SOI 90-nm Ring Oscillator Sub-ps Model-Hardware Correlation and Parasitic-aware Optimization Leading to 1.94-ps Switching Delay

Jean-Olivier Plouchart, Jonghae Kim, Blaine Jeffrey Gross, Kun Wu , Robert Trzcinski, Victor Karam[1], Paul Hyde, Richard Williams, Myung-Hee Na, Judith Mc Cullen, and William Clark

IBM Semiconductor Research and Development Center, Hopewell Junction, NY12533 USA
[1]Carleton University, Department of Electronics, Ottawa, Ontario, Canada K1S 5B6
Phone: +1-845-892-9323, e-mail: plouchar@us.ibm.com

Abstract

This paper reports the SOI 90 nm statistical model to hardware correlation achieved over a broad voltage, temperature and a variety of five different ring oscillators. Monte Carlo simulations were performed and compared with the measured circuit statistical population. A sub-ps Model to Hardware Correlation accuracy was achieved between the mean hardware and simulated delays. Based on the model validation, a parasitic aware layout optimization was performed on a constrained and an unconstrained inverter leading to a 5 and 66 % reduction in delay of the inverter reference circuit. The unconstrained parasitic aware optimization achieves a record inverter switching delay of 1.94 ps.

Introduction

Electrical models and circuit level Model and Hardware Correlation (MHC) are very challenging for nanometer semi-conductor technologies. For example, the 2004 Modeling and Simulation ITRS roadmap recommends better than 5% on the circuit delay accuracy. For the advanced nanometer technologies unloaded ring oscillators delay below 10 ps are currently in production (1), therefore 5% accuracy is equivalent to better than 500 fs accuracy on the delay. This is extremely demanding on the device and interconnections models accuracy. For the future technologies nominal unloaded ring delay of 5 ps will be achieved, which is equivalent in the frequency domain to a 100 GHz bandwidth (0.5/delay) for the switching stage (2). We are clearly moving from RF digital to mm-wave digital regime. Also, because of the broad variations affecting nanometer technologies, statistical analysis is the only viable approach. Therefore, not only the mean but also the variations must be predicted accurately. Statistical Model and Hardware Correlation (MHC) accuracy is also required for parasitic aware optimization, and Chip Limited Yield improvement. This paper presents the challenges, the MHC and the parasitic aware optimization results achieved for our 90 nm SOI technology.

90nm SOI Hardware and Model Correlation Challenges

In order to achieve sub-ps MHC, near FET wiring must be extracted with good accuracy. Figure 1 shows the technology cross-section as well as the near FET wiring parasitic. The extrinsic parameters C_1, C_2, C_3 and R_1 and R_2 are critical parameter for high-speed design. For example, figure 2 shows the simulated variation of the loaded ring delay as a function of the gate (PC) to the local interconnect (MC) capacitor. A sensitivity as high as 14 ps/fF is simulated for a typical 70 aF/um MC-PC capacitor. If the MC-PC capacitor is ignored during the parasitic extraction a 6% faster delay is simulated. The extrinsic $C_{gs,ext}$ (C_1) and miller capacitor $C_{gd,ext}$ (C_2) (fig 1) can reduce significantly the inverter performance. This sensitivity analysis demonstrates also the importance of wire to wire as well as via to via or wire capacitor extraction. Another challenge shown in figure 3 is that the layout representation seen by designers is less and less representative of what is printed on the wafer. Therefore the GDS as the primary source to predict circuit performance is less and less reliable. For example each individual metal wire fill were extracted individually using Calibre xRC (fig 3). A 0.7% (100 fs) and 0.9 % (78 fs) added delay was simulated due to the added capacitor from the metal 1 and 2 fill for a loaded and unloaded ring oscillator, respectively. Another level of complexity specific to SOI is the initial body potential for the floating FETs impact on switching delay. As shown in table I, the simulated delay increases with time for the initial body potential resulting from the DC solution. After several micro-seconds of simulation, the delay will converge to the steady state. The steady state must be determined accurately because this is what is measured by our testers. In order to find the closed loop ring oscillator steady state, the SpectreRF Periodic Steady State analysis for oscillators was performed. The simulation results of the PSS analysis indicate an unloaded ring delay of 7.24 ps, which is 3.6 % slower than the 6.99 ps simulated with the first cycles of the regular transient simulation (DC solution body potential). Another way to initialize the body potential to the steady state is to perform an indirect body initialization by initializing the inverter gates to half VDD. This is a simpler approach than initializing the body of each FET to the steady state potential. As shown in table one, this technique leads to the same result than the PSS analysis.

Results of Ring- Oscillators Monte-Carlo Simulations and Hardware Correlation

A loaded ring oscillator delay and current consumption were measured on a wafer for 60 different chips at different voltages and temperature. The ring oscillator layout was extracted using wire and via resistor extraction, as well as wire to wire and wire to substrate parasitic extraction. The Monte Carlo simulations were performed on the full ring. One of the very unique features of our Erie internal parasitic extraction tool is that it allows Monte Carlo simulation of not only the FETs but the wiring parasitic as well. Therefore the wiring parasitic distribution across the wafer can be taken into account. Figure 4 (a) shows the measured and simulated Monte Carlo active current distribution. The model and hardware results overlap demonstrate that the active current variations can be modeled accurately. Figure 4 (b) shows also some good MHC agreement for a large number of chips of the leakage and delay distribution. Because of the exponential dependency of leakages, this distribution is very difficult to correlate. Figure 4 (c) shows that excellent MHC is achieved for ring oscillator effective capacitor defined by Ceff=2*(Idda-Iddq)*delay/VDD. Figure 5 shows the hardware and simulation Gaussian delay distribution at 1.1V. An excellent mean MHC is achieved. The simulated standard deviation is wider than for the hardware, because the Monte Carlo parameters were allowed to vary broadly. Figure 4 (d) and (e) show the variation at 25C of the simulated and measured mean across a very broad voltage of 0.4 to 1.3 V for the loaded inverter stage delay and active current consumption. Excellent mean MHC is achieved between 0.7 and 1.1 V. Figure 4 (f) shows the variation of the mean Ceff across voltage and the good MHC achieved. From 0.4 to 0.6 V Ceff increases because the gate capacitor goes into inversion. From 1 to 1.3 V Ceff increases again rapidly, because of the fast increased in the shoot through current. Figure 6 (a) and (b) are 3-D graph of the number of chips versus active current consumption and loaded ring delay for the Monte Carlo simulations and the measured chip statistical population, respectively. This graph shows that most of the chips are concentrated on a delay area of ±1 ps and that for a given delay there is a significant spread of the current or power consumption. Inversely, for a given power consumption, a large delay spread can be measured and simulated. Therefore different set of electrical parameters can generate a similar measured parameter. Figure 7 (a) and (b) show the variation of the ratio of the simulated

0-7803-9269-8/05/$20.00 (c) 2005 IEEE

and measured mean delay for a Fan Out of 3, a NAND, a NOR, a loaded inverter and a body contacted inverter ring oscillators, at 85C and 25C, respectively. Across a broad voltage (0.6-1.2 V), temperature, and a variety of rings a maximum MHC discrepancy of +5.9 to -5.1 % is achieved. This is equivalent to a sub-ps MHC delay accuracy for the FO3, NAND, NOR, over a voltage of 0.7 to 1.3V at 25 and 85C. Due to limited simulation data up to 1.1V, the sub-ps MHC delay accuracy is achieved from 0.7 to 1.1V for the loaded inverter. For the body contacted inverter the sub-ps MHC delay accuracy is achieved from 0.9 to 1.3V, because the body contacted FET is more difficult to model.

Parasitic Aware Layout Optimization for Switching Delay Reduction

Once all the electrical models and the simulation flow are validated by MHC, it is possible to use a parasitic aware layout optimization to reduce parasitic and reduce switching delay. For example as shown in figure 8, two inverter layouts are compared. Both layouts have the same diffusion dimensions, and same gate length and width. The optimized layout (fig 8 (b)) does not used poly or local interconnect wiring. Instead, the wiring is done with the high-speed copper wiring. Also the extrinsic PC to MC capacitors and diffusion resistors are reduced by decreasing the size of the diffusion contact and by placing them in the middle of the diffusion. Both ring oscillators were fabricated on the same wafer, and the mean delay of both statistical delay distributions were compared. Figure 9 shows a comparison of the measured results performed on the same wafer for both rings. The optimized layout mean delay is 5% lower than the reference layout, thus validating the importance of parasitic aware layout optimization of CMOS mm-wave digital circuits. Figure 9 shows also that the lowest performance chips exhibit almost no improvements, however the high performance chips exhibit up to 15% improvement in delay over the mean reference design. This measured 5% delay improvement is achieved under the constraint of using the same diffusion and poly width. Without such a strong constraint, we investigated the possibility to decrease further the switching inverter delay by using the extrinsic FET cut-off frequency of the current (Ft) and power gain (Fmax). The extrinsic Ft and Fmax parameters capture gm, as well as the intrinsic and extrinsic Cgs, Rg and Cgd parasitic. The extrinsic FET Ft and Fmax depend of the gate finger width. From the gate finger minimum width, Ft increases monotonically to an asymptotic value equal to gm/(2 *Cgs,intrinsic). On the other hand the Fmax exhibits a simulated maximum at 0.1 μm gate finger width for our technology. For longer or shorter gate finger width Fmax decreases because of the gate resistor increased and extrinsic parasitic limitation, respectively. Therefore, there is trade-off between Ft and Fmax. Figure 10 shows that the minimum inverter delay is achieved at a gate finger width of 1.4 um and a Fmax to Ft ratio of 1.7 and 1.8 for the N and PFET respectively. The Ft and Fmax are simulated at 0.5V Vgs and Vds and include all the near FET wiring parasitic. Once the extrinsic FET layout is optimized, we optimized the inverter load by replacing the PFET with a resistor and by leveraging the wire inductive effect. The wire inductor can compensate at high-frequency the input capacitor Cgs of the following stage as shown in figure 11. A common source buffer is used to drive the external 50 measurement equipment loads. Figure 12 show the chip microphotograph and detailed implementation. All the measurements were done on-wafer using a 50 GHz spectrum analyzer with external V-band wave-guide mixer to extend the band from 50 to 75 GHz. The maximum measured oscillation frequency of 51.42 GHz is shown in Fig. 13. As shown in figure 14, the measured minimum delay per stage measured is 1.94 ps at a 1.2 V power supply voltage, for a 7.4 mW dissipated active power per stage. This is to the author's knowledge the lowest inverter delay reported to date for a semi-conductor technology. The lowest inverter delay reported is 1.95 ps and uses InP DHBT technology (3). Most of the reported fastest inverter delays for semi-conductor technologies are HBT compound with measured delay above 3 ps (4)(5). A reference fan-out of one static CMOS ring oscillator was measured on the same chip with a delay of 5.7 pS at 1 V. Therefore at 1 V, we measured a 66% improvement in inverter delay by using the parasitic aware optimization and inductive loading to compensate for capacitive parasitic. The measured minimum switching energy is 0.48 fJ at 0.35 V and 2.11 ps delay. The maximum switching energy is 14.48 fJ at 1.2 V and 1.94 ps delay (fig 15). This is more than 2 to 70x lower energy than state of the art compound semi-conductor inverter, achieved at even higher clock frequency (1/(2*delay)) (4)(5). The reference (3) could not be added because the power par inverter stage was not reported in the paper. These results demonstrate the importance of parasitic aware optimization and the capability of nanometer SOI technology. The inverter architecture and optimized layout, as well as the low-voltage operating capability and low parasitic capacitance of the SOI CMOS technology, are responsible for the excellent trade-off between power and equivalent clock frequency. With bipolar technologies (4)(5), low-voltage operation is more difficult to achieve since the transistor Vt (0.8-1 V) cannot be adjusted.

Summary

In this paper we presented the high-performance nanometer technology challenges for MHC, which are sub-ps accuracy, increased discrepancy between physical wafer and layout designer representation, detailed parasitic extraction of near FET wiring, broad process variations requiring statistical analysis and steady state computation for the floating body FET. A variety of ring oscillators was measured and simulated using Monte Carlo simulations. We report the first results of MHC for the SOI 90nm technology. Sub-ps MHC delay was achieved over a broad variety of rings, voltage and temperature. A parasitic aware layout optimization was also presented. Constrained optimizations lead to a 5 % reduction in switching delay for a static CMOS inverter. Unconstrained optimizations lead to a 66 % reduction in switching delay, and a measured record 1.94 ps inverter switching delay. This demonstrates the importance of model and design flow accuracy as well as parasitic aware layout optimization for high-speed digital design. These techniques can also be applied to high-speed analog circuit design.

Acknowledgements

We acknowledge T. Sandwick, M. Bhushan, M. Ketchen, and the SRDC Engineers for their critical supports and contributions.

References

(1) M. Khare, et. al., "A high performance 90nm SOI technology with 0.992 μ^2 6T-SRAM cell," Digest of International Electron Devices Meeting, Dec. 2002, pp. 407-410

(2) J.-O. Plouchart, et. al., "A 243-GHz Ft and 208-GHz Fmax, 90-nm SOI CMOS SoC Technology with Low-Power Millimeter-Wave Digital and RF Circuit Capability," Transactions on Electron Devices, Vol. 52, Issue 7, July 2005, pp. 1370-1375

(3) G. He, et. al., "Self-Aligned InP DHBT with ft and fmax Over 300 GHz in a New Manufacturable Technology," IEEE Electron Dev. Lett., vol. 25, Aug. 2004, pp. 520-522

(4) M. Khater, et. al., "SiGe HBT Technology with fmax/fT = 350/300 GHz and Gate Delay Below 3.3 ps," Digest of International Electron Devices Meeting, Dec 2004, pp. 247-250

(5) K. Ishii, et. al., "3.21 ps ECL gate using InP/InGaAs DHBT technology," Electronics Letter, Vol. 39, Oct. 2003, pp. 1434-1436

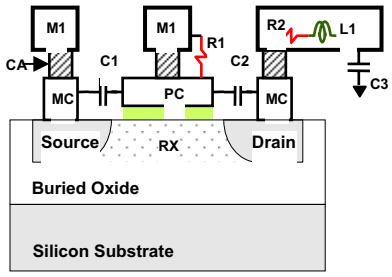


Fig. 1 Cross-section of FET with parasitic components

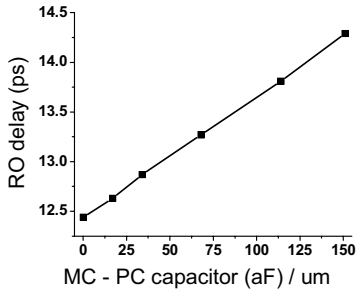


Fig. 2 Simulated delay versus MC to PC capacitor

TABLE I
STEADY-STATE IN SOI AND DELAY IMPACT

Time	Cycle	Delay without IBI (nS)	Delay with IBI (nS)
0.7	10	6.987	7.241
0.5	50	7.014	7.241
7	100	7.031	7.241
35	500	7.075	7.239
70	1000	7.092	7.237
350	5000	7.115	7.222

(IBI stands for indirect body initialization.)

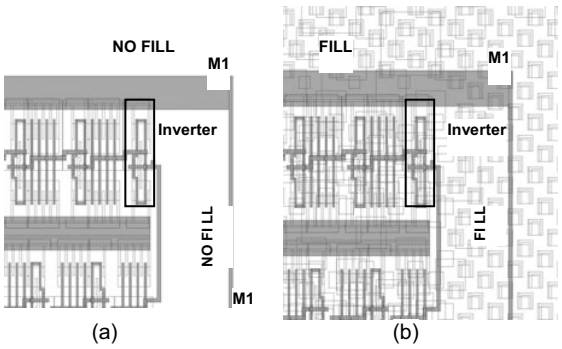


Fig. 3 Fill effect in the layout (a) Layout view without fills (b) Layout view with fills

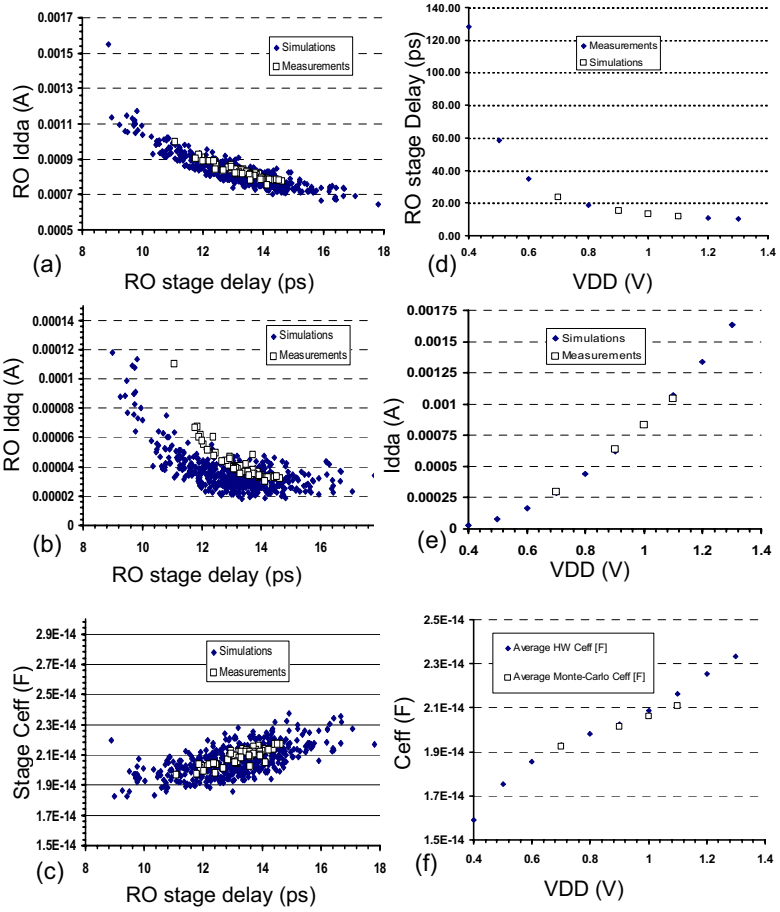
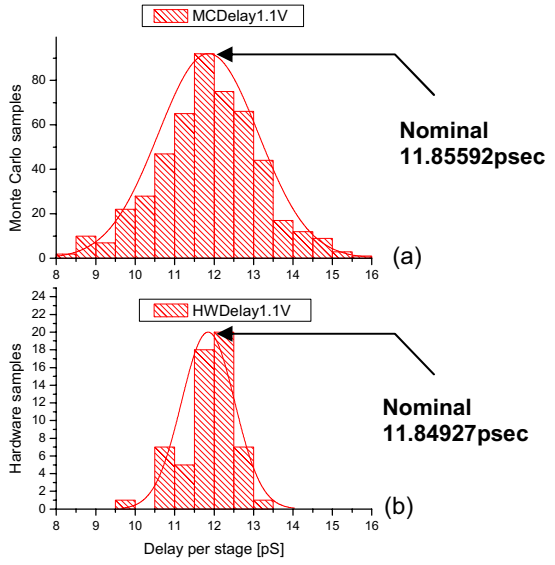


Fig. 4 Statistical RO MHC results at 25C and 1V VDD (a) Idda (b) Iddq (c) Ceff versus stage delay, and mean distribution for (d) Stage delay (e) Idda (f) Ceff versus VDD (Idda is total active current consumption, Iddq is total quiescent current, Ceff is the effective stage capacitance.)



(c)

	Nominal (ps)	Standard deviation (ps)	Frequency deviation (ps)
MC-Delay 1.1V (1:500)	**11.85592**	**1.27362**	**0.05696**
HW-Delay 1.1V (1:59)	**11.84927**	**0.67376**	**0.08772**

Fig. 5 Delay histograms of (a) Monte-Carlo simulation (b) H/W measurement and (c) Summary table of Gaussian fit

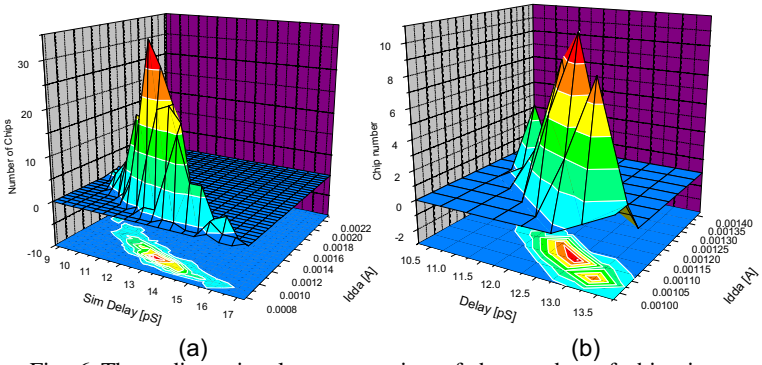


Fig. 6 Three dimensional representation of the number of chips in a delay and current bucket (a) Monte-Carlo simulation (b) H/W measurement

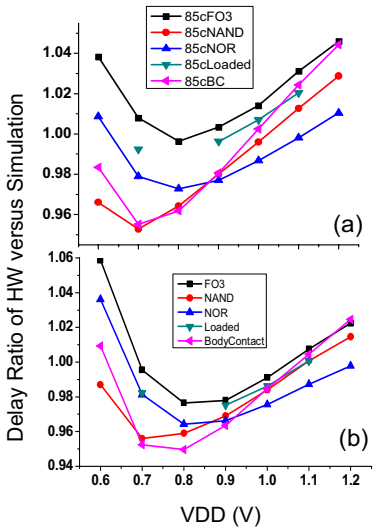


Fig. 7 Mean delay ratio of H/W and Simulation across VDD and variety of RO (a) T=85C (b) T=25C

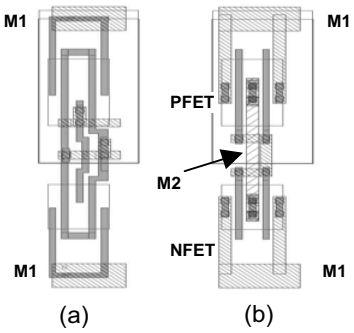


Fig. 8 Inverter layout optimization (a) Reference (b) Optimized version

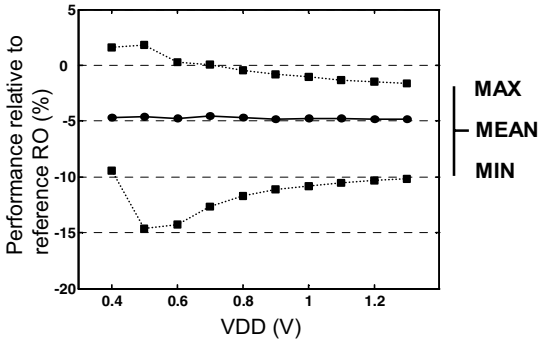


Fig. 9 Relative statistical performance of the ring oscillators (The negative sign means that the optimized RO has lower delay.)

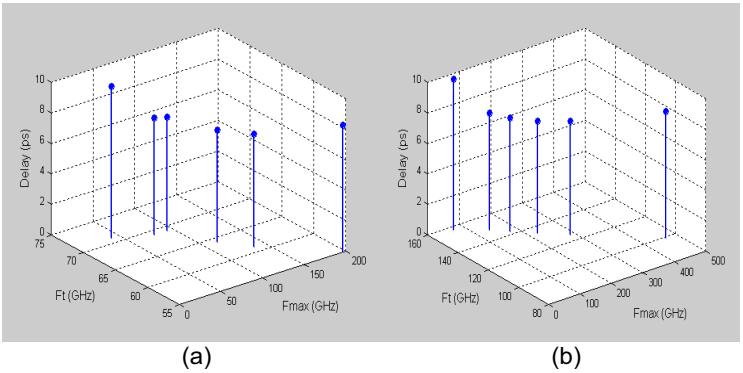


Fig. 10 Simulated Ft, Fmax FET correlation with inverter delay (a) NFET (b) PFET

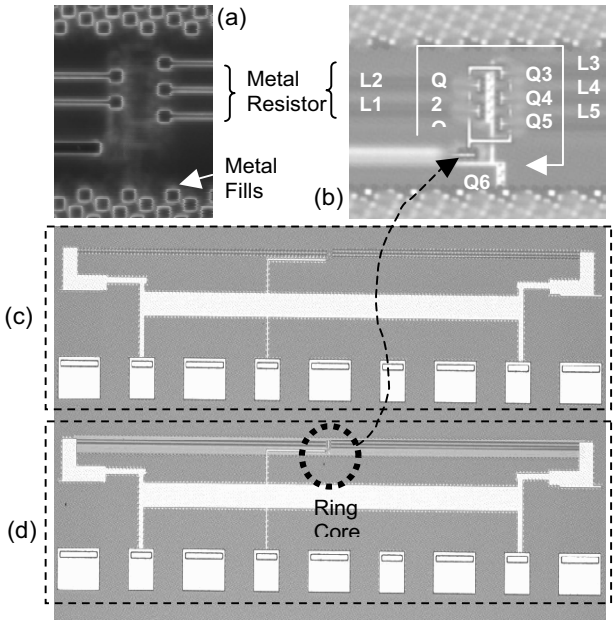


Fig. 12 Ring oscillator microphotograph (die size is 0.3mm x 1.3mm) (a) Metal resistors and metal fills (b) Core circuits micrograph (c) Ring oscillator with fills (d) Ring oscillator without fills

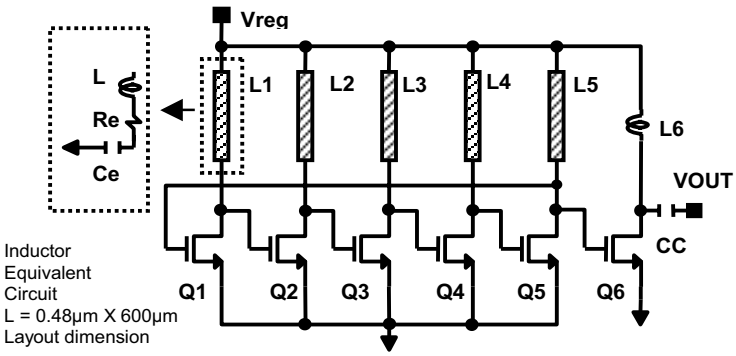


Fig. 11 5-stage ring oscillator schematic. Transistors Q1 through Q6 all have dimensions W=11.34µm / L=90nm.

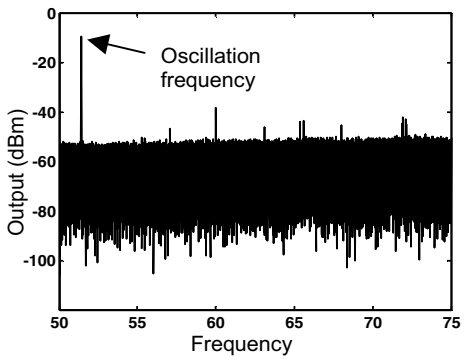


Fig. 13 Output frequency spectrum in 25 GHz span (Vreg = 1.2V and Istage = 6.2mA).

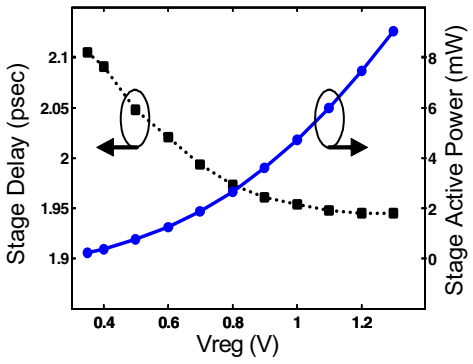


Fig. 14 Delay and active power per stage versus control voltage Vreg
(Stage delay = 1 / (2•frequency•stages))

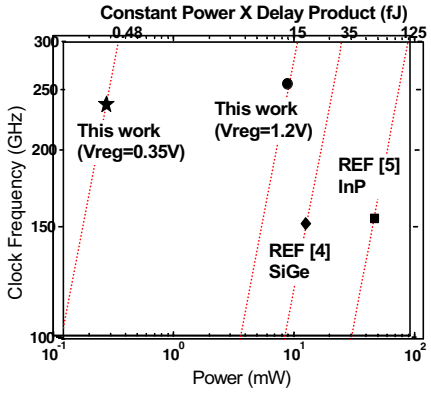


Fig. 15 State-of-the-art of the ring delay