Chapter 1 Dual Work Function Silicon Nanowire MOS Transistors

1.1 Device Design

1.1.1 Introduction to Design Process

In the past, there were several attempts to develop alternative technologies, including molecular technologies [1, 2], that were aimed to replace the current VLSI technology. However, conventional silicon-based technologies prevailed as solid choices over the newcomers for fabricating low power nano devices and circuits without sacrificing high performance. As today's chips require larger die areas to accommodate complex System-On-Chip (SOC) designs, reducing overall power dissipation has been accepted as the major design objective, replacing the need for faster circuit performance. Recent modeling studies in undoped, double-gated SOI MOS transistors revealed that these transistors could produce an order of magnitude less leakage current compared to conventional bulk silicon MOS transistors for achieving ultralow power consumption [3]. However, fabricating ultra-thin transistors sandwiched between two gates with adjustable work function is highly questionable in a production environment since both gates have to be made out of metal in order to produce proper threshold voltage and therefore to maintain a healthy circuit operation. Other studies on this device showed the effect of body thickness to alter the threshold voltage [4] and the variations of the back oxide thickness to decrease overall power dissipation [5]. Two-dimensional analytical modeling [6] and quantum mechanical modeling [7, 8] were also performed to better predict this device's performance and leakage current under different biasing conditions.

Another good candidate is a nano-scale, triple-gated SOI transistors or FINFETs. Theoretical studies conducted on these transistors explored the possibility of increasing transistor performance without increasing power consumption [9]. Recent experimental studies showed close-to-ideal subthreshold slope and Drain-Induced-Barrier-Lowering (DIBL) [10], both of which are important factors to reduce OFF current and power consumption [11]. Besides these promising

DOI 10.1007/978-3-319-27177-4_1





technologies, Silicon Nanowire MOS Transistors (SNT) also offers significant reduction in static and dynamic power consumption and compact layout area without sacrificing circuit performance. SNTs are vertically built on Silicon-On-Insulator (SOI) substrate and cylindrical in shape with a gate surrounding the entire perimeter of the transistor body as shown in Fig. 1.1. The source contact is placed at the bottom of the cylindrical body standing on the SOI substrate while the drain is placed at the top of the device interfacing the first metal layer. The primary objective of this chapter is to design SNTs with dual work function gates and use them in ultra-compact digital CMOS circuits that dissipate minimal static and dynamic power but perform equally or better than the state-of-the-art CMOS circuits.

The design criteria for each NMOS and PMOS SNT are outlined below:

- NMOS and PMOS transistors need to have 300 mV threshold voltage for 1 V CMOS circuit operation for good noise immunity and low OFF current; therefore, different gate metals (dual work function) need to be used for each transistor.
- 2. The static OFF current has to be under 1 pA in each transistor.

However, to achieve these objectives requires changing the body geometry (channel length and radius) of both transistors and determining the device characteristics at each change.

This chapter also discusses the strengths and weaknesses of this technology. Low static and dynamic power dissipation, suppression of Short Channel Effects (SCE), and surface mobility enhancement may be considered as the advantages of SNTs. Alternative placement of NMOS and PMOS transistors as a crossbar configuration may also be counted as an advantage to simplify the layout; however, this method also increases the layout area. Other issues, such as source and drain contact resistances due to small body radius, source contact extension producing high source resistance, and fixed body dimensions resulting in non-adjustable ON currents (therefore limiting transient performance), are definite disadvantages of this technology.

One can trace the foundations of silicon nanowire technology in much earlier studies that investigate material properties and circuits. Silicon nanowires grown by Vapor–liquid–Solid (VLS) mechanism [12] and Chemical Vapor Deposition

1.1 Device Design

(CVD) [13, 14] can be used to fabricate vertical SNTs. In fact, it was demonstrated that silicon nanowires could be used in Static Random Access Memory (SRAM) [15] and high-speed logic circuits [16]. Theoretical studies investigated the bulk and transport properties of silicon nanowires [17] and device properties as a function of wire diameter [18]. Circuit performance and power dissipation of SNTs were briefly studied in 3-D DNA architectures [19, 20].

1.1.2 The Criteria for Low Static Power Dissipation

There are three major components that result in low static power dissipation:

- 1. Junction leakage
- 2. Subthreshold leakage
- 3. Gate-Induced-Drain-Leakage (GIDL) current

Junction leakage current primarily depends on DIBL factor as shown in Eq. 1.1.

$$\text{DIBL} = \left| \frac{V_{\text{TSAT}} - V_{\text{TLIN}}}{(V_{\text{DS}} = V_{\text{DD}}) - (V_{\text{DS}} = 50\text{mV})} \right|$$
(1.1)

where V_{TLIN} and V_{TSAT} are the threshold voltages at $V_{DS} = 50 \text{ mV}$ and $V_{DS} = V_{DD}$, respectively.

Subthreshold leakage current is a function of subthreshold slope, S, and saturation threshold voltage, V_{TSAT} , as expressed in Eq. 1.2.

$$\mathbf{I}_{\text{SUB}} = \text{Io.10}^{\frac{-V_{\text{TSAT}}}{8}} \tag{1.2}$$

Here, Io is the drain current at $V_{GS} = V_{TLIN}$ and S is given in Eq. 1.3 [21].

$$S = \frac{kT}{q} \log \left(1 + \frac{C_D}{C_{OX}} \right)$$
(1.3)

In this equation, C_D and C_{OX} are the channel depletion region and gate oxide capacitances, respectively.

The third component, GIDL current, is a strong function of transverse electric field, ES, at the semiconductor surface perpendicular to the device axis as given by Eq. 1.4 [22].

$$IGIDL = A.E_{s.exp}\left(-\frac{B}{E_{s}}\right)$$
(1.4)

where

$$E_{S} = \frac{V_{DG} - V_{FB} - 1.2}{3tox}$$
(1.5)

A is pre-exponential constant, B is a physically based exponential parameter suggested by [23], V_{DG} is the drain-to-gate potential, V_{FB} is the flat band voltage, and tox is the oxide thickness.

Therefore, the OFF current can be reduced by decreasing DIBL, tox, body doping concentration, and E_S . In this work, tox is set to minimum value of 1.5 nm to maintain the gate leakage current to a negligible level with respect to I_{OFF} as suggested by [3]; the body doping concentration is reduced to intrinsic level to minimize C_D , and E_S is also kept small due to non-overlapping gate-drain region and sub-10 nm wire radius [22].

1.1.3 Device Structure

Both NMOS and PMOS transistors are designed as enhancement type with uniform, undoped silicon bodies constructed perpendicular to the substrate. Both have the same body radius and effective channel length. Source/drain (S/D) contacts are assumed to have ohmic contacts. Both NMOS and PMOS transistors have metal gates and 1.5 nm thick gate oxide.

Device simulations are performed using Silvaco's 3-D ATLAS device simulation environment with a 1 V power supply voltage. Half of the device is constructed in a 2-D platform and then rotated around the y-axis to create a 3-D cylindrical structure for simulations. The device radius is changed from 1 nm to 25 nm while its effective channel length is varied between 5 nm and 250 nm.

1.1.4 Physical Models Used in Device Simulations

Even though sub-100 nm device geometry requires inclusion of Schrödinger's equation to calculate effective electron/hole masses and density of states due to the perturbations in the silicon conduction and valance bands, ATLAS simulator is limited to the full usage of such quantum mechanical effects. Instead, this study follows a semiclassical approach in which the semiconductor surface potential and density of states are corrected using density gradient method [24].

Mobility models are composed of two parts to estimate the effects of low and high electric fields. Lombardi's vertical and horizontal electric field dependent mobility model is used for low electric field effects [25]. Velocity saturation and high electric field effects are estimated by Caughey's drift velocity model [26]. Mobility degradation due to lattice temperature is included using Arora's model [27].

Concentration-dependent Shockley–Read–Hall recombination and surface recombination models are included to estimate the recombination rates in the bulk and at the silicon/oxide interface, respectively. Serberherr's impact ionization model constitutes the only generation model in the simulations [28]. Gate oxide tunneling mechanisms and hot carrier injection are ignored because these mechanisms largely depend on oxide growth and composition, and change from one processing condition to another.

1.1.5 Determining Metal Gate Work Function Values for NMOS and PMOS Transistors

The first task in this design process is to determine an individual metal work function for each NMOS and PMOS transistor at a minimum channel length of 5 nm in order to produce a threshold voltage of approximately 300 mV. This value constitutes 30 % of the 1 V power supply voltage and provides sufficient noise immunity for any CMOS gate. Threshold voltage of each NMOS and PMOS transistor is measured as a function of work function for the body radius from 1 nm to 25 nm as shown in Fig. 1.2. Longer channel length devices yield marginally higher threshold voltage and improve noise margin slightly. The intersection of threshold voltage with 300 mV level in Fig. 1.2 is projected to the x-axis to yield an individual work function value for each NMOS and PMOS transistor at a different body radius. Threshold voltages are measured using two different methods: the first method extrapolates the maximum slope of I_D-V_{GS} curve towards V_{GS} -axis and



Fig. 1.2 Threshold voltages of NMOS and PMOS nanowire transistors as a function of metal work function at a minimum effective channel length of 5 nm. Radius of both NMOS and PMOS transistors is changed between 1 and 25 nm

defines the intercept as the threshold voltage; the second method determines the threshold voltage from the gate voltage at $I_{DS} = \zeta(W/L)$ for $V_{DS} = 50$ mV, where ζ is 10^{-7} A for NMOS and 10^{-8} A for PMOS transistors. This method is suggested by Liu et al. [29] and consistently produced 11 % and 3 % lower threshold voltages for NMOS and PMOS transistors, respectively.

1.1.6 The OFF Current Requirement

The leakage current is an important factor towards lowering overall standby power consumption; both NMOS and PMOS transistors are designed to have static leakage currents smaller than 1 pA, which is significantly smaller than SOI transistors in earlier modeling studies [3, 5, 9] and several orders of magnitude smaller than the technology trend predicted by Sery et al. [30]. Therefore, while most transistors with 1 nm to 5 nm radius produced I_{OFF} less than 1 pA and were considered as potential candidates for an optimum transistor design, transistors with larger radii were eliminated because their leakage currents exceeded 1 pA as shown in Fig. 1.3. In this figure, the transistor geometries closest to the dashed line are considered potential candidates since they produce higher ON currents for a given value of I_{OFF} .

1.1.7 Intrinsic Transient Time

Following the device selection process for low power dissipation in Fig. 1.3, the intrinsic transient time, τ , of each "selected" transistor is measured and then plotted as a function of I_{ON} in Fig. 1.4. Intrinsic transient time determines the time interval for a transistor to charge (or discharge) the gate capacitance of an identical transistor when it is fully on ($V_{DS} = V_{GS} = 1$ V) and it is a quick way of understanding the transient characteristics of an individual transistor without building any circuitry. In Fig. 1.4, ON currents of the selected NMOS and PMOS transistors start diverging from each other after 4 nm radius and 40 nm effective channel length; larger wire radius provides higher I_{ON} values for NMOS transistors, but it reaches a saturation plateau for PMOS transistors. Therefore, the 4 nm radius and 40 nm effective channel length combination is considered an optimal choice to produce approximately equal drive currents and intrinsic transient times for both NMOS and PMOS transistors.

Figure 1.5 shows the ON currents for NMOS and PMOS transistors as a function of L_{EFF} for different wire radii and helps to explain the ON current behavior of each device in Fig. 1.4. For wire radius greater than 5 nm, ON currents of both NMOS and PMOS transistors increase with decreasing L_{EFF} as shown in Eq. 1.6 [31, 32].



Fig. 1.3 The ON versus OFF current of NMOS and PMOS nanowire transistors. The radius of both transistors is changed between 1 and 25 nm while their effective lengths are varied between 5 and 250 nm. Each transistor has a specific gate work function value for each radius as specified in Fig. 1.2. Note that transistors with shorter effective channel lengths produce higher OFF currents

$$I_{ON} = \frac{\mu_{EFF}.\varepsilon ox.W}{2L_{EFF}.t'ox} \left\{ (V_{GS} - V_T)^2 - \frac{16k^2T^2}{q^2} \cdot \frac{t'ox.R.\varepsilon s}{\varepsilon ox} \cdot exp \left[\frac{q}{kT} (V_{GS} - V_{TO} - V_{DS}) \right] \right\}$$
(1.6)

where

$$t'ox = R.ln\left(1 + \frac{tox}{R}\right)$$
(1.7)

$$V_{TO} = V_{FB} + \frac{kT}{q} ln \left(\frac{kT \epsilon s N}{q^2 n i^2} \right)$$
(1.8)

Here, N and R are the body doping concentration and radius of the SNT, respectively.



However, as the wire radius is further reduced from 5 nm to 1 nm, ON currents become independent of L_{EFF}. This behavior is not supported by the expression in Eq. 1.6. If inversion charge concentration, Qi, and drift velocity of electrons (holes) are examined throughout the body of small radius devices, one observes that both charge distribution and velocity are uniform. For example, if $V_{GS} = V_{DS} = 1$ V is applied to an NMOS transistor whose radius is smaller than 5 nm, the value of Qi approaches 10^{19} cm⁻³ and electron drift velocity becomes equal to 10^7 cm/s for device lengths between 5 and 150 nm. These observations suggest that electrons

Fig. 1.4 The ON current versus intrinsic transient time of the "selected" NMOS and PMOS nanowire transistors whose leakage currents are below 1 pA

travel with saturation drift velocity across the transistor body and becomes independent of L_{EFF} as given by Eq. 1.9 [21].

$$I_{ON} = \pi . R^2 vsatQi \tag{1.9}$$

The validity of this statement can be further verified by computing the ON current ratios of small radius devices in Fig. 1.5 and comparing them against the square of body radius ratios. For example, I_{ON} of 1 nm, 2.5 nm, and 5 nm radius NMOS transistors are 0.54 μ A, 3.3 μ A, and 13 μ A, respectively. When we compute the ratio of the ON currents of the R = 5 nm device to the R = 2.5 nm device, we obtain 3.94. If the same ratio is computed using Eq. 1.9, the result becomes equal to 4, assuming Qi in both devices is equal. Similarly, the ratio of ON currents of the R = 2.5 nm device to the R = 1 nm device produces 6.11 from Fig. 1.5 while the same ratio produces 6.25 according to Eq. 1.9.

For large wire radius shown in Fig. 1.5, I_{ON} follows Eq. 1.6 and the ratio of ON currents becomes equal to the ratio of effective electron and hole mobilities in NMOS and PMOS transistors if minor deviations in threshold voltages are ignored. For example, default effective electron and hole mobilities in Silvaco's ATLAS design simulation environment produce an ON current ratio of 3.9, whereas the ON current ratio extracted from Fig. 1.5 is equal to 3.38 for 10 nm radius NMOS and PMOS transistors. However, as the wire radius is reduced and the transistor bulk effect disappears, the ON current follows Eq. 1.9 and the ratio of ON currents becomes approximately proportional to the square of the NMOS and PMOS transistors produce 0.54 μ A and 0.48 μ A ON currents, respectively. The ratio of ON currents approaches unity rather than approaching to the ratio of effective electron to hole mobilities as in large radius devices.

1.1.8 DC Device Characteristics

Figure 1.6 shows the threshold voltage roll-off of the 4 nm radius transistors with effective channel lengths ranging between 40 nm and 150 nm. The figure also includes earlier bulk and SOI transistor data for comparison [33–37]. The amount of ΔV_T is 6 mV for the NMOS and 11 mV for the PMOS transistors. These results are more than an order of magnitude smaller than the values of bulk silicon transistors which require heavily doped substrates to prevent SCE but consequently suffer from early impact ionization and large leakage currents.

The threshold voltage behavior of SNTs in Fig. 1.6 is not surprising because the same trend can also be seen in Fig. 1.7. This figure illustrates that the SCE gradually disappears as wire radius decreases towards 1 nm; larger radius devices are affected by the SCE and exhibit in excess of 100 mV threshold voltage change. This shows that bulk transistors or transistors fabricated on an SOI substrate thicker than 5 nm are still susceptible to threshold voltage variations as a function of device geometry.



Silicon wire transistors, having a radial gate configuration, controls and suppresses SCE simply by reducing wire radius.

The amount of DIBL is 57 mV/V for the NMOS and 53 mV/V for the PMOS transistors with 4 nm radius and 40 nm effective channel length. These values are shown in Fig. 1.8 and compared with previously published data [3, 10, 34, 36, 38, 39].



Subthreshold slope is 62 mV/dec for NMOS and 62.5 mV/dec for PMOS transistors at a drain voltage of 1 V. These results are plotted in Fig. 1.9 and show close-to-ideal characteristics in comparison with Kim's modeling results on double-gated SOI transistors [3] and previously published experimental data [10, 36–43].

Figure 1.10 shows the three components of the OFF current for NMOS transistor. Subthreshold leakage is the first component when there is no impact ionization in the device body. Junction leakage is the result of impact ionization at high lateral electric fields and it doubles the total OFF current. Band-to-band leakage or GIDL component is small compared to junction leakage because of three factors. The first factor is the absence of a gate-drain overlap region in the proposed device structure: only fringing component of the transverse electric field emanating from the edge of the gate may induce GIDL. The second factor is the decrease in transverse electric field with respect to a bulk device with a single gate: surface potential in a bulk or partially depleted SOI device is appreciable to promote GIDL current generation [22]. The third factor is the magnitude of the power supply voltage: the drain-togate potential being less than the silicon band gap is not an effective way to create enough band-bending at the semiconductor surface to allow the valance band electrons to tunnel into the conduction band. Gate oxide tunneling using Concannon's model [44], on the other hand, may increase the OFF current beyond its designed limit as it almost doubles the total OFF current as shown in Fig. 1.10. However, the magnitude of this current primarily depends on processing conditions including gate oxide composition, quality and defect levels during growth, and it is not included in this study.

Figure 1.11 shows the ON current of an NMOS transistor with and without quantum approximations. Van Dort's model is designed for thin gate oxide devices and empirically corrects the surface potential by broadening the energy band gap [45]. Density gradient model is calibrated with the Poisson–Schrodinger equation in simulations and it calculates position-dependent potential energy from the semiconductor surface towards the current transport axis according to higher derivatives of carrier distribution in the channel [24]. Potential energy corrections consequently modify electron and hole distributions in the channel and compute electron and hole current densities.

Figure 1.12 shows the output I–V characteristics of NMOS SNTs whose radius values are between 1 nm and 25 nm at $L_{EFF} = 40$ nm. The drain–source breakdown



Fig. 1.12 Output I–V characteristics of NMOS nanowire transistors with an effective channel length of 40 nm at $V_{GS} = 1$ V. Each nanowire transistor has a leakage current below 1 pA. Dashed lines show the output I–V characteristics of nanowire transistors with "virtual" body contacts



voltage moves towards higher values and finally disappears as radius is reduced towards 1 nm. Note that the breakdown voltage is a function of ionized electron-hole pairs throughout the device body, and it forms a "soft" kink effect in each I–V curve. The kink effect disappears when the transistor bulk is connected to a "virtual" ground in simulations as indicated by the dashed lines in Fig. 1.12.

Surface mobility enhancement in silicon nanowire transistors is also investigated. Semiconductor surface potential in a radial transistor body diminishes as the body radius is reduced. Surface potential in bulk transistors, on the other hand, reaches its maximum value because of the substrate contact and degrades the device mobility. An NMOS transistor with 4 nm radius and 40 nm effective channel length produces an ON current of 8.3 μ A, whereas a bulk NMOS transistor with the same body configuration produces only 6 μ A.

Figure 1.13 illustrates the first circuit-related result and shows the inverter transfer characteristics produced by 4 nm radius and 40 nm effective channel length NMOS and PMOS SNTs. The inverter threshold voltage computed by the projection of $V_{OUT} = 0.5$ V to the x-axis is slightly off-center at 410 mV due to the slightly higher NMOS drive current, but the inverter still produces sufficient low and high noise margins at 340 and 570 mV, respectively, for noise-free circuit operation.

1.2 Circuit Simulations and Performance

1.2.1 Parasitic Extraction and Post-layout Issues

To understand the circuit performance, power dissipation, and layout, several primitive gates, including an inverter, 2-input and 3-input NAND, NOR, XOR gates, and a full adder were built. All measurements were conducted before and after parasitic layout extraction and compared with each other to understand the effects of parasitic wire resistance, capacitance, and contact resistance on circuit performance. Since these transistors are constructed perpendicular to the substrate,



Fig. 1.14 Copper resistivity as a function of width. Srivastava's scattering model is extrapolated to obtain the resistivity value for 6.4 nm copper wires in this study. Prior experimental data are included for comparison



the minimum exposed transistor feature on the layout is 4 nm wire radius to make contacts. Copper wires with 6.4 nm width and 1.4 aspect ratio (wire height to width) are used for interconnects and 2.4 nm by 2.4 nm vias are used for contacts. Since sub-10 nm range copper wire electrical characteristics do not exist in the literature, copper resistivity was extrapolated from Srivastava's model on 1.4 aspect ratio wires [46]. Figure 1.14 shows copper resistivity as a function of wire width for aspect ratios of 1.4 and 1.6, and also contains experimental results for comparison purposes [47-51]; 20 µohm-cm resistivity was subsequently used to calculate the sheet resistance for 6.4 nm wide interconnects. Similarly, contact resistance was extrapolated from the experimental data for 100 nm and larger via diameters and resulted in 18.5 Ω for each metal contact as shown in Fig. 1.15 [48–51]. The estimations on contact resistance and wire resistivity likely contain errors since these parameters are extracted either from a technology that supports 100 nm wire features or extrapolated from a simplified scattering model that does not take into account crucial scattering mechanisms such as interface (wire surface) and grain boundary scattering [52]. Especially, 6.4 nm wire width is exposed to all such mechanisms because this dimension is below a typical grain size of copper (approximately 10 nm) and much lower than the mean free path of electrons (40 nm). N-well/P-well extension to form a source contact also introduces a series resistance to the transistor. The measured value of source extension for the NMOS transistor is 650 Ω and for the PMOS is approximately 2.5 k Ω . However, the change in overall circuit delay as a result of interconnect sheet resistivity, contact resistance and source extension is not significant because the equivalent SNT channel resistance is much larger than the sum of all these extrinsic resistances. A simple RC calculation on inverter rise and fall times reveals approximately $34 \text{ k}\Omega$ for PMOS and 7 k Ω for NMOS transistor channel resistance. If one limits the total contact, wire and source extension resistances to be 10 % of the equivalent NMOS channel resistance (or 790 Ω) to avoid interconnect-related delays, then the discharge path can accommodate 90 nm long copper wire between two contacts. The maximum wire length in the inverter layout is less than 50 nm long. The inverter charge path can even support more wire resistance since the equivalent PMOS channel resistance is 34 k Ω instead of 7 k Ω . More complex circuits containing multiple transistors in series can tolerate higher number of contacts and wire lengths in the charge and discharge paths. However, wire lengths outside the cell boundary in the form of long chip-level routes have the greatest sensitivity to resistivity errors and become an important issue when considering overall circuit delays and slow logic transitions. Fortunately, in the upper-metal routing, design rules are more relaxed, allowing wider and thicker wires.

Area, fringe, and coupling capacitances of metal 1 and metal 2 wires per unit length are calculated using Ansoft's 2-D electrostatic solver. These capacitance values are used to extract metal-to-metal and metal-to-substrate parasitic capacitances from layouts for circuit simulations.

1.2.2 Transient Performance

Post-layout transient characteristics of various CMOS gates composed of 4 nm wire radius and 40 nm channel length transistors are shown in Figs. 1.16 and 1.17 in terms of worst-case transient time and worst-case delay, respectively. The worst-case transient time is determined by projecting 10 and 90 % of the output voltage onto the time-axis and measuring the difference. Similarly, the worst-case delay is determined by projecting 50 % of the input and 50 % of the output voltage values onto the time-axis and measuring the difference. Each transient characteristic is plotted as a function of the output capacitance. Considering the gate capacitance of a single transistor is 32 aF, maximum output capacitance of 200 aF in simulations corresponds to a fan-out of approximately six identical transistors.

The worst-case transient time is essentially equivalent to the rise time of a gate since a PMOS transistor has almost five times higher resistance compared to an NMOS transistor as discussed earlier. The worst-case transient times of the inverter, 2-input and 3-input NAND-gates in Fig. 1.16 overlap each other primarily due to



the single PMOS transistor charging the output capacitance. The worst-case transient times of the 2-input NOR and XOR circuits cluster together because there are two PMOS transistors in series charging the output load. The full adder and 3-input NOR circuits are in close proximity and reveal the highest transient times because the number of PMOS transistors in series increases from two to three in the critical charging path. The worst-case transient time of the full adder is expressed as $T = 4.74 + 0.23C_L$ in picoseconds, where C_L is the output load capacitance in aF.

The worst-case delay in Fig. 1.17 behaves similar to the worst-case transient time in Fig. 1.16 because the worst-case delay uses the same critical charging and discharging paths. The worst-case delay of the full adder circuit is expressed as $T_D = 8.50 + 0.15C_L$ in picoseconds.

The worst-case gate delay values of silicon nanowire technology are comparable to SOI but smaller than bulk silicon technologies. Kim et al. [3] obtained 4 ns and 5 ns individual inverter delays from a chain of double-gated SOI and bulk silicon inverters, respectively. The worst-case inverter delay in this study is approximately 2.5 ps when the inverter output is connected to the input of an identical inverter.

1.2.3 Power Dissipation

The worst-case power dissipation is composed of static power dissipation discussed earlier and dynamic power dissipation which is a function of frequency of operation, fop, power supply voltage, V_{DD} , and load capacitance, C_L as shown in Eq. 1.10.

$$Pdyn = fop.C_L.V_{DD}^2$$
(1.10)

When V_{DD} and fop are adjusted to achieve the optimum circuit performance and noise margin, the only possible variable to reduce Pdyn is the load capacitance. Even though the dimensions of a bulk transistor can be changed to have the same gate capacitance of a single nanowire transistor, impact ionization, punch-through effect, and high S/D capacitance are still potential problems for the bulk device. Dual-gated SOI transistors benefit the same advantages as the SNTs, but their gate capacitance, and therefore the dynamic power dissipation, doubles as shown in Eq. 1.10.

The worst-case post-layout power dissipation of various logic gates with 10 aF capacitive load is shown in Fig. 1.18 as a function of frequency. The worst-case power dissipation is obtained by considering all possible input combinations to a logic gate, measuring the average value of the power supply current within one clock period for each combination (activity factor = 1 %) and finally selecting the combination that yields the maximum average current. Each current waveform is averaged in one clock period during charging and discharging the output capacitance. The worst-case power dissipation of a 2-input NAND gate is 36.9 nW at





1 GHz and increases by 24.9 nW/GHz for a 10 aF output load. The worst-case power dissipation increases with number of transistors, layout complexity, and the number of "parallel" charging or discharging paths to a capacitive load. A full adder, as a more complex circuit, dissipates 85.0 nW at 1 GHz and the power dissipation increases by 51.2 nW/GHz for a 10 aF output load.

Figure 1.19 shows the worst-case post-layout power dissipation of each gate as a function of load capacitance at 1 GHz. The worst-case power dissipation for the 2-input NAND gate and full adder is $P = 14.1 + 2.19C_L$ and $23.6 + 4.04C_L$, respectively, in nanowatts.

1.2.4 Cell Layout and Gate Area Estimations

Layouts of various gates including an inverter, 2-input and 3-input NAND, NOR, and XOR circuits, and a full adder are designed using 4 nm radius and 40 nm effective channel length nanowire transistors. Figure 1.20 shows the cross section and the corresponding layout of an SNT. The active region defines the circular body of the SNT, which is surrounded by an N-well if the transistor is an NMOS device or P-well if it is a PMOS transistor. The outmost circle represents the metal gate. All contacts are indicated by 2.4 nm by 2.4 nm black squares touching the drain, the source and the gate of the transistor. Figure 1.21 shows the layout of a full adder. All interconnects between transistors are established by 6.4 nm wide wires. Further area reduction in this layout is possible if more than two metal layers are used to connect all its inputs and outputs with adjacent cells. Layout areas of the primitive gates used in this study are listed in Table 1.1. A recent 6-transistor SRAM cell designed in a 65 nm technology occupied a cell area of 0.57 μ m² [41]. The 30-transistor full adder in this study has a cell area of approximately 0.11 μ m², which is about 5 times smaller than the SRAM cell and contains five times more



Fig. 1.20 Cross section and layout topology of a single, 40 nm effective channel length and 4 nm body radius NMOS transistor. Note that the source (drain) contact via is not shown on the cross section. A separate N-well completely surrounds the P-well of the PMOS transistor to prevent latch-up



Fig. 1.21 The full adder layout using 40 nm effective channel length and 4 nm body radius NMOS and PMOS nanowire transistors. A, B, and C are the two inputs of the full adder and the carry-in, respectively. \overline{A} , \overline{B} , and \overline{C} correspond to the two complemented inputs of the full adder and the complemented carry-in, respectively

transistors. There are two limiting factors which prevents further layout area reduction in SNTs: source contact extension and gate metal thickness. The former can be minimized by employing small contacts at the expense of increasing contact resistance; the latter has a thickness limit below which metallic grains separate from each other, forming a discontinuous metallic film. This study uses 10 nm gate metal thickness.

110.000

Table 1.1 Layout area of various gates built with 40 nm effective channel length and 4 nm body radius NMOS and PMOS nanowire transistors	Gate	Area (nm ²)
	Inverter	7400
	2-input NAND	14,800
	3-input NAND	19,000
	2-input NOR	14,800
	3-input NOR	19,000
	2-input XOR	24,000

 Table 1.2
 Circuit performance, power dissipation, and layout area of full adder circuits in this study and earlier work

Full adder

Lg (nm)	V _{dd} (V)	Fop (MHz)	P _T (nW)	Delay (ps)	Area (µm ²)	References
350	3.3	-	164,000	227	-	[55]
350	1.2	50	2490	2037	387	[53]
350	1.8	50	6090	827	387	[53]
350	2.5	50	12,820	528	387	[53]
350	3.3	50	24,120	406	387	[53]
350	3.3	-	65,000	400	-	[56]
250	3.3	-	58,000	300	-	[56]
180	3.3	-	30,000	100	-	[56]
180	1.0	100	2500	650	-	[57]
180	1.8	100	6230	292	100	[54]
180	1.0	100	1450	756	100	[54]
180	1.8	300	345	195	-	[58]
180	1.8	50	11	327	-	[59]
40	1.0	1000	556 ^a	28 ^a	0.11 ^a	This work

^aAn output load of 130 aF (4 transistor gates)

A comparison of the SNT full adder and the earlier conventional adders is provided in Table 1.2 in terms of transient performance, power dissipation, and layout area [53–59].

1.2.5 Manufacturability

Silicon nanowire transistors can be manufactured relatively easier compared to the dual-gated SOI transistors. The processing steps in Fig. 1.22 shows a method to fabricate NMOS SNT using chemical–mechanical polishing and other conventional processing methods. Dual-gated SOI transistors require metal gates on both sides of the thin transistor body and their manufacturability may not be possible with conventional processing tools.



gate oxide CVD N+ oxide CVD oxide N+ metal etch N+ contact metal N+ N+

Fig. 1.22 Processing steps for NMOS transistor. (a) Intrinsic Si wire is grown, (b) Gate oxide is grown, and source junction is formed; anisotropic PECVD oxide is deposited, (c) Metal gate is deposited, (d) Thick oxide is deposited to the length of the wire and CMP is applied until wire end is detected, (e) Thick oxide is recessed to drain gate junction by preferential Reactive Ion Etching (RIE), (f) Metal gate is wet etched to define physical gate length, (g) Second thick oxide is deposited, CMP is applied, drain junction is formed by ion implantation, (h) Gate and drain contacts are formed

CVD

oxide



Fig. 1.23 Complete NMOS transistor cross section obtained by Athena process simulator

In order to verify the process flow in Fig. 1.22, two-dimensional processing simulations are carried out in Silvaco's Athena process design environment. The processing steps from Fig. 1.22a through Fig. 1.22h produced the NMOS transistor whose final cross section is shown in Fig. 1.23.

Initially, 450 nm long, intrinsic silicon nanowires are grown perpendicular to the heavily doped N-well (P-well) regions that define the source contacts for the NMOS (PMOS) transistor with Au catalyst. Both the phosphorus-doped N-well and boron-doped P-well define the source region of the NMOS and PMOS transistors, respectively.

After the wire growth, the Au catalyst is stripped from the top of the grown wire using a wet etch, as shown in Fig. 1.22a. This step is followed by a 5 nm gate oxide growth at 975 °C for 30 min. Next, 100 nm CVD oxide is deposited anisotropically to define the gate–source boundary [22]; anisotropic CVD oxide deposits only horizontally to the substrate, but it does not attach to silicon wire walls as shown in Fig. 1.22b. The oxidation step causes phosphorus atoms from the N-well (boron atoms from the P-well) to diffuse approximately 10 nm into the silicon wire from

the substrate surface. A 100 nm thick tungsten layer is deposited as the gate material and defined as shown in Fig. 1.22c. Isotropic CVD oxide is deposited, and a Chemical-mechanical Polish (CMP) step is applied until tungsten is seen, as shown in Fig. 1.22d. CVD oxide is recessed to define the depth of the drain region or drain-gate boundary as shown in Fig. 1.22e. Exposed tungsten is wet etched, as shown in Fig. 1.22f. A second CVD oxide layer is deposited followed by a CMP step that stops at silicon. Low energy phosphorus (boron) is implanted perpendicular to the wire end to form the drain contact for the NMOS (PMOS) transistor as shown in Fig. 1.22g. A 20 s Rapid Thermal Annealing (RTA) step is applied at 900 °C to activate phosphorus (boron) implants without damaging the gate metal. A 100 nm CVD oxide is subsequently deposited. Aluminum via contacts are formed using a unidirectional Reactive Ion Etch (RIE) step. A 140 nm thick aluminum layer is deposited as the metal 1 layer as shown in Fig. 1.22h.

1.3 Summary

In this exploratory work, silicon nanowire CMOS circuits are studied for low power and high density VLSI applications. Three-dimensional undoped NMOS and PMOS nanowire transistors are designed and optimized in Silvaco's ATLAS device design environment to maximize the ON current and to keep the OFF current below 1 pA as a function of device geometry. Threshold voltage of each transistor is adjusted by an individual gate metal work function. As the body radius is reduced from 25 nm towards 1 nm, the variation in threshold voltage is observed to decrease from 140 mV to approximately 6 mV for NMOS transistors and from 130 to 11 mV for PMOS transistors, both of which are indications of diminishing SCE. The ON current is also observed to be independent of the channel length for small radius transistors due to the influence of large lateral electric field forcing carriers in the inversion region to travel with saturation drift velocity. Threshold voltage roll-off, DIBL, and subthreshold slope of silicon nanowire NMOS and PMOS transistors are measured and compared with earlier studies. Transient circuit performance, power dissipation and layout area of an inverter, 2-input and 3-input NAND, NOR, XOR gates and full adder circuits are measured and analyzed. As a specific case, simulation results show that the worst-case delay of a full adder circuit is 8.5 ps at no load and it increases by 0.15 ps/aF. The worst-case power dissipation of the same circuit is 23.6 nW at no load and increases approximately by 4.04 nW/aF. The layout area of the full adder is also measured to be 0.11 μ m² which is 5 times smaller than a 6-transistor SRAM cell laid out using a 65 nm technology node. Compared to the results reported previously for silicon bulk and double-gated SOI transistors, this study indicates the silicon nanowire technology may be a potential choice for the future of VLSI circuits because of its low power dissipation in a compact layout area.

References

- 1. Collier CP, Wong EW, Belohradsky M, Raymo FM, Stoddard JF, Kuekes PJ, Williams RS, Heath JR (1999) Electronically configurable molecular-based logic gates. Science 285:391–394
- Goldstein SC, Budiu M (2001) Nanofabrics: spatial computing using molecular electronics. Proc 28th Annu Int Symp Comp Archit: 178–189
- 3. Kim K, Das KK, Joshi RV, Chuang C-T (2005) Leakage power analysis of 25-nm double-gate CMOS devices and circuits. IEEE Trans Electron Devices 52(5):980–986
- Tsutsui G, Saitoh M, Nagumo T, Hiramoto T (2005) Impact of SOI thickness fluctuation on threshold voltage variation in ultrathin body SOI MOSFETs. IEEE Trans Nanotechnol 4(3):369–373
- Zhang R, Roy K, Janes DB (2001) Double-gate fully-depleted SOI transistors for low-power high performance nano-scale circuit design. Proc Int Symp Low Power Electron Design: 213–218
- Reddy GV, Kumar MJ (2005) A new dual-material double-gate (DMDG) nanoscale SOI MOSFET-two dimensional analytical modeling and simulation. IEEE Trans Nanotechnol 4(2):260–268
- 7. Kumar A, Kedzierski J, Laux SE (2005) Quantum-based simulation analysis of scaling in ultrathin body device structures. IEEE Trans Electron Devices 52(4):614–617
- Vasileska D, Ahmed SS (2005) Narrow-Width SOI devices: the role of quantum-mechanical size quantization effect and unintentional doping on device operation. IEEE Trans Electron Devices 52(2):227–236
- 9. Yang J-W, Fossum J (2005) On the feasibility of nanoscale triple-gate CMOS transistors. IEEE Trans Electron Devices 52(6):1159–1164
- Yu B, Chang L, Ahmed S, Wang H, Bell S, Yang C-Y, Tabery C, Ho C, Xiang Q, King T-J, Bokor J, Hu C, Lin M-R, Kyser D (2002) FinFET scaling to 10 nm length. Tech Dig IEDM: 251–254
- 11. Choi Y-K, Chang L, Ranade P, Lee J-S, Ha D, Balasubramanian S, Agarwal A, Ameen M, King T-J, Bokor J (2002) FinFET process refinements for improved mobility and gate work function engineering. Tech Dig IEDM: 259–262
- 12. Wagner RS, Ellis WC (1964) Vapor–liquid–Solid mechanism of single crystal growth. Appl Phys Lett 4(5):89–90
- Kamins TI, Williams SR, Basile DP, Hesjedal T, Harris JS (2001) Ti-catalyzed Si nanowires by chemical vapor deposition: Microscopy and growth mechanisms. J Appl Phys 89(2):1008–1016
- 14. Islam MS, Sharma S, Kamins TI, Williams RS (2004) Ultra-high-density silicon nanobridges formed between two vertical silicon surfaces. Nanotechnology 15:L5–L8
- 15. Kikuchi T, Moriya S, Nakatsuka Y, Matsuoka H, Nakazato K, Nishida A, Chakihara H, Matsuoka M, Moniwa M (2004) A new vertically stacked poly-Si MOSFET for 533 MHz high speed 64 mbit SRAM. Tech Dig IEDM: 923–926
- 16. Takato H, Sunouchi K, Okabe N, Nitayama A, Hieda K, Horiguchi F, Masuoka F (1991) Impact of surrounding gate transistor (SGT) for ultra-high-density LSI. IEEE Trans Electron Devices 38(3):573–578
- Zheng Y, Rivas C, Lake R, Alam K, Boykin TB, Klimeck G (2005) Electronic properties of silicon nanowires. IEEE Trans Electron Devices 52(6):1097–1103
- Miyano S, Hirose M, Masuoka F (1992) Numerical analysis of a cylindrical thin-pillar transistor (CYNTHIA). IEEE Trans Electron Devices 39(8):1876–1881
- Dwyer C, Vicci L, Poulton J, Erie D, Superfine R, Washburn S, Taylor RM (2004) The design of DNA self-assembled computing circuitry. IEEE Trans VLSI Syst 12(11):1214–1220
- Dwyer C, Vicci L, Taylor RM (2003) Performance simulation of nanoscale silicon rod fieldeffect transistor logic. IEEE Trans Nanotechnol 2(2):69–74
- 21. Sze SM (1981) Physics of semiconductor devices, 2nd edn. Wiley, New York

- 22. Choi Y-K, Ha D, King T-J, Bokor J (2003) Investigation of gate induced drain leakage (GIDL) current in thin body devices: single-gate ultrathin body, symmetrical double gate, and asymmetrical double gate MOSFETs. Jpn J Appl Phys 42:2073–2076
- Semenov O, Pradzynski A, Sachdev M (2002) Impact of gate induced drain leakage on overall leakage of submicrometer CMOS VLSI circuits. IEEE Trans Sem Manufac 15(1):9–18
- 24. Wettstein A, Schenk A, Fichtner W (2001) Quantum device-simulation with density gradient model on unstructured grids. IEEE Trans Electron Devices 48(2):279–283
- 25. Lombardi C, Manzini S, Saporito A, Vanzi M (1988) A physically based mobility model for numerical simulation of nonplanar devices. IEEE Trans Comput Aided Design Integr Circuits Syst 7(11):1164–1171
- Caughey DM, Thomas RE (1967) Carrier mobilities in silicon empirically related to doping and field. Proc IEEE 55(no. 12):2192–2193
- Arora ND, Hauser JR, Roulston DJ (1982) Electron and hole mobilities in silicon as a function of concentration and temperature. IEEE Trans Electron Devices ED-29:292–295
- 28. Serberherr S (1984) Process and device modeling for VLSI. Microelec Reliab 24(2):225-257
- Liu Z-H, Hu C, Huang J-H, Chan T-Y, Jeng M-C, Ko PK, Cheng YC (1993) Threshold voltage model for deep-submicrometer mosfets. IEEE Trans Electron Devices 40(1):86–95
- 30. Sery G, Borkar S, De V (2002) Life is CMOS: why chase the life after. Proc Design Autom Conf: 78–83
- Chiang TK (2005) New current–voltage model for surrounding-gate metal-oxide-semiconductor field effect transistors. Jpn J Appl Phys 44(9A):6446–6451
- Chen Q, Harrell EM, Meindl JD (2003) A physical short-channel threshold voltage model for undoped symmetric double-gate MOSFETs. IEEE Trans Electron Devices 50(7):1631–1637
- Boeuf F et al (2004) A conventional 45 nm CMOS node low-cost platform for general purpose and low power applications. Tech Dig IEDM: 425–428
- 34. Numata T et al (2004) Performance enhancement of partially and fully-depleted strained-SOI MOSFETs and characterization of strained Si device parameters. Tech Dig IEDM: 177–180
- 35. Shima A, Ashihara H, Hiraiwa A, Mine T, Goto Y (2005) Ultra-shallow junction formation by self-limiting LTP and its application to sub-65 nm node MOSFETs. IEEE Trans Electron Devices 52(6):1165–1171
- 36. Wang HCH et al (2004) Low power device technology with SiGe channel, HfSiON, and Poly-Si gate. Tech Dig IEDM: 161–164
- Luo Z et al (2004) High performance and low power transistors integrated in 65 nm bulk CMOS technology. Tech Dig IEDM: 661–664
- Lindert N, Choi YK, Chang L, Anderson E, Lee WC (2001) Quasiplanar NMOS FinFETs with sub-100 nm gate lengths. Proc Device Res Conf: 26–27
- Choi YK, Lindert N, Xuan P, Tang S, Ha D, Anderson E (2001) Sub-20 nm CMOS FinFET technologies. Tech Dig IEDM: 421–424
- Wakabayashi H et al (2004) Transport properties of sub-10 nm planar-bulk-CMOS devices. Tech Dig IEDM: 429–432
- 41. Bai P et al (2004) A 65 nm logic technology featuring 35 nm gate lengths, enhanced strain,
 8 Cu interconnect layers, low-k ILD and 0.57 µm SRAM cell. Tech Dig IEDM: 657–660
- 42. Kedzierski J, Dried DM, Nowak EJ, Kanarsky T, Rankin JH (2001) High performance symmetric-gate and CMOS compatible Vt asymmetric-gate FinFET devices. Tech Dig IEDM: 437–440
- 43. Yang FL, Chen HY, Chen FC, Chan YL (2001) 35 nm CMOS FinFETs. Proc Symp VLSI Tech: 104–105
- 44. Concannon A, Piccinini F, Mathewson A, Lombardi C (1995) The numerical simulation of substrate and gate currents in MOS and EPROMS. Tech Dig IEDM: 289–292
- 45. Dort MJV, Woerlee PH, Walker AJ (1994) A simple model for quantisation effects in heavilydoped silicon MOSFETs at inversion conditions. Solid State Electron 37(3):411–414

- 46. Srivastava N, Banerjee K (2004) A comparative scaling analysis of metallic and carbon nanotube interconnections for nanometer scale VLSI technologies. Proc 21st Int Multilevel Interconnect Conf: 393–398
- 47. Moon P, Dubin V, Johnston S, Leu J, Raol K, Wu C (2003) Process roadmap and challenges for metal barriers. Tech Dig IEDM: 841–844
- 48. Tada M et al (2003) A 65 nm-Node, Cu interconnect technology using porous SiOCH film (k = 2.5) covered with ultra-thin, low-k pore seal (k = 2.7). Tech Dig IEDM: 845–848
- 49. Nakai S et al (2003) A 65 nm CMOS technology with high performance and low-leakage transistor, a 6T-SRAM cell and robust hybrid-ULK/Cu interconnects for mobile multimedia applications. Tech Dig IEDM: 285–288
- 50. Kondo S, Yoon BU, Tokitoh S, Misawa K, Sone S, Shin HJ, Ohashi N, Kobayashi N (2003) Low-pressure CMP for 300-mm ultra low-k (k = 1.6 1.8)/Cu integration. Tech Dig IEDM: 151–154
- 51. Isobayashi A, Enomoto Y, Yamada H, Takahashi S, Kadomura S (2004) Thermally robust Cu interconnects with Cu–Ag alloy for sub 45 nm node. Tech Dig IEDM: 953–956
- 52. Alers GB, Sukamto J, Park S, Harm G, Reid J (2006) Containing the finite size effect in copper lines. Semicond Int 29(5):38–42
- 53. Alioto M, Palumbo G (2002) Analysis and comparison on full adder block in submicron technology. IEEE Trans VLSI Syst 10(6):806–823
- 54. Chang CH, Gu J, Zhang M (2005) A review of 0.18-µm full adder performances for tree structured arithmetic circuits. IEEE Trans VLSI Syst 13(6):686–695
- 55. Shams AM, Darwish TK, Bayoumi MA (2002) Performance analysis of low-power 1-Bit CMOS full adder cells. IEEE Trans VLSI Syst 10(1):20–29
- 56. Sayed M, Badawy W (2002) Performance analysis of single-bit full adder cells using 0.18, 0.25 and 0.35 μm CMOS technologies. IEEE Int Symp Circuits Sys: 559–562
- 57. Chang CH, Zhang M, Gu J (2003) A novel low power low voltage full adder cell. Proc 3rd Int Symp Image Signal Process Anal: 454–458
- 58. Khatibzade AA, Raahemifar K (2004) A study and comparison of full adder cells based on the standard static CMOS logic. Proc Can Conf Electr Comp Eng: 2139–2142
- 59. Goel S, Gollamudi S, Kumar A, Bayoumi M (2004) On the design of low-energy hybrid CMOS 1-Bit full adder cells. Proc 47th Midwest Symp Circuits Syst: 209–212