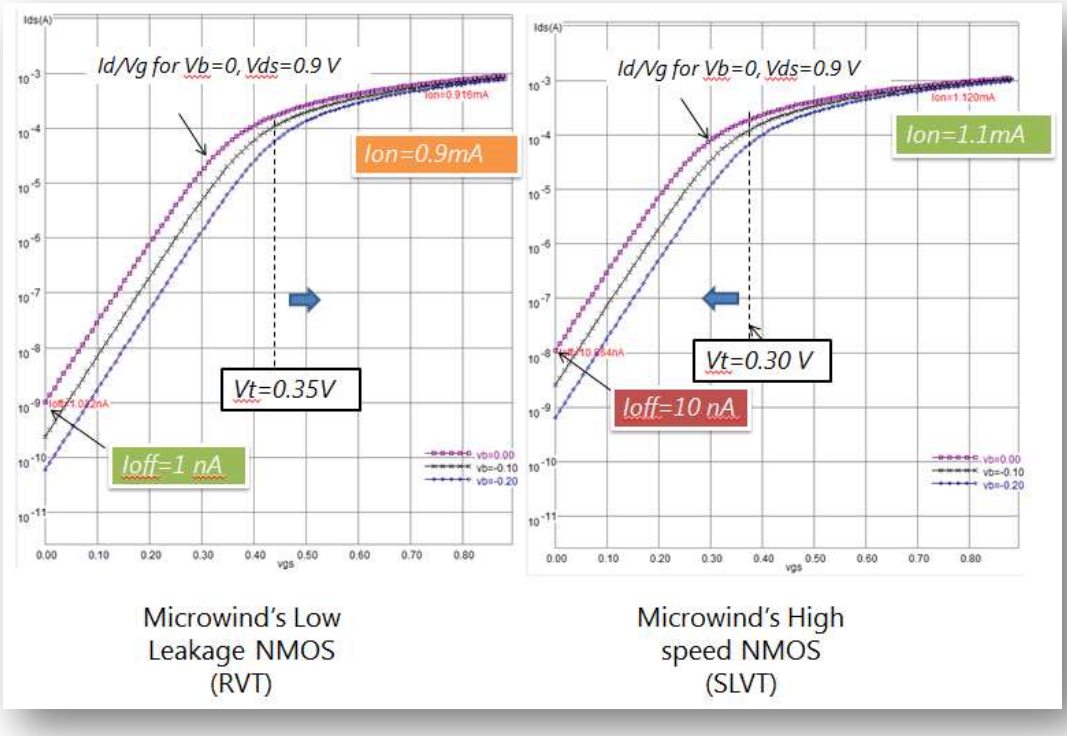


Figure 10: 2D cross-section of the N-channel MOS and P-channel devices [from Sato, 2013]



(a) Low leakage W=1µm, L= 20nm *(b) High speed W=1µm, L=20nm*

Figure 11: Id/Vd characteristics of the low leakage and high speed nMOS devices.



(a) low leakage W=1µm, L= 20nm *(b) high speed MOS W=1µm, L= 20nm*

Figure 12: Id/Vg characteristics (log scale) of the low leakage and high-speed nMOS devices

The I/V characteristics of the low-leakage and high-speed MOS devices (Figs. 11 and 12) are obtained using the MOS model BSIM4 (See [Sicard2007] for more information about this model). The I/V characteristics reported in Fig. 10 demonstrate that the low-leakage NMOS has a drive current capability of around 0.9 mA for W=1.0 µm at a voltage supply of 0.9V. For the high speed NMOS, the drive current rises to 1.1 mA/µm.

The drawback associated with this high current drive is the leakage current which rises from 1 nA/µm (low leakage NMOS) to 10 nA/µm (high speed NMOS), as seen in the Id/Vg curve at the X axis location corresponding to Vg= 0 V (Fig. 11).

From a design view-point, the "option" menu in the MOS generator enables to switch from low leakage to high-speed. In terms of layout, the only difference is the option layer that contains the MOS option information (Fig. 13).

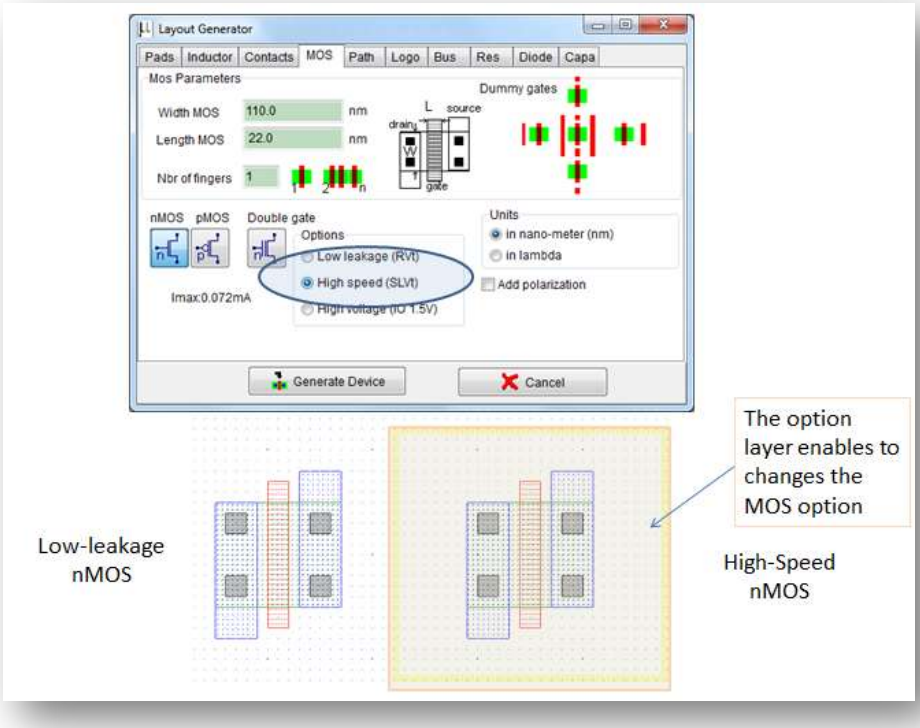


Figure 13: Changing the option of the MOS device from low-leakage to high-speed using the option layer. Double-click in one corner of the option layer to change its properties

P-channel MOS device characteristics

The PMOS drive current in CMOS 20-nm technology is around 0.8mA/µm for the low-leakage MOS and up to 1.0 mA/µm for the high-speed MOS. Note that NMOS and PMOS performances are quite comparable (Fig. 14), thanks to the eSiGe strain engineering for PMOS that nearly compensates the intrinsic mobility degradation of holes (P-channel) vs. electrons (N-channel). The leakage current is around 1 nA/µm for the low-leakage MOS and nearly 10 nA/µm for the high-speed device (Fig. 15).

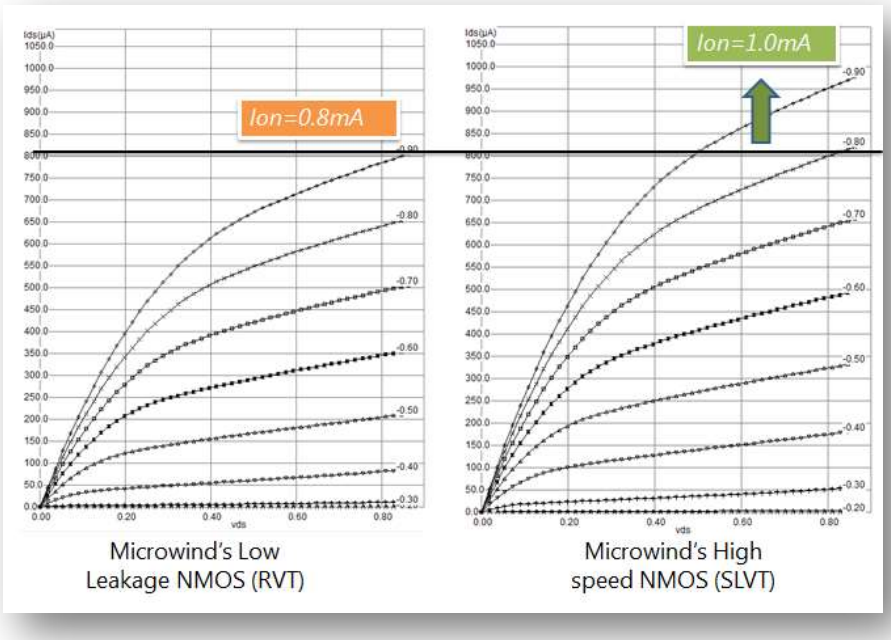


Figure 14: Id/Vg characteristics (log scale) of the low leakage and high-speed pMOS devices

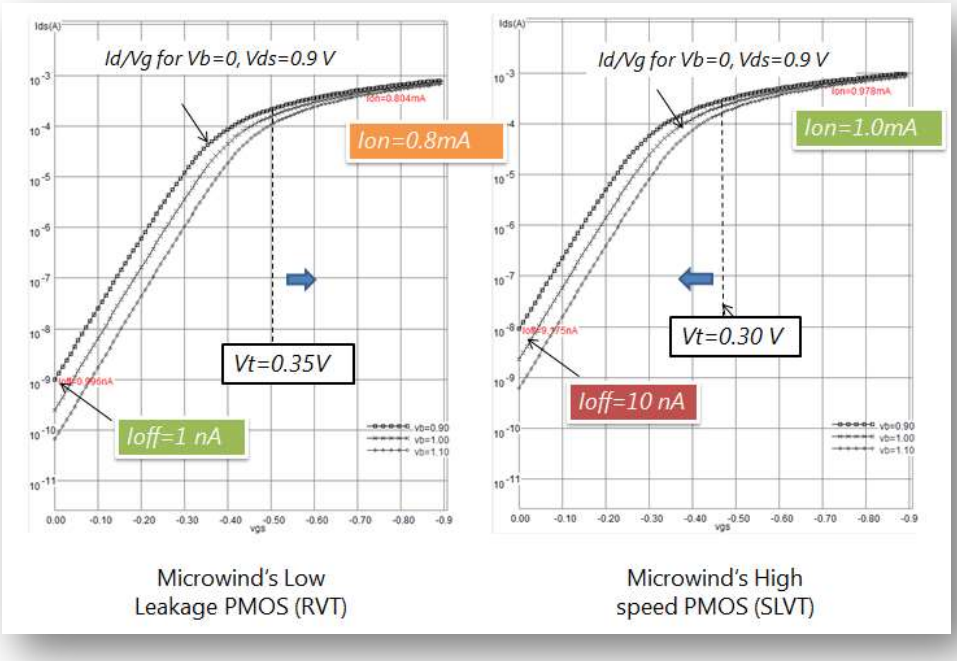


Figure 15: Id/Vg characteristics (log scale) of the low leakage and high-speed pMOS devices

6. MOS Design for Manufacturing

Process Variability

One important challenge in nano-CMOS technology is process variability. The fabrication of millions of MOS devices at nano-scale induces a spreading in switching performances in the same IC.
The effect of process variability on the MOS Ioff/Ion characteristics is plotted using the menu "Ioff vs. Ion" under the "MOS I/V curve" menu (Fig. 16). It can be seen that the MOS devices have a wide variability in performances. The 3 MOS types (low leakage, high speed, high voltage) are situated in well defined space in the Ioff/Ion domain. The low leakage is in the middle (medium Ion, low Ioff), the high speed on the upper right corner (high Ion, high Ioff), and the high voltage is at the lower left side of the graphics (low *Ion*, very low *Ioff*). Note that the exact locations of the dots will change for each MOS characteristics plotted because it is a random process.

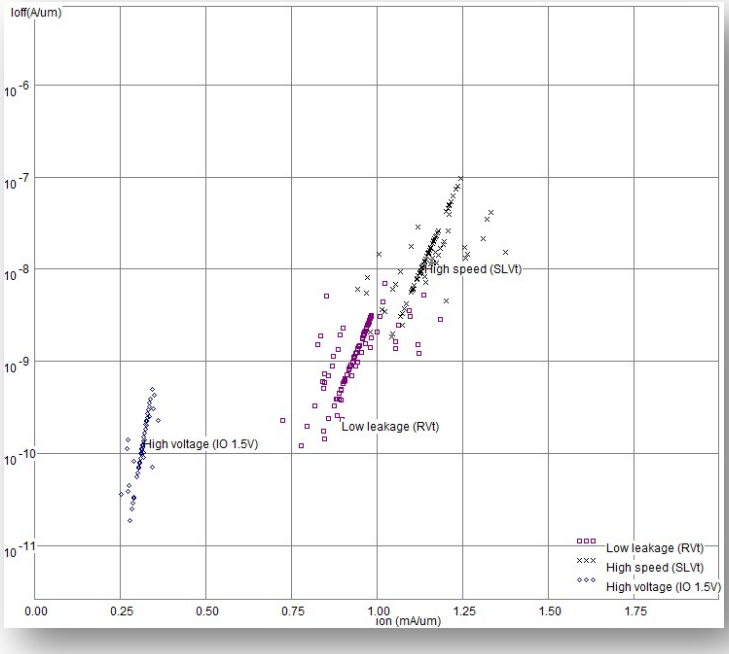


Figure 16: Ioff/Ion calculated by Microwind on 100 samples of n-channel MOS with random distribution of VT, U0, and LINT with a Gaussian distribution around the nominal value

Dummy gates

One solution to reduce MOS performance variability is to design in a regular way the MOS gates. A new option has been inserted in the MOS generator to add "dummy" gates around the active device (Fig. 17). The variability of lithography depending on the environment will significantly affect the dummy gate, but only little the central active gate, as illustrated in Fig. 18

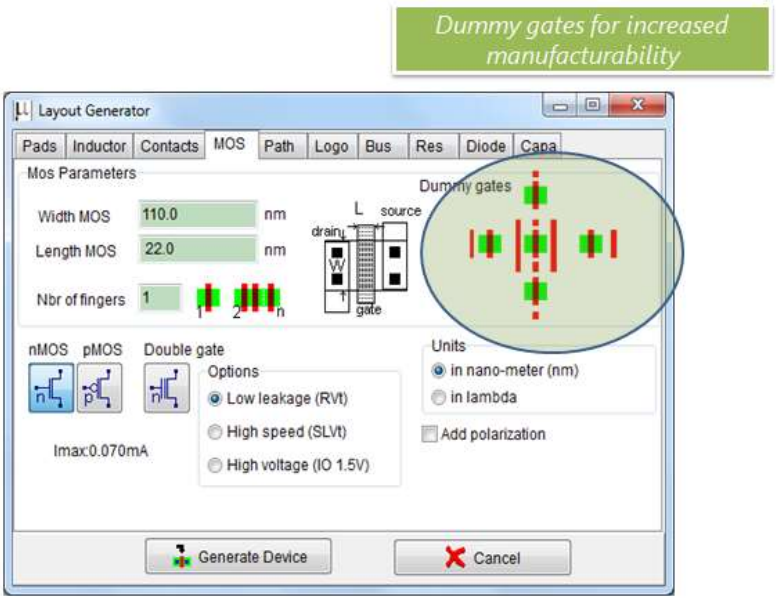


Figure 17: The layout generation includes the dummy gate option at any side (or all sides) of the active MOS device to reduce variability and increase manufacturability

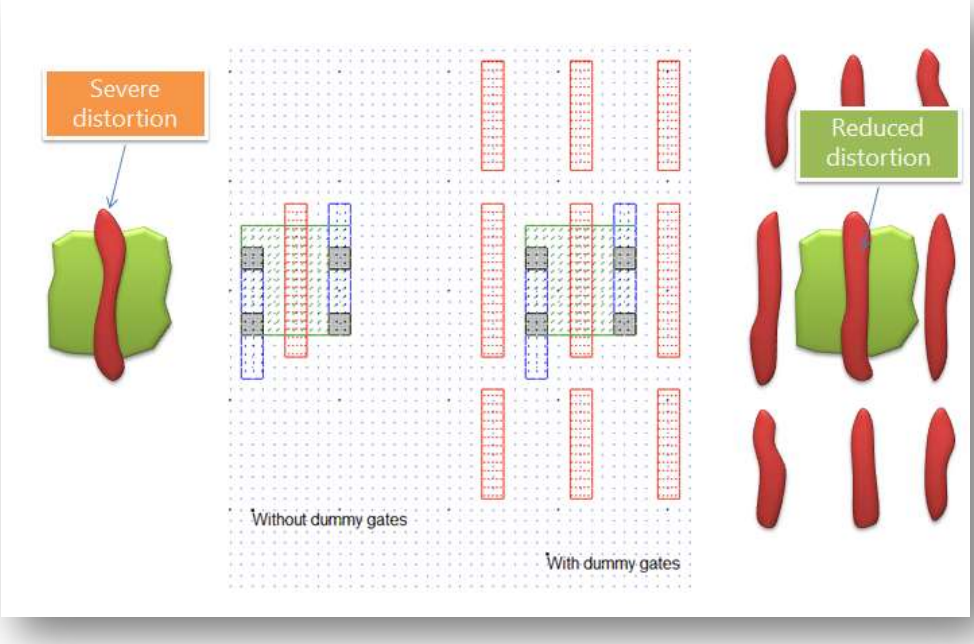


Figure 18 : The sub 40-nm lithography induces severe pattern distortions that may be reduced thanks to dummy components all around the sensitive area (20nm\nmos-dummy-gates.msk)

7. Interconnects

Metal Layers

As seen in table 4, the original 20-nm technology comprises 11 layers, 3 lower ones for short routing, 4 for long routing, 2 for local supply and 2 very thick layers for system supply. In Microwind, we reassigned the *metal1* to *metal8* according to the table 4:

- M1 & M2 are at 6 λ pitch for local routing
- M3 & M4 are at 8 λ pitch for medium routing
- M5 & M6 are at 32 λ pitch for long routing and local supply
- M7 & M8 are dedicated to general supply

Parameter	Pitch (nm)	Thickness (nm)	Pitch in Microwind	Purpose
Middle-of-the-Line (MOL)	64	50	Not supported	Intra-cell routing
M1-M3	64	68	M1-M2: 6 λ (66 nm)	Short routing
M4-M7	80	80	M3-M4: 8 λ (88 nm)	Medium routing
M8-M9	358	150	M5-M6: 32 λ	Block supply and long routing
M10-M11	1000	200	M7-M8: 92 λ	System supply and IO routing
Interconnect layer permittivity K	2.5-2.7			

Table 4: Key features of interconnects available in the 20 nm technology and corresponding values in Microwind

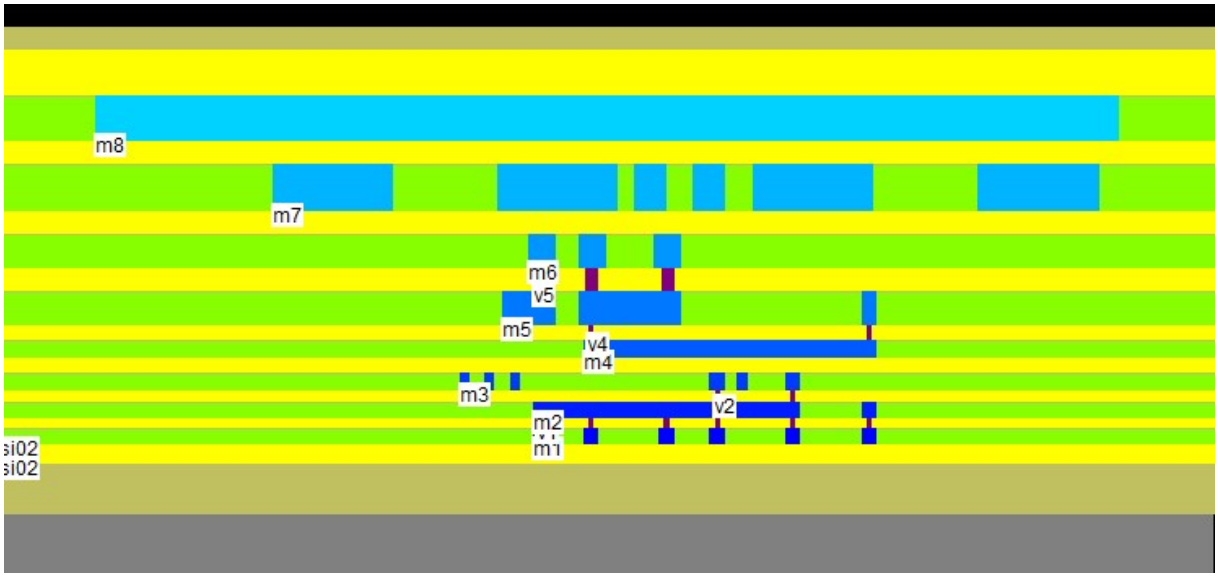


Figure 19: The 8 metal layers of the Microwind implementation of the 20-nm, originally with 11 layers

Layers *metal5* and *metal6* are a little thicker and wider, while layers *metal7* and *metal8* are significantly thicker and wider, to drive high currents for power supplies (Fig. 19).

Interconnect Resistance

At minimum width, the interconnect resistance of the M1-M2 lower metal layers is around 12 Ω/µm (Fig. 20). Metal layers 2 to 4 have relaxed design rules, meaning lower resistance (7 Ω/µm).

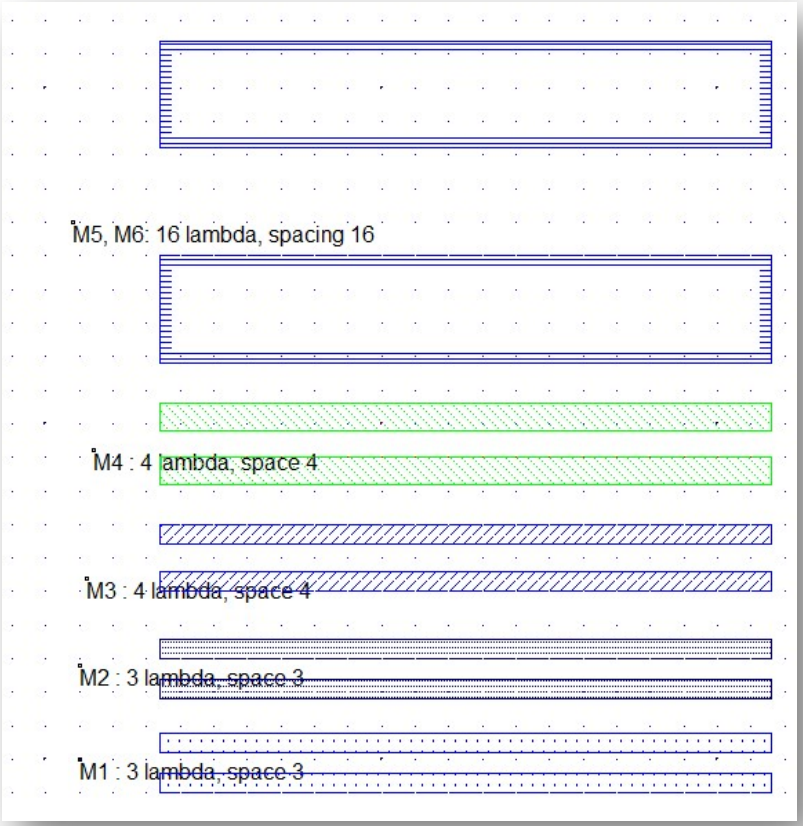


Figure 20: Computing the wire resistance using 1µm length metal tracks at minimum width and spacing (20nm\metalM1-M4.MSK)

Double Patterning for M1-M2 Interconnects

The double patterning is requires for M1 and M2 as the pitch between tracks is smaller than 80nm. Half the patterns go on the first patterning and half go on the second patterning, as illustrated in Fig. 21. In order to ensure an easy selection of metal tracks for the first and second patterning, regular structures with straightforward orientation such as M1 east-west, and M2 south-north are requested (Fig. 22). The other solution is to relax the pitch constraints for an improved manufacturability, at the cost of an extended silicon area. M3-M8 with pitch of 80-nm and higher are still fabricated using simple patterning.

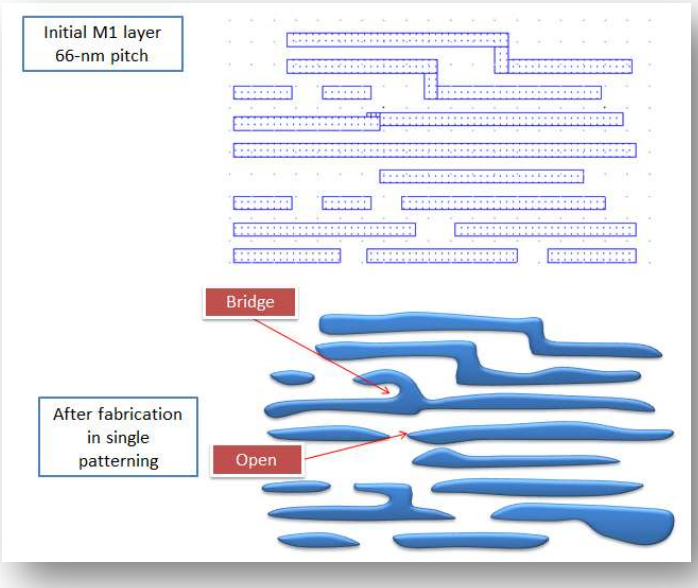


Figure 21: Simple patterning of M1 and M2 may lead to bridges and shorts

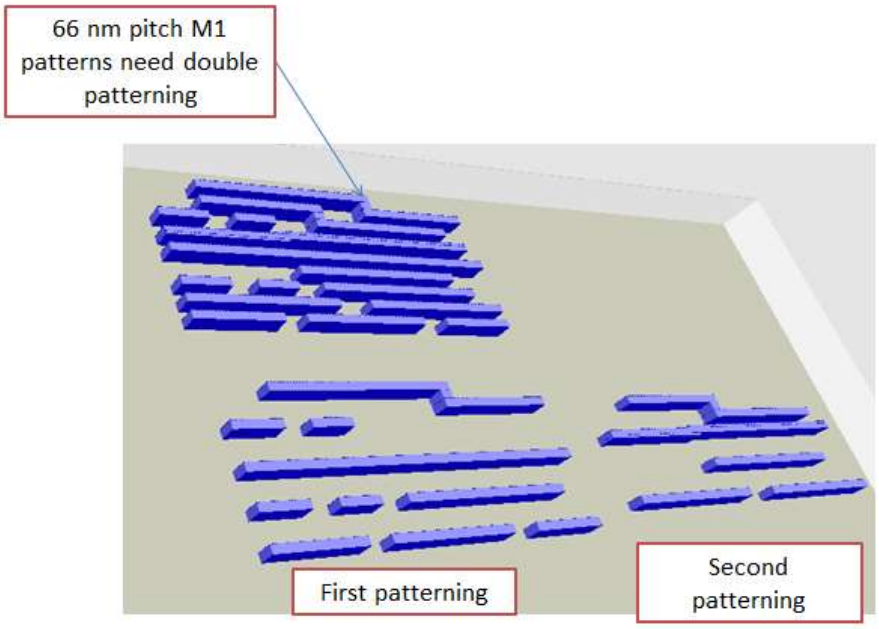


Figure 22: Principles of double patterning (20nm\double-patterning-m1.MSK)

8. Ring Inverter Simulation

The ring oscillator made from 5,7,9,etc.. inverters, has the property of oscillating naturally. The layout of one stage is reported in Fig. 23. It corresponds to the design proposed by [Scholze2011] based on a fanout of 3, meaning that the equivalent capacitance of 3 gates is connected to each node. Two inverters are in parallel for an improved current drive.

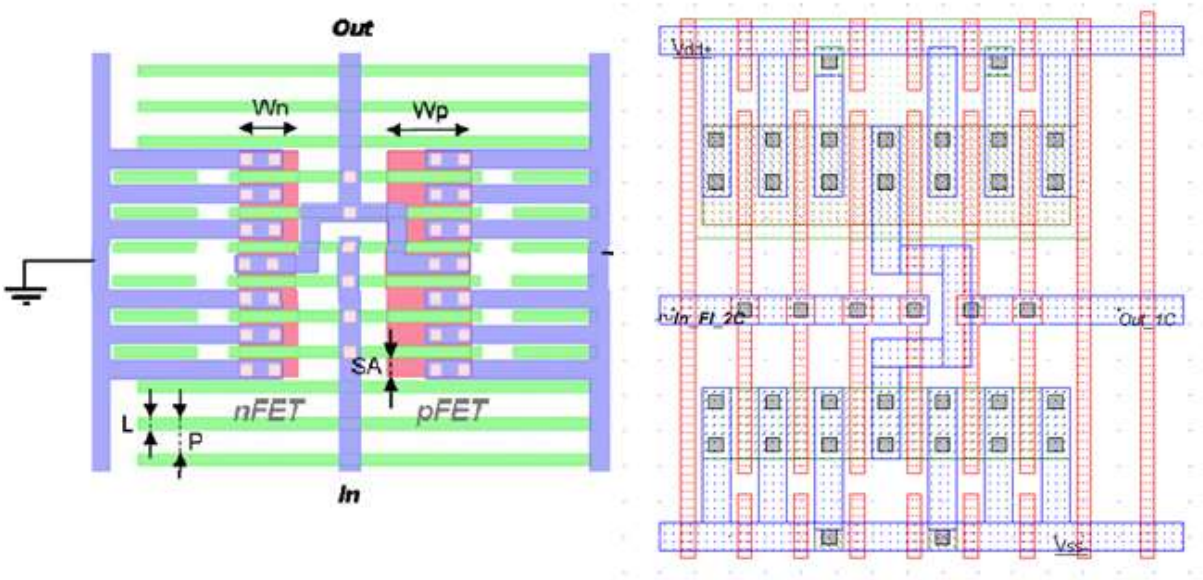


Figure 23: Redesigning a 20-nm inverter based on the layout proposed by [Scholze2011] (20nm\invStage_FO3_Scholze.MSK)

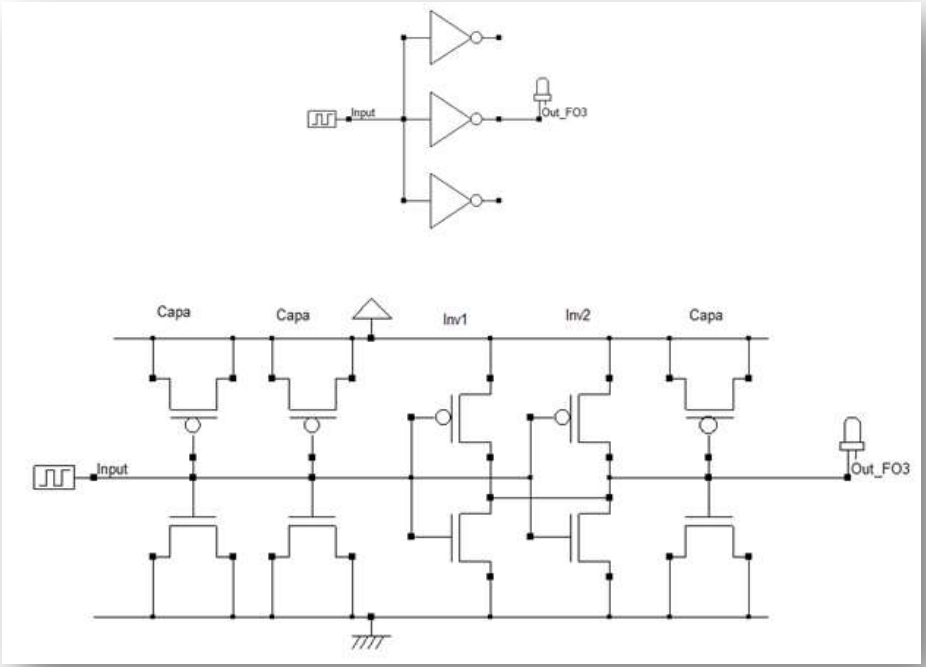


Figure 24: Principles of the Inverter with Fanout 3 using capa cells as loads (2 on the input, one on the output). Two inverters in parallel correspond to one single inverter with double current drive

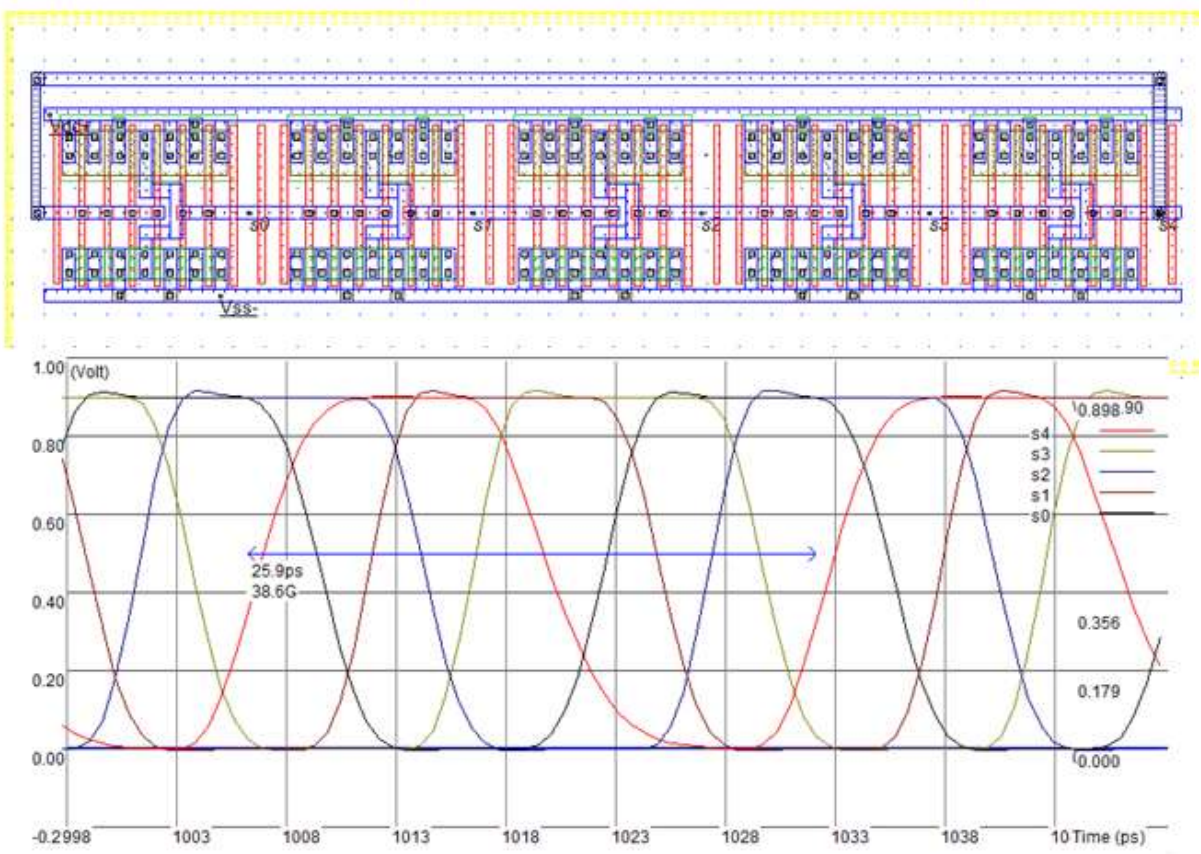
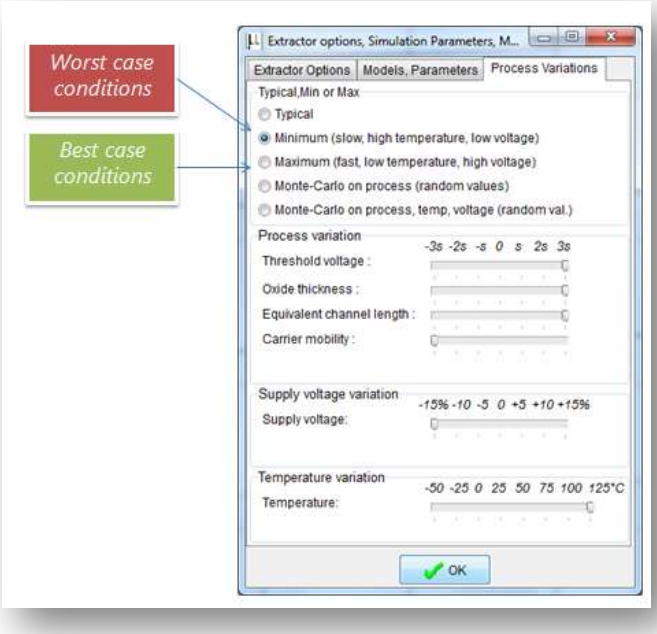


Figure 25: The simulation shows a 26 ps cycle period, that is approx.. 5ps/stage with Fanout 3, which fits with [Scholze2011] (20nm\RingOsc_5Stage_FO3_Scholze.MSK)

Five oscillators are chained (Fig. 25) to create a free oscillation at a rate approaching 40 GHz (aound 5 ps/stage) for a ”High-speed” option, very close to the results published in [Scholze2011]. The FO3 (Fanout3) oscillating frequency is lower than for FO1 (Fanout 1, no extra load capacitance), but closer to real case situation where gates are usually connected to more than one single gate, with interconnects of significant length.

Simulation of Process Variations (PVT)

Microwind gives access to “Process-Votage-Temperature” (PVT) simulation through the command **Simulate → Simulation Parameters → Process Variations**. Direct access from the simulation waveform window is also possible using the button **“Process Var.”.** The most usual simulation consists of simulating extreme situations (Min and Max), as compared to typical conditions, as shown in Fig. 26. More than 300% variation is observed between the oscillating frequency under Min and Max situations (20 to 60 GHz).



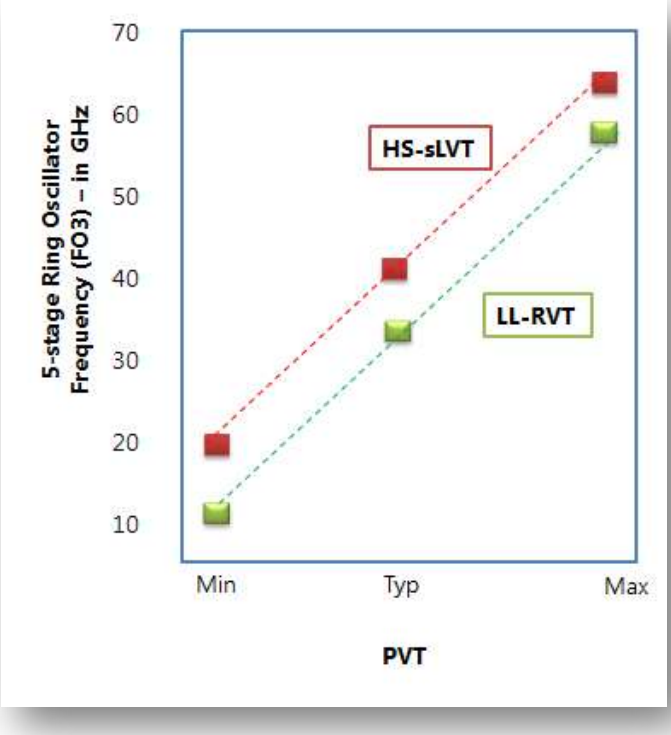


Figure 26: PVT Min-Max simulation showing 300 % variation in oscillating frequency of the ring-oscillator (20nm\RingOsc_5Stage_FO3_Scholze.MSK)

Notice that

- In Min situation, VT is high, mobility U0 is low and the channel is long (LINT>0). The supply is minimum and the temperature is maximum.
- In Max situation, VT is low, mobility U0 is high and the channel is short (LINT<0). The supply is maximum and the temperature is minimum.

Parameter class	Parameter	Symbol (BSIM4)	Unit	Min (20-nm)	Typ (32-nm)	Max (20-nm)
Process	Threshold Voltage	VT	V	0.30	0.25	0.20
	Mobility	U0	$cm.V^{-2}$	500	600	700
	Channel length reduction	LINT	m	$1e^{-9}$	0	$-1e^{-9}$
Voltage	Supply	VDD	V	0.77	0.9	1.03
Temperature	Temperature	TEMP	°C	125	27	-50

Table 5: Variation of process parameters

9. 6-transistor static RAM

One of the most representative designs for comparing technology nodes is the static RAM cell designed using 6 transistors (6T-SRAM). In our implementation in Microwind (see Fig. 27), the layout size is 407 x 176 nm, with a surface area of 0.072 µm^2, which is close to [Huiling2012] 0.08 µm^2. Note that the layout strictly obeys the basic design rules. Most contacts are shared with neighboring cells: the VSS, VDD contacts, the Select and Data lines. It is usual to find more aggressive layout design rules in RAM cell designs, in order to further decrease the cell area.

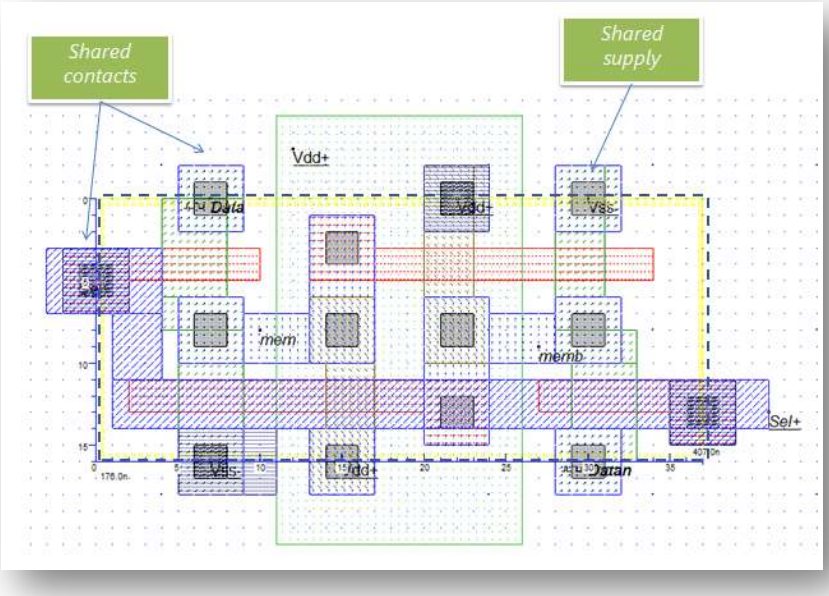


Figure 27: The 6-transistor RAM layout using 20-nm design rules (20nm\SRam_6T_20nm.MSK)

10.Conclusions

This application note has illustrated the trends in CMOS technology and introduced the 20-nm technology generation, based on technology information available from the Joint Development Alliance (JDA) regrouping major integrated circuit manufacturers. The key features of the 20-nm CMOS technology have been illustrated, including the N-channel and P-channel MOS device characteristics, design for manufacturing and double patterning. A 5-stage ring has been used for calibration purpose. Future work will concern the 14-nm technology node which introduces the FinFET device and extend massively the concept of design for manufacturability.

References

[Sicard2005] E. Sicard, "Introducing 90-nm technology in Microwind", software application note, August 2005, www.microwind.org

[Sicard2006] E. Sicard and S. M. Aziz, "Introducing 65-nm technology in Microwind3", software application note, July 2006, www.microwind.org

[Sicard2007] E. Sicard, S. Ben Dhia "Basic CMOS cell design", McGraw Hill, 450 pages, international edition 2007 – ISBN 9780071488396

[Sicard2008] E. Sicard and S. M. Aziz, "Introducing 45-nm technology in Microwind3", software application note, July 2008, www.microwind.org

[Sicard2009] E. Sicard "Microwind User's Manual, lite version 3.5", INSA editor, 2009, www.microwind.org.

[Sicard2010] E. Sicard and S. M. Aziz, "Introducing 32-nm technology in Microwind35", software application note, July 2010, www.microwind.org

[Goldberg2013] Goldberg, "64nm Pitch Interconnects: Optimized for Designability, Manufacturability and Extendibility", 2013 Symposium on VLSI Technology Digest of Technical Papers

[Huiling2012] Huiling Shang, , "High Performance Bulk Planar 20nm CMOS Technology for Low Power Mobile Applications", 2012 Symposium on VLSI Technology Digest of Technical Papers

[Sato2013] F. Sato, "Process and Local Layout Effect interaction on a high performance planar 20nm CMOS", 2013 Symposium on VLSI Technology Digest of Technical Papers

[Scholze2011] A. Scholze, "Exploring MOL Design Options for a 20nm CMOS Technology using TCAD", SISPAD 2011