IET Journals
IET Computers & Digital Techniques

# Hybrid NEMS–CMOS integrated circuits: a novel strategy for energy-efficient designs

## H.F. Dadgour   K. Banerjee

*Electrical and Computer Engineering Department, University of California, Santa Barbara, 93106 CA, USA*
*E-mail: kaustav@ece.ucsb.edu*

**Abstract:** Substantial increase in gate and sub-threshold leakage of complementary metal-oxide-semiconductor (CMOS) devices is making it extremely challenging to achieve energy-efficient designs while continuing their scaling at the same pace as in the past few decades. Designers constantly sacrifice higher levels of performance to limit the ever-increasing leakage power consumption. One possible solution to tackle the leakage issue, which is proposed in this work, is to integrate nano-electro-mechanical switches (NEMS) with CMOS technology. Hybrid NEMS–CMOS technology takes advantage of both near-zero-leakage characteristics of NEMS devices along with high ON current of CMOS transistors. The feasibility of integration of NEMS switches into a CMOS process is illustrated by a practical process flow. Moreover, co-design of hybrid NEMS–CMOS as low-power dynamic OR gates, static random access memory (SRAM) cells and sleep transistors is explored. Simulation results indicate that such hybrid dynamic OR gates can achieve 60–80% lower switching power and almost zero-leakage power consumption with minor delay penalty. However, the hybrid OR gate outperforms its CMOS counterpart both in terms of delay and switching power consumption with increase in fan-in beyond 12. Additionally, it is shown that a hybrid NEMS–CMOS SRAM cell can achieve almost $8\times$ lower standby leakage power consumption with only minor noise margin and latency cost. Finally, application of NEMS devices as sleep transistors results in up to three orders of magnitude lower OFF current with negligible performance degradation as compared to CMOS sleep switches.

# 1   Introduction

During the past two decades, CMOS integrated circuits (ICs) have witnessed unprecedented improvements in their functionality and performance. This was primarily achieved by aggressive technology scaling, which resulted in device density and performance doubling roughly every 18 months as per Moore's law [1] while achieving a remarkable 25% per year improvement in cost per chip function. As CMOS scaled from generation to generation, power dissipation increased proportionately to increasing transistor density and switching speeds.

However, with the minimum feature size of the transistor entering the sub-100 nanometre regime, power dissipation is increasing ominously especially because of a substantial increase in the (sub-threshold) leakage power. Sub-threshold leakage power used to be insignificant for earlier generations of ICs but it is becoming an increasing fraction of the total

power [2, 3]. The increase in sub-threshold leakage power arises due to the fact that power supply ($V_{dd}$) scaling necessitates threshold voltage ($V_{th}$) scaling. This trend, which is forecasted by the International Technology Roadmap for Semiconductors (ITRS) [4], is shown in Fig. 1. Gate leakage also increases significantly as a result of scaling down the oxide thickness. The relative importance of gate and sub-threshold leakage power compared to total power consumption is illustrated in Fig. 2 [5]. Moreover, most leakage mechanisms are strongly temperature dependent. This strong coupling between temperature and leakage can cause further increase in total power dissipation [6].

## 1.1   Theoretical limit of sub-threshold performance of CMOS devices

The characteristics of an ideal switch can be described as one having abrupt transition between ON and OFF states with high ON current and zero OFF current. However, all
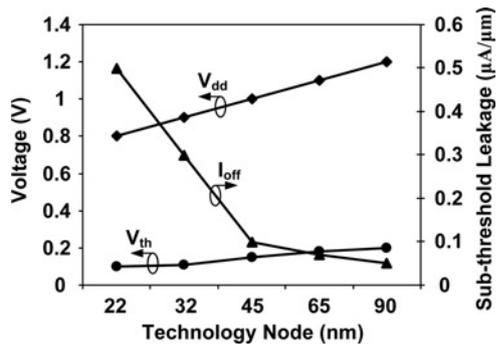
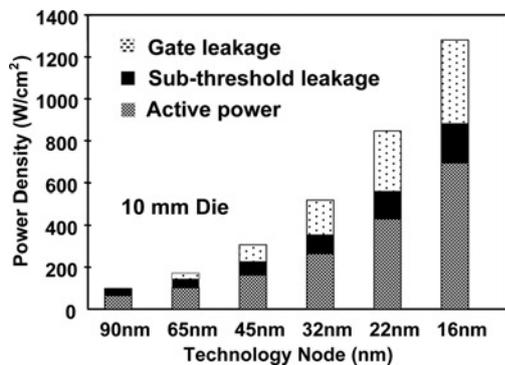**Figure 1** *CMOS technology scaling trends as predicted by ITRS [4]*



**Figure 2** *Contributions of active, sub-threshold and gate leakage power consumption towards total power consumption of chips [5]*

solid-state switches have non-zero sub-threshold current and show gradual switching behaviour (Fig. 3). Inverse of the slope of $\log(I_d) - V_{gs}$ characteristics of solid-state devices, which is called sub-threshold swing, is used as a metric to identify abruptness of switching. Sub-threshold swing ($S$) of a device essentially indicates the amount of gate voltage reduction necessary to reduce the sub-threshold current by one decade ($S = dV_{gs}/d\log I_d$) [7] as shown in Fig. 3.
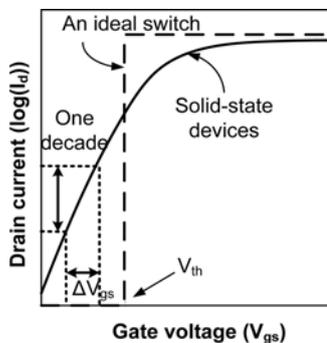


**Figure 3** *Comparison between switching characteristics of an ideal switch and that of solid-state devices*

Slope of the log $I_d - V_{gs}$ curve is defined as the sub-threshold swing; therefore we have $S = dV_{gs}/d\log I_d = \Delta V_{gs}$. This figure has not been drawn to scale

Sub-threshold swing can be analytically calculated for each type of devices. In the case of CMOS devices, sub-threshold leakage current is dominated by diffusion mechanism. It has been shown that in the sub-threshold region, $I_d - V_{gs}$ characteristics of an MOS device can be modelled by (1) [7]

$$I_{ds} = \mu_{eff} C_{ox} \frac{W}{L} (m-1) \left(\frac{kT}{q}\right)^2 e^{q(V_{gs} - V_t)/mkT} \left(1 - e^{-qV_{ds}/kT}\right)$$

(1)

Therefore sub-threshold swing can be derived as follows

$$S = \left(\frac{d(\log_{10} I_{ds})}{dV_{gs}}\right)^{-1} = \left(\frac{\partial V_{gs}}{\partial \psi_s}\right) \left(\frac{\partial \psi_s}{\Delta(\log_{10} I_{ds})}\right)$$

$$= \ln(10) \frac{kT}{q} \left(1 + \frac{C_{dm}}{C_{ox}}\right)$$

(2)

where $\psi_s$ is the surface potential and $C_{dm}$ is the bulk depletion capacitance. Assuming that $C_{dm}$ is zero, one can calculate the theoretical limit for sub-threshold swing of MOS devices from (2) as $S = 2.3 \times 26$ mV/decade $\simeq 60$ mV/decade.

Since in reality, $C_{dm}$ is not zero, sub-threshold swing values are typically around 90 mV/decade for bulk CMOS devices as shown in Fig. 4, where minimum reported sub-threshold swings for various devices are summarised [8–13]. In this figure, $S$ values for non-classical CMOS-based transistors, such as fully depleted SOI (FDSOI) and FinFET devices are also shown. Although conduction in FDSOI and FinFET devices are also based on drift/diffusion mechanisms, their sub-threshold swing is closer to the theoretical limit of 60 mV/decade. The reason is that these types of devices have very low bulk depletion capacitance ($C_{dm} \simeq 0$) because of their un-doped body and hence, according to (2), value of $S$ can be closer to 60 mV/decade.
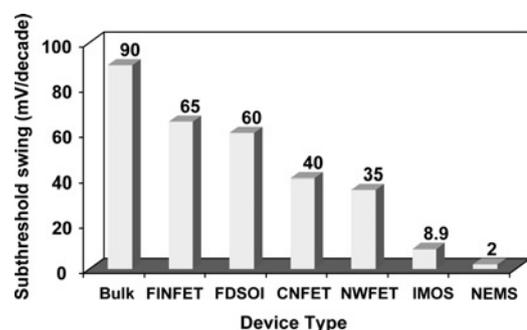


**Figure 4** *Minimum sub-threshold swings reported in the literature for various emerging device structures*

The value reported here for the CNT FET corresponds to the tunnelling-CNFETs, which have the lowest sub-threshold swing

## 1.2 Conduction mechanisms of CMOS and low sub-threshold swing devices

Different device alternatives have been proposed in the literature to achieve sub-threshold swing values lower than 60 mV/decade. As shown in Fig. 4, tunnelling-type carbon nano-tube transistors (T-CNFET), nano-wire-based transistors (NWFET) and impact-ionisation-based MOS (IMOS) show sub-threshold slopes of 40, 35 and 8.9 mV/ decade, respectively. However, it has been shown that tunnelling-based devices have relatively lower ON current [9] and impact-ionisation-based devices require very high $V_{dd}$, which make them less attractive alternatives. Recently, negative gate capacitance ($C_{ox}$ in (2)) has been proposed to lower the value of $S$ below 60 mV/decade [14], but this approach is still under investigation.

On the other hand, it has been experimentally shown that an electromechanical FET exhibits astonishingly low sub-threshold slope of 2 mV/decade [13]. The details of their operation will be discussed in Section 2. In Fig. 5, band diagram of conventional CMOS is illustrated along with those of tunnelling, ionisation and NEMS-based devices in both ON (saturation) and OFF states. In this figure, only n-type devices are illustrated and hence, carriers are always considered to be electrons.

Current conduction in CMOS devices is based on drift and diffusion; however, to move from source to drain, an electron should obtain enough kinetic energy to go over the source-channel energy barrier ($\Delta E_{ON}$ or $\Delta E_{OFF}$) as shown in Fig. 5a. In the OFF state, because of small gate bias, energy barrier is higher ($\Delta E_{OFF}$) so that fewer carriers can obtain enough energy to go over it and the transport is dominated by diffusion. In the ON state, the barrier can be lowered to $\Delta E_{ON}$ by applying positive voltage to the gate and hence, conduction of current increases. In this case, the transport is dominated by drift.

In tunnelling-based transistors, conduction happens because of a phenomenon called band-to-band tunnelling where carriers actually can tunnel through the energy barrier instead of going over it. This phenomenon occurs only when distance between two energy bands ($d_{ON}$ in Fig. 5b) is very small. Amount of current is an exponential function of inverse of the horizontal distance between two bands in the figure ($d_{ON}$).

In the OFF state, by removing the gate bias, two bands become far apart ($d_{OFF}$) and therefore current can be significantly reduced. Since tunnelling current is a strong function of band-to-band distance, tunnelling-based devices can achieve nearly abrupt switching between ON/OFF states and low sub-threshold swing.

Impact-ionisation-based devices employ a phenomenon called avalanche breakdown. These devices are designed in a way that electrons can obtain elevated energy because of
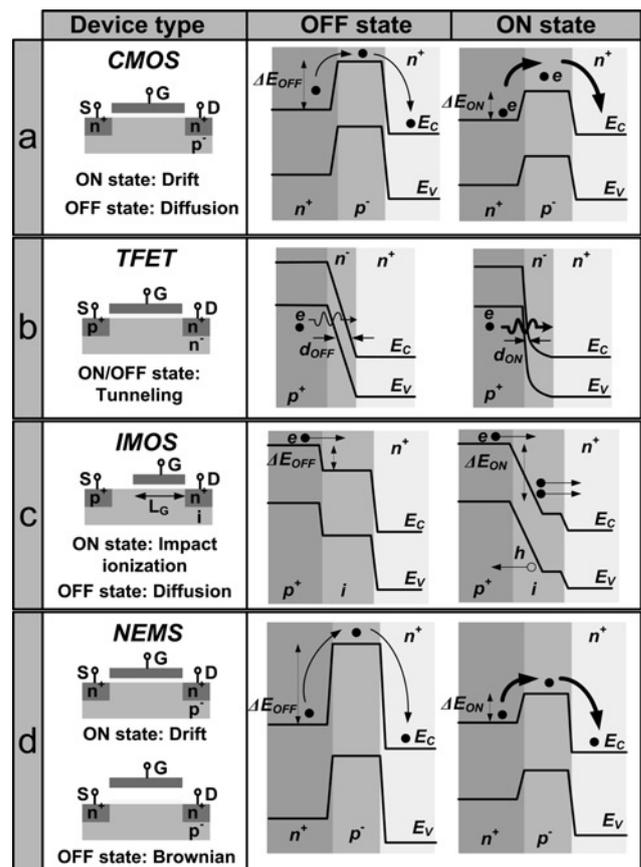


**Figure 5** *Different conduction mechanisms*
*a* CMOS device uses drift/diffusion mechanism
*b* Tunnelling transistors employ band-to-band tunnelling mechanism
*c* Impact-ionisation-based devices use avalanche breakdown mechanism
*d* NEMS devices (in this case, a suspended-gate FET is assumed) employ drift mechanism in ON state and there is no conduction mechanism in OFF state (except for Brownian motion)

the high electric field. When these high-energy electrons collide with lattice atoms, they can ionise them and thereby, generate more electron–hole pairs, which further contribute to current in a chain reaction fashion. This phenomenon is depicted in Fig. 5c where an electron travels from a very high-energy state to a lower one and generates an extra electron–hole pair. Impact-ionisation-based devices can also deliver very low sub-threshold swing because of sharp switching between ON and OFF states caused by inherent positive feedback in the avalanche breakdown phenomenon [15].

Another attractive approach to achieve ultra-low sub-threshold swing values is to employ a nano-electro-mechanical switch (NEMS) that exhibits nearly zero conduction in the OFF state because of 'physical' separation between the components in the switch. In particular, a nano-electro-mechanical field effect transistor (NEMFET) device combines an electro-mechanically controlled part (gate) and a solid-state part (source–substrate–drain) such

595

that in the OFF state, as shown in Fig. 5d, gate is not able to lower the energy barrier because of the existence of an air gap between the gate electrode and the gate dielectric material. In this case, two terminals are separated by an insulator and hence, OFF-state leakage is only limited to Brownian motion displacement current and vacuum tunnelling currents [16, 17]. However, application of a large enough positive voltage initiates a self-enforcing mechanism that very quickly pulls the suspended gate all the way down, touching the dielectric layer beneath it, thereby creating a conduction channel between source and drain. As a result, in the ON state, gate is able to lower the energy barrier and facilitate the conduction of current. This abrupt switching between ON/OFF states results in very small sub-threshold swing (Fig. 5d).

## 1.3 Gate leakage

Gate leakage is also becoming a critical issue in sub-65 nm regime. According to an ITRS prediction, gate leakage will even surpass sub-threshold leakage in 32 nm technology node (see Fig. 2). Metal gate over high-$k$ dielectric materials (MG/HK) has been proposed to counter gate leakage problem. However, there are challenges associated with gate work function engineering and the choice of metal gate materials for high-$k$ dielectrics [18, 19]. NEMFET devices offer lower gate leakage compared to the CMOS transistors in the OFF state because of the existence of an air gap between gate and dielectric material; however, in the ON state gate leakage remains unaffected [13]. It is important to note that even in the ON state, although the air gap vanishes completely, gate leakage is still lower than that of a CMOS device with similar dielectric thickness. The reason is that the actual voltage drop across the dielectric layer of a NEMFET is less compared to the voltage difference that appears across the dielectric layer in the CMOS device. This issue will be discussed in more detail in Section 2.2.

## 1.4 Scope of this work

Although NEMS devices have extremely low OFF current and sub-threshold swing, a NEMS device does not offer as high ON current as CMOS transistors do. Therefore we have recently proposed the idea of hybrid NEMS–CMOS technology to combine near-zero-leakage characteristics of NEMS with high ON current of CMOS transistors to simultaneously achieve ultra-low-power and high-performance operation [20].

The hybrid NEMS–CMOS technology can potentially revolutionise the IC industry by shifting the IC power trends to a new (and much lower) plane (much like what CMOS did to the bipolar junction transistor (BJT) power trend). Although NEMS based on nanometre-scale semiconductor structures (including silicon, silicon-on-insulator, GaAs/AlGaAs systems, SiC on Si, aluminium nitride on Si as well as carbon nanotubes (CNTs)) have been reported [21–24], there is no report in the literature

regarding co-design and integration of such NEMS structures with CMOS circuits and technology.

We believe that integrating NEMS devices into the nanoscale CMOS process can have significant implications for the IC industry. Certainly, there will be some challenges in integration of NEMS devices with CMOS; however, integration of mechanical devices with CMOS circuits has already been accomplished at the micron-scale [25, 26]. For example, a chip consisting of numerous arrays of micro-electro-mechanical systems (MEMS)-based micro-mirrors, known as a digital micro-mirror device (DMD), located on top of a CMOS-based SRAM chip has been commercially fabricated [25]. Each mirror can be rotated by loading the required state into an SRAM cell located beneath it. Furthermore, integration of a MEMS resonator with bulk N-well CMOS process has also been demonstrated [26].

To explore potential applications of hybrid NEMS–CMOS circuits, we have focused on three most critical applications, namely dynamic OR gate, SRAM cell and sleep transistor design, which are known to be most affected by increasing leakage [2]. HSPICE simulations are carried out to compare NEMS–CMOS circuits against their pure-CMOS counterparts. Device models for the NEMFET (based on a suspended-gate FET or simply SG-FET) structures were generated using equivalent circuit models that were calibrated against device simulation results for SG-FETs with effective channel length of 45 nm [27]. BSIM models [28] were employed for CMOS transistors. Simulation results suggest that a hybrid dynamic OR gate can achieve 60–80% lower switching power and near-zero-leakage power consumption with minor delay penalty and a hybrid SRAM cell can achieve 7.7× lower standby leakage power consumption with only 14 and 23% penalty on noise margin and latency, respectively. Also, in the sleep transistor application, NEMS devices exhibit up to three orders of magnitude lower leakage current with minor performance degradation compared to the CMOS sleep transistors.

In summary, we highlight in this paper

1. The concept of hybrid NEMS–CMOS circuits for lowering leakage and achieving energy-efficient designs in nano-scaled CMOS technologies and a simplified process flow for their fabrication.

2. The design and analysis of low-power and high-performance dynamic OR gates, SRAM cells and sleep transistors using hybrid NEMS–CMOS technology.

This paper is organised as follows: in Section 2 a brief introduction is provided on NEMS including its basic operation and modelling. Section 3 describes an integrated fabrication process for NEMS and CMOS devices. Section 4 includes proposed dynamic OR gate architecture and

corresponding simulation results. In Section 5, a novel NEMS–CMOS-based SRAM cell architecture is proposed and compared against existing works in the literature. Section 6 discusses the application of NEMS devices as sleep transistors and finally concluding remarks are made in Section 7.

# 2 Nano-electro-mechanical switches

MEMS have been fabricated for decades and used for different aerospace, electrical and bioengineering applications. In recent years, more advanced fabrication techniques have been employed to further miniaturise MEMS into the nano-scale regime. These types of structures have been alternatively called nano-electro-mechanical switches (NEMS). NEMS can be employed as switching devices in ICs where flow of current between source and drain terminals can be controlled using interaction of electrical and mechanical means. These devices are usually composed of a moveable beam that can deform in response to an applied electrical bias; thereby aiding or blocking the flow of electrical current.

In Section 2.1, we first demonstrate the possibility of having mechanical switches operating in multi-GHz range and then discuss operation of a suspended-gate based NEMS device (SG-FET) in Section 2.2. In Section 2.3, we discuss the low-leakage characteristics of NEMS devices and then in Section 2.4, we briefly discuss an alternative approach to fabricate NEMFETs, using CNT-based NEMS devices. Since suspended-gate devices are easier to integrate in standard CMOS process (as will be shown later), in this work, we focus on the analysis of SG-FET-based circuit architectures. Therefore in Section 2.5, we discuss HSPICE modelling of suspended-gate devices.

## 2.1 High-frequency operation of NEMS devices

Interestingly, scaling down of mechanical systems into the nanometre regime has very important implications for ICs: mechanical switching speed of NEMS can match that of CMOS transistors. The reason is that intrinsic delay of a mechanical switch is limited by the mass of the moving segment of NEMS device (that scales as cube of the dimensions) whereas switching speed of a CMOS device is restricted by the amount of time that is required for charge to be relocated inside the device to form the conduction channel (that scales as square of the dimensions).

For instance, Fig. 6 shows a scanning electron microscope (SEM) picture of an array of silicon wire resonators. These wires can resonate at very high frequencies and the resonant frequency of each wire is a function of its length. The $\sim 2$-$\mu$m long wire, for example, has a resonant frequency of $\sim 400$ MHz [29]. Since resonant frequency increases
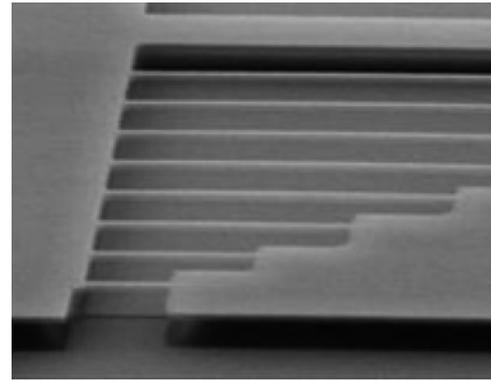


**Figure 6** *Electron micrograph of a series of silicon nanowire-based NEMS resonators fabricated in single-crystal silicon [29]*

rapidly by reducing the length of the wire, GHz operation can be easily guaranteed once nano-scaled wires are fabricated.

It should be noted that the mechanical delay of the beam is equal to the inverse of resonant frequency of the beam, which in turn is equal to $\sqrt{K/m}$, where $K$ is the spring constant of the gate (inverse of its stiffness) and $m$ is the mass of the suspended gate. The spring constant ($K$) itself can be expressed as $K = (16Ewh^3/l^3)$, where $E$ is the Young's modulus; $l$, $h$ and $w$ are the length, thickness and width of the beam, respectively. Also, the mass of the beam can be expressed as $m = lwh\rho$, where $\rho$ is the mass density of the beam [29]. Substituting these values in the resonant frequency formula ($\sqrt{K/m}$), one arrives at

$$f = \frac{h}{l^2}\sqrt{\frac{16E}{\rho}} \qquad (3)$$

According to this formula, the resonant frequency is proportional to $1/l^2$; it increases dramatically as the device is scaled down. Since both $K$ and $m$ are proportional to the width of the beam ($w$), resonant frequency of the beam (inverse of its intrinsic delay) is independent of its width. Fig. 7 plots resonant frequency and intrinsic delay of suspended-gate devices with different beam lengths, where thickness ($h$) is assumed to be 5 or 10 nm.

Note that a fraction of the length of the beam, which resides between source and drain terminals, is the width of NEMS device considered in this work (Fig. 9), which will be discussed in the following subsection. It can be observed that for device lengths less than 40 nm, intrinsic delays can be as low as 10 ps. Although this value is higher than the intrinsic delay of the same-sized CMOS devices ($\simeq 1$–$2$ ps), it is possible for suspended-gate-based circuits to be employed in high-performance applications. The reason is that the total gate delay is equal to the sum of the intrinsic
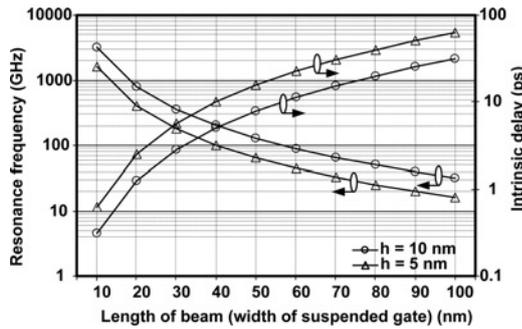
**Figure 7** *Resonant frequency and intrinsic delay of a suspended-gate device are function of its length (l), from (3).*

Smaller devices exhibit much lower intrinsic delay. In this figure, Young's modulus (*E*) and mass density (*ρ*) of silicon are assumed to be 150 GPa and 2330 kg/m$^3$, respectively. The thickness of the beam (*h*) is 5 nm (triangle) for one set of curves and 10 nm (circle) for the other

delay and the delay associated with charging/discharging of the load capacitance. On the other hand, for logic gates that always drive a capacitive load (fan-out > 0), the contribution of the load delay to the total delay is generally much higher than that of the intrinsic delay. Therefore although the intrinsic delay of NEMS-based gates is relatively high compared to that of CMOS gates, the overall delay of a hybrid circuit can end up being close (though higher) to that of CMOS circuits, as will also be shown by our simulations results.

For example, Fig. 8 exhibits the total delay of NEMS- and CMOS-based inverters for different values of fan-out, where intrinsic delay of each gate corresponds to its delay when fan-out = 0 (zero load delay). It can be observed that the difference between the intrinsic delay values ($\Delta d_1$) is much higher than the difference between total delays of the two gates when their fan-out is 5 ($\Delta d_2$).
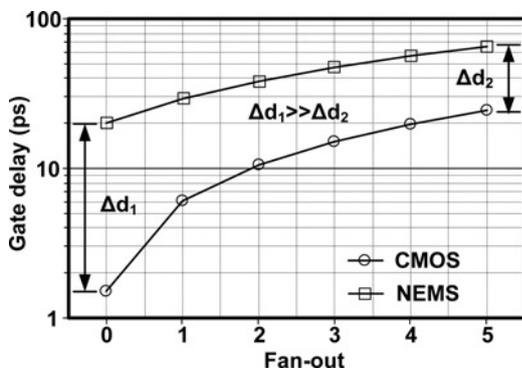


**Figure 8** *Total gate delay (intrinsic + load delay) of NEMS- and CMOS-based inverters for different values of fan-out*

The difference between intrinsic delay of NEMS and CMOS inverters ($\Delta d_1$) is high. However, by increasing the fan-out value, this difference decreases dramatically ($\Delta d_1 \gg \Delta d_2$)

## 2.2 Suspended-gate NEMS device

The first NEMFET device, in the form of a suspended-gate metal-oxide-semiconductor field-effect transistor (SG-MOSFET), was proposed in [30] and then fabricated in [13]. An SEM micrograph of the fabricated SG-MOSFET device is shown in Fig. 9. In this device, the gate is physically suspended with four supporting arms (sides) over the channel area. To better understand the device operation, in Fig. 10, a cross-sectional view of the device in Fig. 9 is shown (along the broken lines in Fig. 9).

Figs. 10*a* and *b* illustrate the position of the suspended gate in both ON and OFF states according to cross-section #1 and cross-section #2 in Fig. 9. As shown in Fig. 10, in the absence of a gate bias, an air gap exists between the gate material and the gate dielectric layer. Hence, there is no conduction channel; however, when sufficient gate voltage is applied to the gate terminal, the suspended-gate bends over from its original straight position. As a result, the gate touches the underlying dielectric layer; attracting carriers of opposite charge and forming a conducting channel between source and drain in the underlying silicon substrate.

Since the operation of an SG-MOSFET involves both mechanical and electrical interactions, we discuss the device behaviour in ON and OFF states in more detail considering relation between mechanical and electrical components. In Fig. 11, the suspended gate is connected to
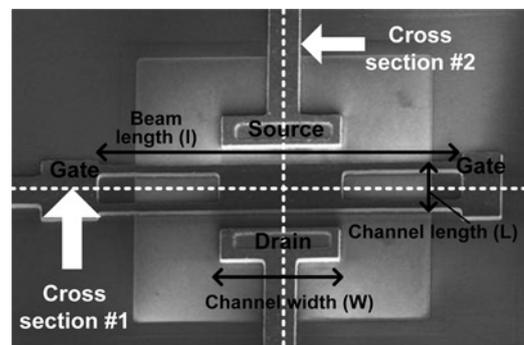


**Figure 9** *SEM micrograph of an SG-MOSFET switch with a 20 μm wide (=W) and 10 μm long (=L) suspended gate with four arms (source: [13])*
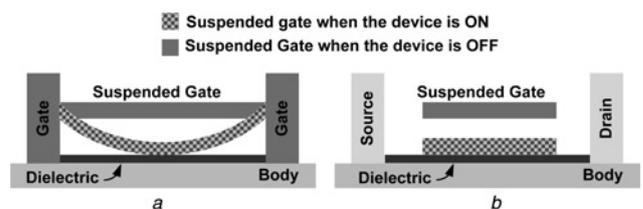


**Figure 10** *Two cross-sectional views of the suspended gate of Fig. 9*

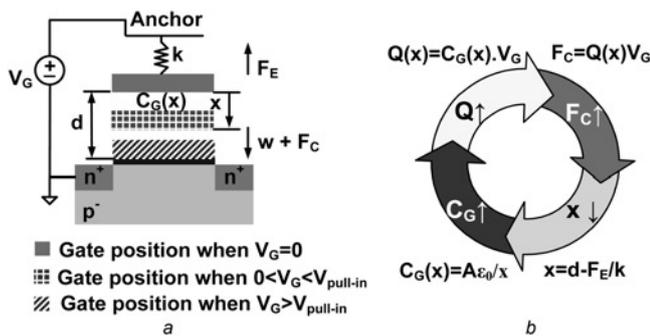*a* Along cross-section #1
*b* Along cross-section #2

**Figure 11** *Basic operation of a suspended-gate NEMFET in ON and OFF states*

*a* Position of the suspended gate in ON/OFF states
*b* Positive feedback involved in SG-MOSFET

an anchor through a spring; however, it should be noted that there is no physical anchor or spring in the device structure and those concepts are solely used to model and demonstrate the movable suspended gate (which is essentially a bendable beam).

When no bias voltage is applied to the device ($V_G = 0$), the suspended gate is separated from the gate dielectric layer by the air gap of size $d$ and elastic force ($F_E$) of the suspended gate is in equilibrium with its weight ($w$), and there is no channel formed inside silicon (OFF state).

As we apply a small positive gate voltage, charge appears on both plates of the capacitor, $C_G(x)$, which is composed of the suspended gate and a thin layer inside silicon. This creates an electrostatic force ($F_C$) between positive charges in the gate and negative ones in the silicon layer, which causes the suspended gate to move downward in the $x$-direction. Once the distance between the two plates of the capacitor decreases, $C_G(x)$ increases and hence, more charge appears on its plates and causes the suspended gate to move further downward. It is possible to prove through mathematical analysis that this mechanical system is inherently unstable for $V_g > V_{\text{pull-in}}$, where $V_{\text{pull-in}}$ is called pull-in voltage (equivalent to the threshold voltage of CMOS transistors) [13]. This means that once the gate voltage is larger than $V_{\text{pull-in}}$, the only stable solution for the system is when the gate collapses and the air gap vanishes. In other words, when $V_G > V_{\text{pull-in}}$, sum of the electrostatic and the weight forces ($w + F_C$) becomes larger than the maximum elastic force that can be supported by elasticity ($F_E$) of the suspended gate and hence, the suspended gate is pulled all the way down until the air gap vanishes (ON state). On the other hand, when $V_g < V_{\text{pull-in}}$, the system is not inherently unstable and can have stable solution, which corresponds to having an air gap smaller than the original air gap at $V_g = 0$, and the suspended gate settles in a position closer to the dielectric ($d-x$).

It should also be noted that the elastic force ($F_E$) exists whenever the gate (beam) is deflected. No matter whether

the device is ON or OFF, as long as the beam is deformed, there must be an equal force in the opposite direction to compensate for it. Therefore after gate collapses, charges (and hence, voltage) are required to support the bent gate. It should be noted that if this was not the case, in the ON state, current of a same-sized NEMS and CMOS devices would be equal. (It is obvious from device simulations [20] that ON current is lower for NEMS devices.) Since in the ON state, a fraction of gate voltage is used up to keep the suspended gate bent, effective gate voltage that appears across the gate dielectric layer is less than the actual gate voltage. Hence, ON current of an SG-FET is lower than that of CMOS devices for a given $V_g$. The voltage difference necessary for deforming the suspended gate is modelled as a voltage-controlled voltage source in Section 2.5.

## 2.3 Low leakage characteristics of NEMS

Unlike leakage current across source and drain terminals in solid-state devices, the OFF-state leakage current in a NEMS device is determined only by the Brownian motion displacement current and vacuum tunnelling currents [16, 17]. The reason is that in OFF state, the gate terminal has no control over the channel area because of the existence of an air gap (as will be discussed in the following section) and hence, it is as if the source and drain terminals are separated by a piece of lightly doped semiconductor. Since conductivity of lightly doped Si is extremely low, leakage current in the OFF state is limited only to the Brownian motion displacement current (Fig. 5d). Note that this is not the case for CMOS devices where even in the sub-threshold region of operation, the gate terminal retains weak control over the channel. Moreover, during the OFF state, there is no physical contact between gate terminal and the dielectric layer below it, which confines the gate leakage to only vacuum tunnelling currents.

## 2.4 CNT-based NEMS

An alternative approach for fabricating NEMS switches is CNT-based cantilever NEMFET as shown in Fig. 12 [31]. The electro-mechanical principle of operation is similar; however, instead of a moveable gate, the conducting channel between the source and drain is made moveable. In this structure, source terminal is connected to a conductive bendable cantilever (or a CNT), which is
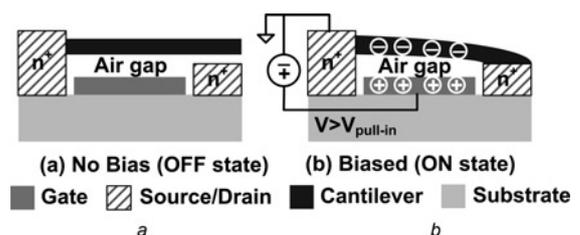


**Figure 12** *CNT-based cantilever NEMFET*
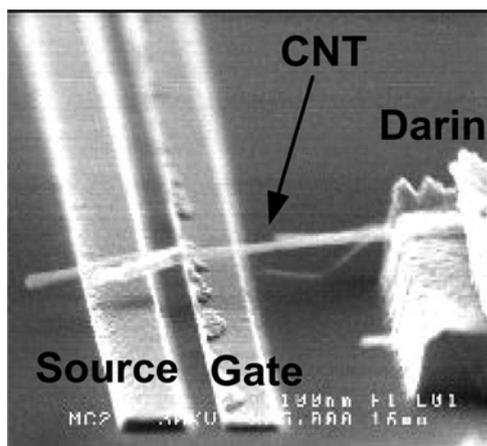
*a* OFF state
*b* ON state

**Figure 13** *SEM picture of nanotube NEMS device [32]*

suspended over the gate terminal. In the OFF state, as shown in Fig. 12a, there is no connection between the drain and source terminals. However, if sufficient voltage is applied between the gate and source, the cantilever can deform and make a connection between the source and drain terminals (Fig. 12b). An SEM picture of a fabricated CNT-based relay is shown in Fig. 13 [32]. It should be noted that integration of CNT-based electromechanical devices in standard CMOS fabrication could be more challenging than incorporation of the SG-FET devices in the CMOS process. Recently, we have presented a detailed study of the scaling analysis of CNT-NEMS devices, highlighting the manufacturability issues of CNT-based NEMS structures [33].

## 2.5 Modelling of SG-FET devices

Accurate modelling of NEMS devices is essential for performance comparison as well as co-design with existing CMOS technology. Precise modelling of NEMS devices is particularly complicated because of the interaction between electrical and mechanical components. However, using duality between electrical and mechanical systems, an entirely electrical SPICE model has been proposed in [34] for NEMS devices. These two kinds of systems are dual because of the similarity of the differential equations that govern electrical and mechanical systems as shown in Fig. 14a. Considering these equations, for each variable in the mechanical system, a dual variable can be identified in the electrical system. For example, the last row in this figure shows that force ($f$), mass ($m$) and velocity ($v$) are related by Newton's second law of motion $f = m\,dv/dt$ and voltage ($V$), current ($I$) and inductance ($L$) are related by $V = L\,dI/dt$. Hence, it can be argued that force, mass and velocity are dual variables of voltage, inductance and current, respectively. Dual variables are also shown in Fig. 14b.

In this model, each mechanical variable (shown in Fig. 15a) is replaced with its electrical equivalent, which leads to the SPICE model presented in Fig. 15b. The damping factor ($c$) of the suspended gate is modelled with



**Figure 14** *Analogy between mechanical and electrical systems*
*a* Analogous basic equations for both systems
*b* Dual parameters for those systems

a resistance ($R$) and mass of the moveable gate ($m$) is replaced with an inductance ($L$). Additionally, different forces ($f$), which pull the suspended gate towards the substrate, are modelled with a voltage-controlled voltage source ($f(V_g)$). Since $f(V_g)$ is a complicated function of several parameters, a polynomial approximation is used through curve fitting [34].

As mentioned earlier, the switching delay of NEMS devices consists of two parts: intrinsic and load delay. The intrinsic delay is mainly associated with the mechanical delay of the device (time required for the NEMS device to become deformed and touch the underlying dielectric layer). Load delay is the delay corresponding to the time required for charging/discharging of the load capacitances. Both intrinsic and load delays are included in Fig. 15. Intrinsic (mechanical) delay is accounted for by including two components: $R$ (which resembles mechanical damping of the beam) and $L$ (which models the mass of the beam). Therefore propagation delay of signal from the node $V_g$ through $R$ and $L$ elements is equivalent to considering mechanical delay of the NEMS device. The load delay is also considered by including a traditional CMOS device in the model (Table 1).

The power consumption because of mechanical movement of the beam is also taken into account in this model. Electrical power consumption by the two components ($R$ and $L$) resembles equivalent power consumption in the
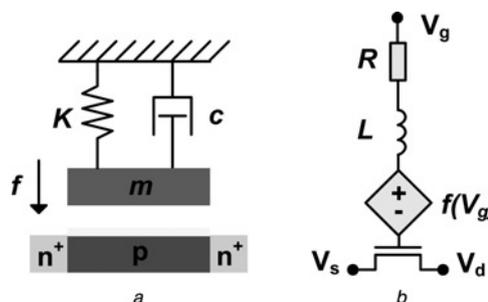


**Figure 15** *SG-MOSFET modelling*
*a* Mechanical variables of a suspended-gate structure
*b* Equivalent electrical SPICE model of the suspended-gate device

**Table 1** Model parameters of Fig. 15 used in this work

| Parameter | Value | Description |
|---|---|---|
| $L$ | 12 fH | mass of the beam |
| $R$ | 250 k$\Omega$ | damping factor of the beam |
| $f(V_g)$ | $0.4V_g - 2V_g^2$ | fitting parameter of the model |

The $f(V_g)$ is obtained during model calibration with a similar approach as in [34]; $L$ and $R$ values are calculated based on formulas presented in [35].

mechanical system because of physical movement of the movable gate. However, it should also be noted that power consumption because of mechanical movement of the beam is negligible. The reason is that the mass of the beam is very small in nano-scaled devices and hence, as shown in [20], mechanical power consumption can be as low as only 6% of the total power consumption.

According to this model, ON current of the NEMS device is lower than that of an identically sized CMOS device because actual gate voltage for the NEMS transistor is smaller than $V_g$ due to the voltage drop caused by $f(V_g)$ (Fig. 15$b$). This electrical model is then calibrated with reported data from nano-scaled SG-MOSFET device simulations [27]. Reported values in [27] are for SG-MOSFET devices with effective channel length of 45 nm, which is comparable to that of a CMOS device fabricated in 90 nm technology node. It should be noted that although the device in [27] is a depleted mode (normally ON) suspended-gate SOI device, its operation is exactly similar to the enhancement mode (normally OFF) devices employed in this work. Therefore the only difference between the two devices is that the device from [27] is ON when a low voltage is applied to its gate. This makes no difference in the overall functionality if one uses the inverted version of signals for the depleted mode NEMS devices. For example, if a CMOS device is connected to signal '$A$', its NEMS counterpart must be connected to the inverse of $A$, and the operation of the circuit will be accurate.

The $I_{ON}$ and $I_{OFF}$ values, which were used for the calibration of the NEMS model, along with $I_{ON}$ and $I_{OFF}$ parameters of 90 nm CMOS devices are summarised in Table 2. The calibration procedure is simple because there is only one fitting parameter, which is $f(V_g)$, in the model. That is because for each gate voltage value, its corresponding current value is known from the simulated $I-V$ curve [27] and also, $R$ and $L$ parameters can be

**Table 2** $I_{ON}$ and $I_{OFF}$ values for NEMS and CMOS devices

| Device | $I_{ON}$ | $I_{OFF}$ |
|---|---|---|
| CMOS [4] | 1110 $\mu$A/$\mu$m | 50 nA/$\mu$m |
| NEMS [27] | 330 $\mu$A/$\mu$m | 110 pA/$\mu$m |

calculated as they are only function of the physical properties of the suspended-gate device. To calibrate the model, for each point on the $I-V$ curve (say ($I_1$, $V_1$)), that particular gate voltage ($V_1$) should be applied to the node $V_g$ of the model (as shown in Fig. 15$b$) and then, $f(V_g)$ must be altered until the drain-to-source current of the model circuit matches the current of the simulated device ($I_1$). In all simulations reported in the following sections, the 90 nm BSIM [28] models were used for the CMOS devices and the HSPICE model of Fig. 15$b$ was used for the same-sized NEMS transistors.

It should be noted that $I-V$ characteristics of the suspend-gate device inherently exhibit hysteresis. It means that its $I-V$ curve is path dependent and as a result, gate voltage ($V_g$), which makes the suspended-gate device switch from OFF to ON state ($V_{pull-in}$), is not equal to the voltage level required for toggling it from ON to OFF ($V_{pull-out}$). The model presented in Fig. 15 does not account for such a path dependency of $I-V$ characteristics. Note that using this model does not affect the operation of the circuits proposed in this work. However, it is possible to use this property of NEMS devices in the design of novel circuit structures such as Schmitt Triggers.

## 3 Fabrication feasibility of hybrid NEMS–CMOS circuits

A simplified fabrication process depicted in Fig. 16 is proposed to fabricate the necessary components of the NEMS switch along with CMOS transistors. In this figure, we combined fabrication steps of SG-MOSFET devices as reported in [13] with processing steps of the standard CMOS technology. The first step is to define polysilicon gates of CMOS devices (Fig. 16$a$). The next step, as shown in Fig. 16$b$, is self-aligned definition of
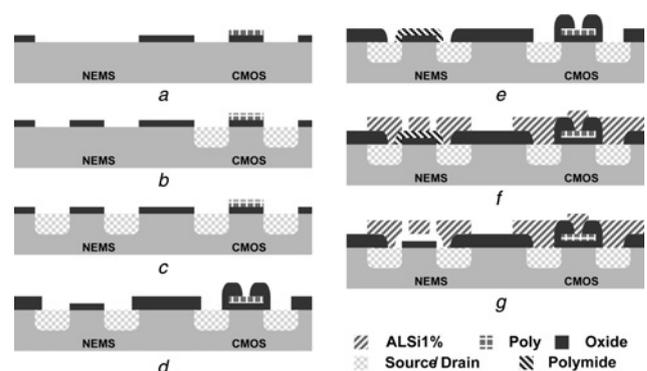


**Figure 16** Simplified process flow for fabrication of hybrid NEMS–CMOS circuits

$a$ Formation of poly gate for CMOS
$b$ Definition of active area for CMOS
$c$ Definition of active area for NEMS
$d$ Growth of thick oxide layer
$e$ Sacrificial layer deposition
$f$ AlSi metalisation and dry etching
$g$ Release of metal gate

source and drain area for CMOS devices. It should be noted that since NEMFET employs a suspended gate, its active area cannot be fabricated in a self-aligned fashion. Therefore in the third step (Fig. 16c), source and drain regions must be implanted for the NEMS devices. Then the thick-field oxide layer can be grown (Fig. 16d).

The next step is deposition of a sacrificial layer, which in the following steps will be removed for the gate release. Cured polymide can serve as a sacrificial layer to form the gap between the gate oxide and the suspended gate (Fig. 16e). Polymide can be patterned through a photolithography process or by dry oxygen etching and the polysilicon uses an $SF_6$ plasma etching. Both options require a two-step chemical mechanical polishing (CMP) to flatten the device. It is important to note that the modelling study presented in [27] showed the need to form gaps of a few nanometres to optimise the pull-in voltage values. Hence, dry etching might be a more suitable alternative for obtaining nm-order gaps. High-quality poly-Si, the most commonly used material for surface micromachined MEMS compatible with CMOS technology, cannot be used for the sacrificial layer, since it requires high-temperature annealing for stress reduction and dopant activation [21].

Thus, after patterning any of these sacrificial materials, a layer of AlSi can be sputtered to build the mechanical component (the bendable beam that will operate as the gate terminal) and then patterned with chlorine plasma etch (Fig. 16f). Finally, the suspended gate can be released with an isotropic dry oxygen plasma (for polymide) or $SF_6$ plasma (for polysilicon) (Fig. 16g).

# 4 Application of hybrid NEMS–CMOS as dynamic OR gates

Owing to near-zero-leakage characteristics of NEMFETs, these devices are ideal for low-power applications. However, since NEMFET devices have relatively low ON current characteristics compared to their CMOS counterparts, it is more desirable to build hybrid NEMS–CMOS circuits to combine good characteristics of each type of devices [20]. Therefore implementation of hybrid NEMS–CMOS circuits is explored in this section. First, we discuss limitations of pure CMOS dynamic OR gates. The next sub-section includes details of our proposed dynamic OR gate, which is then followed by HSPICE simulation results.

## 4.1 Limitations of pure CMOS dynamic OR gates

Dynamic implementation of wide fan-in OR gates offers low latency, because it does not require a p-type metal-oxide-semiconductor (PMOS) transistor stack unlike their static CMOS counterparts. However, the major disadvantage of dynamic gates is their low noise margin, which is conventionally improved by employing a PMOS keeper (Fig. 17a). But, under increasing process variation and
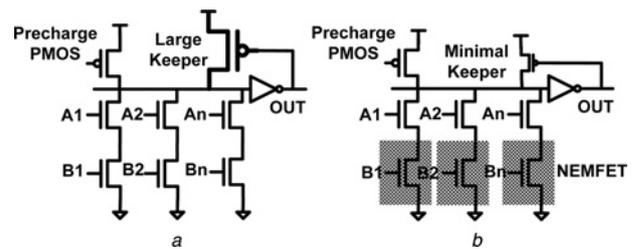


**Figure 17** *Dynamic OR gates*
a Conventional
b Proposed

higher leakage levels, keeper size must be increased to meet the noise margin criterion, which on the other hand, degrades performance of dynamic gates.

The trade-off between noise margin and performance is shown graphically in Fig. 18 for an eight-input dynamic OR gate, where Y- and X-axis represent normalised worst-case delay and noise margin in volts, respectively. Three curves show different levels of process variation measured in terms of standard deviation of threshold voltage ($\sigma_{Vth}$) as percentage of its nominal value ($\mu_{Vth}$) [36].

## 4.2 Hybrid NEMS–CMOS dynamic OR gates

To combine low leakage characteristics of NEMFET devices and high ON current of MOSFET devices, a novel dynamic gate architecture is presented as shown in (Fig. 17b). In this architecture, NEMFETs are placed in series below the standard NMOS devices in the pull-down network. Since NEMS devices are known to have much lower sub-threshold leakage than their CMOS counterparts, this will tremendously reduce leakage current of the pull-down network. As a result, the size of the keeper can be made to be minimal. Since contention between keeper circuit and pull-down network causes increased delay and power consumption, considerable improvement is expected because of minimisation of keeper circuit. On the other hand, because of low ON current of NEMS devices, pull-down
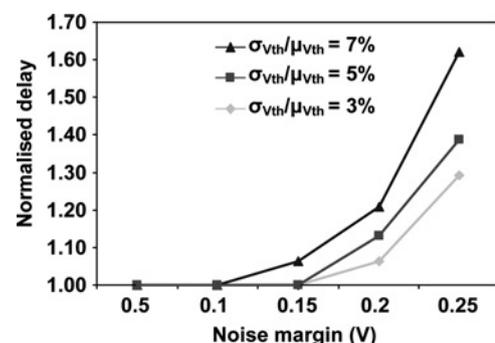


**Figure 18** *Normalised delay of an eight-input dynamic OR gate against its noise margin under different levels of process variation*

network of this architecture poses higher ON resistance compared to the conventional all-CMOS pull-down circuits. Hence, performance of the proposed gate is comparable to that of an equal-sized conventional dynamic gate.

A major advantage of the proposed circuit is its superior low-power operation. Switching power consumption of the new gate is much lower than conventional dynamic gates because of minimal contention between keeper and pull-down network. Moreover, leakage power consumption of the new gate is much smaller because of almost zero-leakage characteristics of NEMS switches.

## 4.3 Simulation results

Proposed dynamic OR circuit was simulated using HSPICE to investigate its potential advantages over pure CMOS counterpart. In Fig. 19, simulation results for two eight-input dynamic OR gates, one realised using only CMOS devices (Fig. 17*a*) and the other with hybrid NEMS–CMOS architecture (Fig. 17*b*), are shown. Switching power and the worst-case delay of two gates are plotted on the two vertical *Y*-axes for different fan-out values (plotted on *X*-axis). It can be observed that hybrid NEMS–CMOS gate shows 10 and 20% higher delay when gate is loaded with fan-out of 1 and 5, respectively. However, hybrid NEMS–CMOS gate consumes up to 60–80% less switching power compared to CMOS gate at the same fan-out level.

In Fig. 20, simulation results are shown for the two dynamic OR gates discussed above with different number of inputs (fan-in). Similar to the previous figure, switching power consumption and delay of gates are plotted on the vertical axes, whereas *X*-axis shows fan-in of the gates. For small gates (four- and eight-input) CMOS circuits exhibit lower delay; however, switching power consumption of the proposed hybrid gate is much lower than that of the conventional dynamic gate. It is interesting to note that as we move towards larger fan-in (12- and 16-input), the
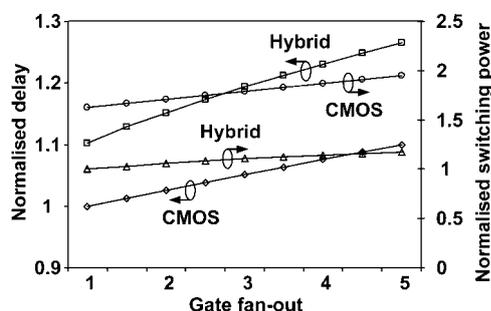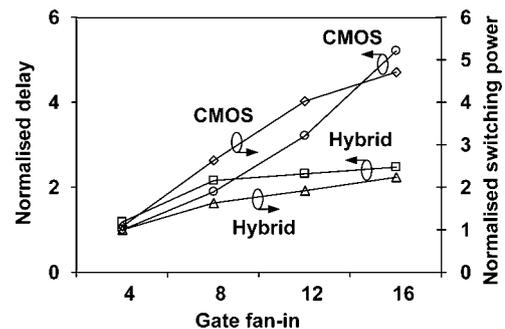


**Figure 20** *Normalised switching power consumption and the worst-case delay of hybrid NEMS−CMOS and CMOS dynamic OR gate for different fan-in levels*

Fan-out is kept constant at 3. Switching power and delay values are normalised with respect to the power and delay value of the hybrid gate with fan-in of 4

hybrid gate outperforms its CMOS counterpart both in terms of delay and switching power consumption.

It should be noted that although intrinsic delay of a mechanical-based device is several times higher than that of CMOS devices (as stated in Section 2.1), the delay difference between hybrid and CMOS dynamic gates in Figs. 19 and 20 is not that high. The reason is that for large dynamic gates, the contribution of intrinsic delay to total delay compared to that of load delay is negligible. Therefore it is possible for hybrid gate to have a comparable delay. Moreover, one must notice that hybrid NEMS circuit is able to compensate for its higher intrinsic delay and lower ON current, because it reduces the amount of contention between the keeper and pull-down circuit (in case of the dynamic gate). Therefore the reason that hybrid circuit results in lower delay is not because the device itself is faster than CMOS counterparts, but due of the fact that it is easier for hybrid circuit to pull down the dynamic node.

To capture contributions of delay and power consumption on overall performance of these two types of gates, power−delay product metric was calculated. In order to incorporate the impact of both idle-state leakage and switching power consumption, the total power consumption of the circuits was evaluated using (4). Here, $\alpha$ denotes the activity factor of the dynamic circuits, $P_L$ denotes leakage power of each gate, $P_S$ is the switching power consumption and $D$ is the worst-case delay of each gate

$$\text{P.D.} = ((1 - \alpha)P_L + \alpha P_S)D \qquad (4)$$

In Fig. 21, the power−delay product metric is plotted for the two dynamic gates with two different output load values of $C_L = 1$ and 3, which denote fan-outs of 1 and 3, respectively. In this figure, activity factor ($\alpha$) is varied from 0 to 1, which is plotted on the *X*-axis. As can be easily observed, the proposed hybrid architecture strongly surpasses the CMOS gate in terms of power−delay product in both cases.



**Figure 19** *Normalised switching power consumption and the worst-case delay of two eight-input hybrid NEMS− CMOS and CMOS dynamic OR gates for different fan-out levels*

Switching power and delay values are normalised with respect to the power of the hybrid gate and delay of the CMOS gate with fan-out of one, respectively
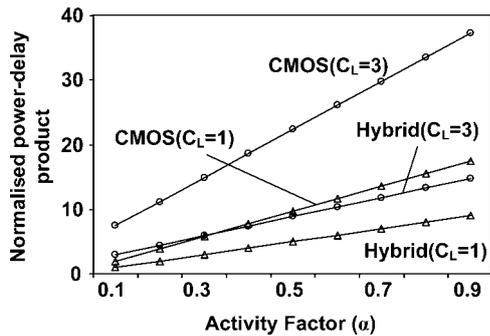
**Figure 21** *Power−delay product metric comparison for hybrid NEMS−CMOS and pure CMOS dynamic OR gates with different output capacitances*

# 5 Application of hybrid NEMS−CMOS in SRAM cell design

SRAM cell design is known to be one of the most challenging tasks of low-power designers [3]. In this section, we introduce a novel SRAM architecture based on the proposed NEMS−CMOS technology. Also, we compare the proposed SRAM cell against various existing low-power SRAM cells and the conventional SRAM circuit.

## 5.1 Leakage issue in pure CMOS SRAM cells

Performance of modern microprocessors is strongly dependent on the size of their on-chip cache memories. The cache memories are composed of arrays of SRAM cells (Fig. 22a). As technology scales down, more and more SRAM cells can be integrated in the same area. However, because of the increased leakage current of transistors, the relative fraction of leakage power consumption (as compared to the fraction of switching power) is increasing. Moreover, the probability of read failures (toggling of stored value during read operation) and read latency (delay between accessing the cell and sensing the voltage change on bit lines) also degrades with scaling.

Read stability degrades because current of the access transistor AR (in saturation region) increases at a higher pace than the current of the pull-down transistor NR (in linear region) as shown in Fig. 22a. Also, the higher leakage current of access transistors (in other cells that are connected to the bit line bar (BLB)) makes it tougher for the access transistors to create the necessary voltage difference for sense amplifiers. Therefore read access delay is also adversely affected by leakage. Hence, using low-leakage devices can potentially lead to better performance and stability as well as lower power consumption [2].

## 5.2 Low-leakage CMOS SRAM cells

Low-leakage SRAM cell architectures must be able to reduce the leakage power consumption; however, they also must have minimal impact on the read and write latency of cache
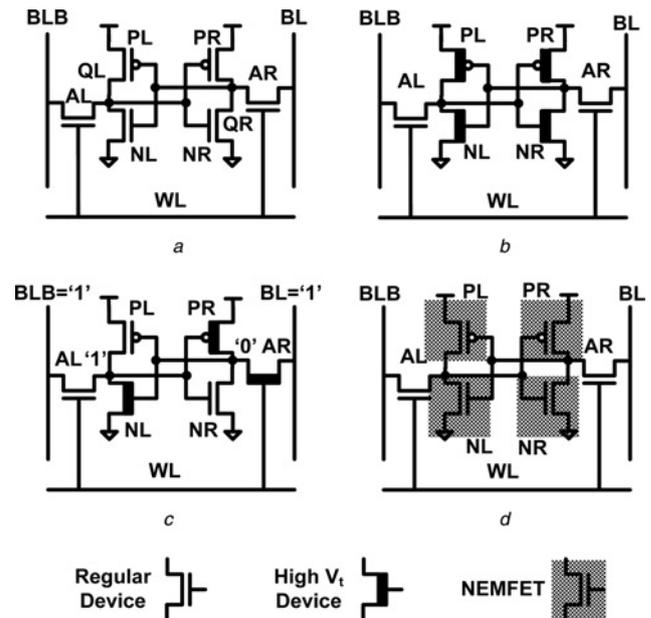


**Figure 22** *Different SRAM cell architectures in the literature*

*a* Conventional
*b* Dual $V_t$ [37]
*c* Asymmetric [38]
*d* Proposed hybrid NEMS−CMOS SRAM cell
AR, AL, PR, PL, NR and NL refer to different transistors; QR and QL refer to the output node of the inverter located on right and left side of the SRAM cell, respectively

memory. Recently, a dual-$V_t$ SRAM cell has been proposed in [37], where both high- and low-$V_t$ transistors were employed to reduce leakage current at the cost of lower cell stability and performance (Fig. 22b). Also, an asymmetrical SRAM cell architecture has been proposed using dual-threshold voltage technology [38] (Fig. 22c). It is argued that since data stored in cache are more likely to be zeros than ones, high-$V_t$ devices can be used to reduce the leakage in zero storing state and low-$V_t$ transistors can be used for the one state.

It should be noted that all existing low-power SRAM cells achieve low-leakage characteristics with some noise margin and latency cost. In the following sub-section, we propose employing hybrid NEMS−CMOS structures and show that although the stability and performance loss in our proposed architecture are comparable to those of other approaches, superior low-leakage characteristics can be achieved with hybrid cells (Fig. 22d).

## 5.3 Hybrid NEMS−CMOS SRAM cells

In this section, a hybrid NEMS−CMOS SRAM cell is proposed as shown in Fig. 22d. Compared to the conventional cell, NMOS pull-down transistors (NR and NL) along with the PMOS pull-up devices (PL and PR) are replaced with their NEMS counterparts. Since NEMS devices have very low OFF current, NEMS transistors draw minimal leakage current from power supplies;

however, low ON current of NEMS devices degrades the stability of the cell against read failure errors. It should be noted that replacing access transistors (AR and AL) with NEMS devices is not a good idea because of their huge impact on latency.

Another alternative to the proposed architecture is to only replace the PMOS pull-up transistors (PL and PR in Fig. 22d) with NEMS devices. Since PMOS devices are OFF during the read operation, low ON current of PMOS NEMS devices does not affect the read latency. However, in this case, leakage power saving is not as much as that of the proposed version because of the leaky NMOS devices. Different trade-offs of employing hybrid NEMS–CMOS in SRAM cells are discussed in the following sub-section.

## 5.4 Simulations and comparisons

Different SRAM cell structures of Fig. 22 are simulated using HSPICE. As in the previous section, for NEMS devices, $I-V$ characteristics were taken from [27]. Three major metrics are used for comparison: (a) static noise margin (SNM), (b) read latency and (c) standby leakage power consumption of SRAM cells. Four different SRAM cell architectures of Fig. 22 are examined. In the simulation results, we call architectures of Figs. 22a, b, c and d as 'Conv.', 'Dual $V_t$', 'Asym.' and 'Hybrid', respectively. In Fig. 23, butterfly curves are shown for different SRAM cells and measured values for SNM are reported. It can be observed that the proposed hybrid architecture exhibits 14% lower noise margin as compared to that of the conventional cell; however, its noise margin is slightly higher than those of the other two low-leakage SRAM cell architectures.

In Fig. 24, the read latency and standby leakage current of various SRAM cells are compared. It should be noted that the read latency and leakage values are normalised to those values of the conventional SRAM cell for ease of comparison. Also, since the asymmetric cell shows different read latencies for zero and one values, an average value is used in this graph. As can be observed, all three low-leakage SRAM cells have higher latency compared to the
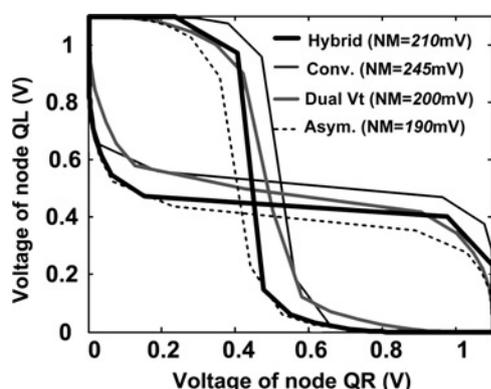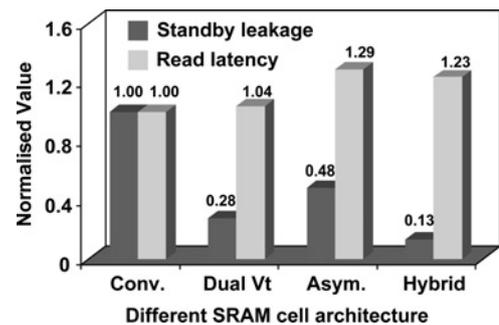


**Figure 24** *Comparison between conventional, existing low-leakage and proposed SRAM cell architectures*

conventional design. The proposed hybrid design has 23% higher read latency. However, it is possible to further reduce the latency of the hybrid cell via proper transistor and circuit optimisation. Finally, from a leakage perspective, the hybrid SRAM cell outperforms all other circuits and has almost 8× lower leakage power consumption compared to the conventional SRAM cell.

# 6 Application of NEMS devices as sleep transistors

Low-leakage characteristics of NEMS devices can also be exploited in designing more efficient sleep transistors. Sleep transistors refer to switches that are placed between the power supply and the circuit to reduce leakage current [39] (Fig. 25). During the normal operation, the sleep transistor is turned ON to connect the circuit to power supply and it is turned OFF to reduce leakage current during the idle periods when circuit is not active. There are two different types of sleep transistors: header-type (Fig. 25a), which is composed of a PMOS device and is placed between the actual $V_{dd}$ and the circuit and a footer-type (Fig. 25b), where an NMOS device is placed between the actual ground and the circuit. Sleep transistors can also be categorised as fine- and coarse-grain types. In the case of
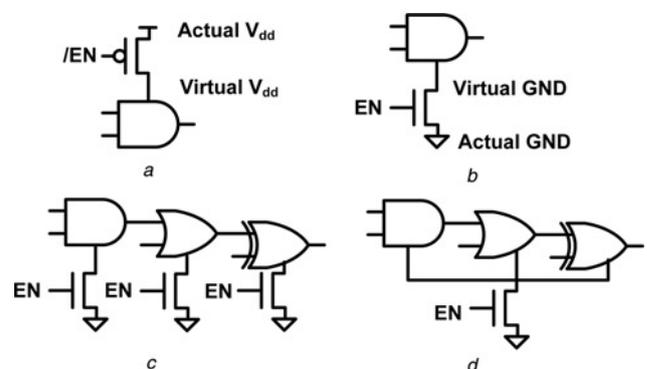


**Figure 23** *SRAM butterfly curves used for measuring SNM values for different SRAM cell architectures of Fig. 22*



**Figure 25** *Different types of sleep transistors*

a Header-type
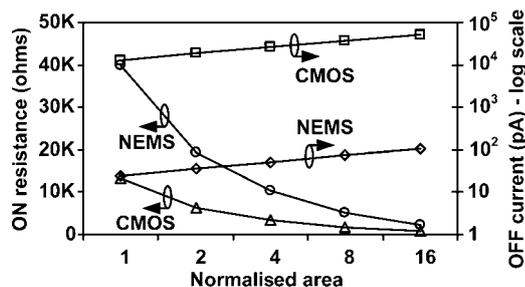b Footer-type
c Fine-grain
d Coarse-grain

**Figure 26** *Comparison between ON resistance and OFF current of NEMS and CMOS sleep transistors*

All areas are normalised with respect to the area of a CMOS device with $W/L = 5$ in 90 nm technology

fine-grain sleep transistor (Fig. 25c), each digital gate is separated from the power supply independently by one sleep device; however, in the coarse-grain implementation (Fig. 25d), one sleep transistor is responsible for separation of a block of digital gates.

Since NEMS devices demonstrate ultra-low leakage characteristics, they are an ideal choice as sleep transistors. There are three important features of each sleep transistor [39]. First, it must have low sub-threshold leakage current; second, it must have low ON state resistance so that the voltage difference between virtual and actual power supply node (Fig. 25) remains minimal. Third, its area should be preferably small compared to the rest of the circuit. A comparison between ON-state resistance and OFF-state leakage of NEMS- and CMOS-based sleep transistors for different device areas is shown in Fig. 26. It can be observed that NEMS-based sleep device outperforms CMOS-based sleep transistor in terms of OFF-state leakage current. However, it can also be noticed that NEMS devices exhibit higher ON resistance for equal area size, because of lower drain current of NEMS devices. It can be observed from Fig. 26 that for larger sleep transistors, the difference between the ON resistance of NEMS and CMOS sleep transistors become minimal. Hence, the NEMS sleep transistors can be sized-up to offer up to three orders of magnitude lower leakage current with negligible performance degradation compared to that of their CMOS counterparts.

# 7 Conclusions

The hybrid NEMS–CMOS technology is proposed as an alternative to the current CMOS technology, which is troubled by ever-increasing leakage power consumption. The hybrid NEMS–CMOS technology integrates NEMS with CMOS devices to combine near-zero leakage characteristics of NEMS switches with high ON current of CMOS transistors and simultaneously offers ultra-low-power and high-performance operation. This technology can have interesting implications for mobile applications where low-leakage power consumption and energy efficiency are extremely critical. To illustrate the feasibility

of the hybrid NEMS–CMOS technology, a simplified process flow has been outlined that shows different steps for fabrication of a NEMS device along with a CMOS transistor. Various low-power applications of the proposed hybrid technology have been explored through judicious design of hybrid NEMS–CMOS dynamic OR gates and SRAM cells at the 90 nm node. Simulation results indicate that such dynamic gates can achieve 60–80% lower switching power consumption and almost zero-leakage power consumption with minor delay penalty. Most importantly, the hybrid gate outperforms its CMOS counterpart both in terms of delay and switching power consumption with increase in fan-in beyond 12. Additionally, it is shown that a hybrid SRAM cell can achieve nearly $8\times$ lower standby leakage power consumption with only minor noise margin and latency cost, which can be further reduced with adequate device/circuit optimisation. Finally, as sleep transistors, NEMS devices offer up to three orders of magnitude lower OFF current with negligible performance degradation compared to their CMOS counterparts.

# 8 Acknowledgment

# 9 References

[1] MOORE G.E.: 'Cramming more components onto integrated circuits', *Electronics*, 1965, **38**, pp. 114–117

[2] DE V., BORKAR S.: 'Technology and design challenges for low power and high performance microprocessors'. Proc. IEEE Int. Symp. Low Power Electronics and Design, 1999, pp. 163–168

[3] BORKAR S., KARNIK T., DE V.: 'Design and reliability challenges in nanometer technologies'. IEEE/ACM Design Automation Conf., 2004, p. 75

[4] International Technology Roadmap for Semiconductors (ITRS), http://public.itrs.net

[5] BORKAR S.: 'Exponential challenges, exponential rewards – the future of Moore's law'. VLSI-SOC, 2003: 2

[6] BANERJEE K., LIN S.-C., KESHAVARZI A., NARENDRA S., DE V.: 'A self-consistent junction temperature estimation methodology for nanometer scale ICs with implications for performance and thermal management', *IEDM Tech. Digest*, 2003, pp. 887–890

[7] TAUR Y., NING T.: 'Fundamentals of modern VLSI devices' (Cambridge University Press, 1998)

[8] APPENZELLER J., LIN Y.M., KNOCH J., AVOURIS PH.: 'Band-to-band tunneling in carbon nanotube field-effect transistors', *Phys. Rev. Lett.*, 2004, **93**, p. 196805

[9] APPENZELLER J., YU-MING L., KNOCH J., ZHIHONG C., AVOURIS P.: 'Comparing carbon nanotube transistors – the ideal choice: a novel tunneling device design', *IEEE Trans. Electron Devices*, 2005, **52**, pp. 2568–2576

[10] WOUTERS D.J., COLINGE J.P., MAES H.E.: 'Subthreshold slope in thin-film SOI MOSFETs', *IEEE Trans. Electron Devices*, 1990, **37**, pp. 2022–2033

[11] LIN H.C., LEE M.H., SU C.J., SHEN S.W.: 'Fabrication and characterization of nanowire transistors with solid-phase crystallized poly-si channels', *IEEE Trans. Electron Devices*, 2006, **53**, pp. 2471–2477

[12] GOPALAKRISHNAN K., GRIFFIN P.B., PLUMMER J.D.: 'I-MOS: a novel semiconductor device with a subthreshold slope lower than $kT/q$', *IEDM Tech. Digest*, 2002, pp. 289–292

[13] ABELE N., FRITSCHI R., BOUCART K., ET AL.: 'Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor', *IEDM Tech. Digest*, 2005, pp. 479–481

[14] SALAHUDDIN S., DATTA S.: 'Use of negative capacitance to provide voltage amplification for low power nanoscale devices', *Nano Lett.*, 2008, **8**, pp. 405–410

[15] SZE S.M.: 'Physics of semiconductor devices' (Wiley, New York, 1981, 2nd edn.)

[16] RUECKES T., KIM K., JOSELEVICH E., TSENG G.Y., CHEUNG C.L., LIEBER C.M.: 'Carbon nanotube-based nonvolatile random access memory for molecular computing', *Science*, 2000, **289**, pp. 94–97

[17] DEQUESNES M., ROTKIN S.V., ALURU N.R.: 'Calculation of pull-in voltages for carbon nanotube-based nanoelectromechanical switches', *Nanotechnology*, 2002, **13**, pp. 120–131

[18] YEO Y.-C., RANADE P., KING T.-J., HU C.: 'Effects of high-*K* gate dielectric materials on metal and silicon gate workfunctions', *IEEE Electron Dev. Lett.*, 2002, **23**, pp. 342–344

[19] AUTH C., CAPPELLANI A., CHUN J.-S., ET AL.: '45 nm high-$k$ + metal-gate strain-enhanced transistors', *Intel Technol. J.*, 2008, **12**, pp. 77–86

[20] DADGOUR H.F., BANERJEE K.: 'Design and analysis of hybrid NEMS−CMOS circuits for ultra low-power applications'. IEEE/ACM Design Automation Conf., 2007, pp. 306–311

[21] TAKEUCHI H., WUNG A., SUN X., ET AL.: 'Thermal budget limits of quarter-micrometer foundry CMOS for post-processing MEMS devices', *IEEE Trans. Electron Devices*, 2005, **52**, (9), pp. 2081–2086

[22] EKINCI K.L., ROUKES M.L.: 'Nanoelectromechanical systems', *Rev. Sci. Instrum.*, 2005, **76**, Article no. 061101

[23] JANG J.E., CHA S.N., CHOI Y., ET AL.: 'Nanoelectromechanical switches with vertically aligned carbon nanotubes', *Appl. Phys. Lett.*, 2005, **87**, p. 163114

[24] ABELE N., SEGUENI K., BOUCART K., ET AL.: 'Ultra-low voltage MEMS resonator based on RSG-MOSFET'. Proc. 19th IEEE Int. Conf. MEMS, 2006, pp. 882–885

[25] HORNBECK L.J.: 'Digital light processing and MEMS: timely convergence for a bright future'. Proc. SPIE Micromachining and Microfabrication Process Technology Conf., 1995, vol. SPIE 2639, pp. 1–21

[26] GONG J.F., XIAO Z.Y., CHAN P.: 'Integration of an RF MEMS resonator with a bulk CMOS process using a low-temperature and dry-release fabrication method', *J. Micromech. Microeng.*, 2007, **17**, pp. 20–25

[27] KAM H., LEE D.T., HOWE R.T., KING T.-J.: 'A new nano-electro-mechanical field effect transistor (NEMFET) design for low-power electronics', *IEDM Tech. Digest*, 2005, pp. 463–466

[28] http://www-device.eecs.berkeley.edu/~ptm/mosfet.html

[29] CARR D.W., EVOY S., SEKARIC L., CRAIGHEAD H.G., PARPIA J.M.: 'Measurement of mechanical resonance and losses in nanometer scale silicon wires', *Appl. Phys. Lett.*, 1999, **75**, pp. 920–922

[30] IONESCU A.M., POTT V., FRITSCHI R., ET AL.: 'Modeling and design of a low-voltage SOI suspended-gate MOSFET (SG-MOSFET) with a metal-over-gate-architecture'. IEEE Int. Symp. Quality Electronic Design, 2002, pp. 496–501

[31] KINARET J.M., NORD T., VIEFERS S.: 'A carbon nanotube based nanorelay', *Appl. Phys. Lett.*, 2003, **82**, pp. 1287–1289

[32] LEE S., LEE D., MORJAN R., ET AL.: 'A three-terminal carbon nano-relay', *Nano Lett.*, 2004, **4**, pp. 2027–2030

[33] DADGOUR H., CASSELL A.M., BANERJEE K.: 'Scaling and variability analysis of CNT-based NEMS devices and circuits with implications for process design', *IEDM Tech. Digest*, 2008, pp. 529–532

[34] POTT V., IONESCU A., FRITSCHI R., ET AL.: 'The suspended-gate MOSFET (SG-MOSFET): a modeling outlook for the design of RF MEMS switches and tunable capacitors'. Proc. Int. Semiconductor Conf., 2001, pp. 137–140

607

[35] NGUYEN C.T.C., HOWE R.T.: 'An integrated CMOS micromechanical resonator high-$Q$ oscillator', *IEEE J. Solid-State Circuits*, 1999, **34**, pp. 440–455

[36] DADGOUR H.F., JOSHI R.V., BANERJEE K.: 'A novel variation-aware low-power keeper architecture for wide fan-in dynamic gates'. ACM Design Automation Conf. (DAC), 2006, pp. 991–996

[37] HAMZAOGLU F., YE Y., KESHAVARZI A., *ET AL*.: 'Dual $V_t$-SRAM cells with full-swing single-ended bit line sensing for high-performance on-chip cache in 0.13 $\mu$m technology generation'. Proc. IEEE Int. Symp. Low Power Electronics and Design, 2000, pp. 15–19

[38] AZIZI N., NAJM F.N., MOSHOVOS A.: 'Low-leakage asymmetric-cell SRAM', *IEEE Trans. Very Large Scale Integr. Syst.*, 2003, **11**, pp. 701–715

[39] KAIJIAN S., HOWARD D.: 'Challenges in sleep transistor design and implementation in low-power designs'. Proc. IEEE/ACM Design Automation Conf., 2006, pp. 113–116