Weighted Area Technique for Electromechanically Enabled Logic Computation With Cantilever-Based NEMS Switches

Shruti Patil¹, Min-Woo Jang¹, Chia-Ling Chen¹, Dongjin Lee², Zhijang Ye², Walter E Partlo III³, David J. Lilja¹, Stephen A. Campbell¹, and Tianhong Cui²

¹Department of Electrical and Computer Engineering, University of Minnesota
²Department of Mechanical Engineering, University of Minnesota
³Department of Chemistry, University of Minnesota

Abstract—Nanoelectromechanical systems (NEMS) is an emerging nanoscale technology that combines mechanical and electrical effects in devices. A variety of NEMS-based devices have been proposed for integrated chip designs. Amongst them are near-ideal digital switches. The electromechanical principles that are the basis of these switches impart the capability of extremely low power switching characteristics to digital circuits. NEMS switching devices have been mostly used as simple switches to provide digital operation, however, we observe that their unique operation can be used to accomplish logic functions directly. In this paper, we propose a novel technique called ‘weighted area logic’ to design logic circuits with NEMS-based switches. The technique takes advantage of the unique structural configurations possible with the NEMS devices to convert the digital switch from a simple ON-OFF switch to a logical switch. This transformation not only reduces the delay of complex logic units, but also decreases the power and area of the implementation further. To demonstrate this, we show the new designs of the logic functions of NAND, XOR and a three input function $Y = A + B.C$, and compose them into a 32-bit adder. Through simulation, we quantify the power, delay and area advantages of using the weighted area logic technique over a standard CMOS-like design technique applied to NEMS.

I. INTRODUCTION

The emerging technology of Nanoelectromechanical Systems (NEMS) is an advancement of the fabrication of electromechanical devices into nanoscale regimes. Digital switches that are implemented using the NEMS technology show several superior characteristics, such as zero leakage currents, high on-current capabilities and the potential for high-speed, low-Vt (threshold voltage of devices) operation. In the light of the growing power issues in digital designs that use CMOS transistors, the near-ideal OFF-state characteristics of the NEMS switches are particularly attractive.

A number of published works have used NEMS switches in digital designs and shown their potential to reduce power consumption to a great extent [1], [2]. These ideas are based on a variety of device structures that have been proposed for NEMS-based switches([1], [3], [4], [5]). With different device structures, the applicability of the devices in digital design situations changes, however, most structures have been used as simple switches that use a combination of electrostatic and mechanical forces for actuation and release. In this work, we observe that the unique electromechanical operation of the device need not be restricted to a simple switch, but can be used to derive logic functions directly. This idea further allows for novel logic design strategies for NEMS devices. In this paper, we describe a logic technique called ‘weighted area logic’ (WAL) design, with which logic functions are embedded into the device structure itself. By transforming electromechanical properties into a direct logic capability, the device gets converted into a logical switch. This reduces the number of devices required to implement a logic function, which leads to a decrease in delay, power and area of the implementation.

To demonstrate and evaluate the proposed logic design technique, we use Carbon-nano-tube (CNT) based NEMS cantilever switches. Amongst the materials that have been studied for NEMS devices (for example tungsten, poly-silicon [1], CNT [3], [6]), CNT films possess the most promising material properties, such as low mass density, high yield strength and high Young’s modulus [6]. Theoretically, this leads to the potential to operate in Gigahertz ranges with a low-Vt ([7], [8]). Experimentally, two-terminal NEMS switches operating at about 600ps have been demonstrated in ([8], [9]). Using the proposed logic design technique with CNT-based NEMS devices, we show the design of a 32-bit adder. We find that the WAL design strategy has the potential to decrease the delay as well as the power, which together results in a considerable reduction in the energy and energy-delay product of the adder. The proposed logic design technique is intended to be a general concept that applies to any NEMS switching structure which relies on electromechanical principles.

II. CNT BASED NEMS-CANTILEVER DEVICE AND OPERATION

A cantilever-based nano-electro-mechanical (NEMS) switch is actuated by a combination of electrical and mechanical effects in the device. The NEMS-CNT switch that we use in this paper is a 3-terminal structure, with anchor (A), gate (G)
and drain (D) terminals. A cantilever beam fabricated with a single wall carbon nanotube (SWCNT) thin film extends from the anchor to the drain terminal, forming a CNT-based conducting channel in the switch. Fig. 1(a) shows the device structure, where the cantilever beam stands freely above the drain. In order to operate the device, a gate voltage is applied to create a voltage difference between the gate and the anchor. This induces opposite charges on the gate and the cantilever, creating an electrostatic force of attraction between them. If the difference in the voltages on the terminals exceeds a certain threshold value called as pull-in voltage $V_{PI}$, the electrostatic forces exert a sufficient pull-in force on the cantilever that forces it to make contact with the drain. This connects the anchor and drain terminals, thus closing the ‘switch’. When the electrostatic force of attraction disappears, the pull-in force disappears, however, in the closed position, there exists a spring restoring force in the cantilever due to its mechanical properties. Therefore the cantilever springs back to its freely standing initial position, resulting in an ‘open’ switch. To avoid shorting the source to the gate, a thin insulator can be placed on top of the gate electrodes. This is achieved by using a high permittivity material such as $HfO_2$, to minimize any reduction in the electrostatic force [10].

$$V_{dd} = \text{constant}$$

![Switch Structure](image)

Fig. 1. (a)-(c)NEMS-CNT Device Structures; (d)-(e)Voltage biases on terminals for different logic states

The simple cantilever shown in Fig. 1(a) is very dense, but is quite sensitive to the effects of stress in the film, which can lead to significant changes in the zero-bias gap between the gate and the electrode. A more stable structure is shown in Fig. 1(b), where the cantilever beam is supported on both sides by the anchor terminals. This structure, as seen in the top view is shown in Fig. 1(c). The behavior of the device is fundamentally similar to the first structure, but is less sensitive to process variables. Furthermore, it provides important new design capabilities by incorporating multiple gate electrodes into a single device. The biasing of the anchor and gate terminals in the open and closed states is shown in Fig.1(d)-1(e). Voltage $V_{dd}$ is chosen to be $1 - 2$ times that of $V_{PI}$ according to the required noise margin levels. NEMS-CNT switches that operate on this principle have been demonstrated in ([3], [6], [8]). With suitable materials, their potential to operate at pull-in voltages of 100mV has also been investigated [11], making them promising devices for use in low-power digital designs.

**A. Equivalence to a transistor**

Digital designs using the MOSFET technology have traditionally utilized the ON/OFF switch-like behavior of the transistors to accomplish logic functions. In fact, the behavior of a NEMS-CNT device resembles that of a digital CMOS switch in that the voltage between the anchor and the gate acts as the controlling voltage, and determines the ON or OFF state, similar to a transistor operation. Therefore, an equivalent NEMS-CNT-based logic design can be derived by replacing the PMOS and NMOS devices in a CMOS-based design with appropriately biased NEMS-CNT devices. We will refer to this design style as the standard design technique for NEMS. In order to demonstrate that the standard design technique applies to real NEMS devices, we demonstrated a NAND logic gate fabricated using a NEMS-CNT device with a fixed-fixed beam structure. The fabrication process, device details and characterization curves are presented in [10]. This experimentally verified two aspects of designing with NEMS-CNT switches: (1) the behavioral similarity between MOSFET switches and NEMS-CNT devices, and (2) the feasibility of deriving logic circuits by a simple replacement of CMOS devices by NEMS-CNT switches.

**B. Model for circuit simulation**

To assist in circuit simulation, the behavior of the device can be abstracted into a parameterized NEMS device model. Functionally, the NEMS cantilever device is a 3-terminal switch with a finite on-resistance $R_{on}$. Its switching mechanism has a mechanical delay $T_{mech}$. Switching occurs at pull-in voltage $V_{PI}$. For any gate voltage less than $V_{PI}$, the device current is zero. While the gate-cantilever capacitance $C_{GC}$ can be modeled as a dynamic capacitance that describes the electromechanical operation, we note that the internal capacitances of the device are extremely low. For the gate, this is due to the large gap, typically about 20nm, compared to a 1nm equivalent oxide thickness of a fully scaled gate oxide. For the source and drain, the low capacitance is because there are no junctions in the device. It is built in a low dielectric contact material ($SiO_2$), instead of in a semiconductor. Thus, the capacitance of the device as experienced by a circuit stems mainly from the self capacitances of the wire associated with the three terminals of the anchor, gate and the drain, i.e. $C_a, C_g$ and $C_d$ respectively. Another significant behavior exhibited by the NEMS-CNT devices is their asymmetry in rise and fall times [6]. The rise time due to the device switching ON is about 80-90% higher than the fall time measured when the device switches OFF.

Based on this operation, we developed a Verilog-A model for the NEMS devices for circuit-level simulation of complex logic units that can allow us to compare between different designs. It models the NEMS-CNT device as a 3-terminal switch with five parameters $R_{on}, T_{mech} (= T_{ON}), V_T, C_{node}$ and Asymmetry Ratio ($T_{OFF}/T_{ON}$). This simplistic behavioral model allows us to fit measured device values in the
model, while giving the flexibility to observe the sensitivity of the device for different parameters along with preliminary performance evaluation and projection for different applications.

III. CANTILEVER-BASED NANOMECHANICAL LOGIC DESIGNS

While one feasible technique of designing with the NEMS-CNT devices is by using the classic complementary standard design technique, the devices also possess a number of unique properties. For example, the behaviors of the PMOS and NMOS devices in a CMOS implementation can be obtained by applying appropriate biases on the same device structure. This simplifies the fabrication process significantly and reduces the area. Secondly, the pull-in voltage of an individual device can be easily varied by varying the structure of a device. This in turn, enables the devices to be manufactured as more complex structures. On the flip side, the devices suffer from a large mechanical delay relative to an electrical delay. Therefore, it is desired to reduce the total delay of a logic circuit. Taking into account these strengths and challenges of the devices, we propose an electromechanical design strategy called as ‘weighted area logic design’ that relies on the electromechanical operation of the devices to implement NEMS-CNT based logic structures. By incorporating both electrical and mechanical effects to produce a logical result directly, the number of devices in the maximum combinational path are reduced, resulting in faster logic computation. Also, by integrating the logic functionality into the device structure, the logic circuits that are produced result in smaller structures than a corresponding standard implementation.

A. Weighted Area Logic (WAL) Design Concept

The amount of electrostatic force formed between the cantilever and the gate depends upon many factors, including (1) the voltage bias on the two terminals; (2) the area of the cantilever and the gate resulting in the pull-in forces; (3) the distance between the cantilever and the gate. The weighted area logic design strategy exploits the second factor. The pull-in force experienced by the cantilever is proportional to the overlap area between the cantilever and the gate. By changing the cantilever-gate overlap area, the magnitude of effect of the gate on the cantilever can be controlled. The larger the overlap area, the higher is the effect, or the ‘weight’, of the gate. With this ability, instead of relying on switch-based electronic circuit to compute a logical output, we are able to leverage the electromechanical properties of the devices to obtain the logical result directly.

When a larger impact of an input is required on the output, a device with a larger gate-cantilever overlap area is designed. For example, in the expression \( Y = A + B.C \), the input A affects the logic-high output individually, while inputs B and C affect the logic-high result in combination. Thus, the contribution of input A in the function is larger, and necessitates a larger impact on the cantilever by input A alone as compared to that of inputs B and C alone. The contribution of an input to the logic-high and logic-low states in a truth table is evident when the logic design is expressed in the form of a sum-of-products (SOP) (or product-of-sum (POS)) expression. These expressions describe the desired output as a logical function of the inputs. For instance, when inputs are OR-ed together, all inputs affect the output equally and independently. On the other hand, when inputs are AND-ed together, this indicates that the inputs affect the output in combination.

We convert this observation into the weighted area logic design approach. An input that affects the output individually is designed so that it exerts sufficient mechanical pull-in on a cantilever by itself. On the other hand, inputs that affect the output in combination are designed and placed so that they are able to pull-in a cantilever only in combination. To increase the pull of a gate on the cantilever, its area of overlap with the cantilever is increased, while to decrease its pull on the cantilever, its area of overlap is reduced. In terms of weights, the inputs that are OR-ed are allotted equal weightage, each of which is the maximum weightage for the determination of the output. The inputs that are AND-ed are also given equal weightage, however the sum of the weights is the maximum weightage. Thus, a cantilever–based design can be directly derived from an SOP or a POS expression. Since the logical operation is partly incorporated into the physical structure, the resultant design occupies a lesser chip area than a design with devices that act as 3-terminal switches equivalent to a standard MOSFET operation.

B. Device-level simulation of the WAL Concept

To first validate the weighted area theory, we used a MEMS device simulator called Coventorware. The simulator enables a structural verification of the device layout and allows us to see the effects of the weighted gate overlap structures. Fig. 2 shows the four structures that were studied. With a normal structure as a baseline, the gate-cantilever overlap area was varied as \(-28\%\), \(+33\%\) and \(+67\%\) to observe the changes in pull-in voltages for a smaller area as well as
larger areas. For the simulation, three differently sized devices were considered with W/L= 2um/14um, 1um/7um and 0.6um/4.2um respectively. Fig 3 shows the change in pull-in voltage and resonant frequencies observed due to the change in the cantilever-gate overlap area. The results show that within +/-30% of overlapped gate area, the pull-in voltage changed within a range of +/-30%. However, its fundamental resonant frequencies that affect its switching delay did not change as much. This helps to strongly conclude that the weighted gate overlap structures could change their pull-in voltages effectively without much loss of switching speed.

IV. WEIGHTED AREA LOGIC (WAL) DESIGN OF DIGITAL ELEMENTS

A. NAND/NOR Logic Gates

Using the weighted area concept, basic two-input logic functions of NAND and NOR can be directly implemented as electromechanical designs. Fig. 4(a) shows the NAND design employing NEMS-CNT devices using the standard MOSFET design technique. It is derived from CMOS-based implementation of a NAND gate using only the switch-like behavior of the devices. This design is expected to take advantage of the low static power dissipation due to near-ideal OFF characteristics of the devices, as well as the simplicity of fabrication due to same type of devices performing the pull-up and pull-down function (as opposed to p-type and n-type required in MOSFET designs). However, the maximum mechanical delay of the circuit for the worst-case input transition is two units of the mechanical delay of the cantilever beam.

Fig. 3. Effects of weighted gate overlap on device characteristics

By employing the weighted logic design methodology, a smaller design with a delay as small as a single unit of mechanical delay can be obtained. Fig. 4(b) shows a custom design based on the weighted logic design concept that electromechanical forces can combine logically to enable switching. The two cantilevers C1 and C2 operate in parallel biased by power and ground, and are controlled by common input lines A and B. With suitable overlap areas, forces from two gates can either pull in a cantilever individually or in combination. Cantilever C1 experiences pull-in forces from the inputs and makes contact with the output when both inputs are at Vdd. Cantilever C2 is constructed so that the overlap area between the cantilever and each input is much larger than the overlap area of cantilever C1 with each input. Its overlap area with each gate is actually designed to ensure that each input is individually capable of exerting a pull-in force that is sufficient to enable the cantilever to make contact with the output. To increase the gate-cantilever overlap area, the cantilever C2 can be designed as a ‘winged’ structure as shown in Fig. 4(b). However, even with a simply designed cantilever, the NAND function can be obtained by separating the individual gates into individual cantilevers, as shown in Fig. 5(a). With the cantilevers biased at Vdd and Ground respectively, the NAND logic function is obtained. On the other hand when the cantilevers are biased at Ground and Vdd respectively, the NOR logic operation is obtained.

B. Simulation of NAND-WAL circuit

In order to verify the design, we simulated the NEMS-WAL circuit using the device model discussed in Section II-B with SPICE simulation tool. We extracted the sensitivity of the NAND-WAL design to varying values of the device param-
eters. With respect to ON-OFF time asymmetry, the NEMS-CNT devices consume lesser power in the WAL design, since the power dissipation in the short circuit mode when the devices are switching at once ($P_{SC}$) is reduced to almost zero. With respect to varying ON resistance and node capacitance, the WAL design did not show a lot of variation, however it always consumes lesser power and delay than the standard implementation.

We also compared the NAND-WAL design with the standard NAND (NAND-Std) design, using the following device parameters: $C_{node} = 1 fF$, $T_{mech} = 1 ns$, $V_{PP} = 1 V$, $R_{on} = 1 k\Omega$ and $T_{off} = 0.5 T_{on}$. These values were chosen as optimal or average parameters from the sensitivity curves. The load capacitance at the output nodes was assumed to be $1 fF$. For all possible transitions in the NAND truth table, the WAL design reduces the maximum mechanical delay of two units to a single unit, demonstrating the speed advantage of incorporating logic capability within the device structure.

C. XOR gate

The XOR function is a key function that accomplishes the SUM operation in a half-adder. Several interesting device structures have been proposed previously to implement the XOR function effectively in order to reduce the area of an adder, for example, the 4-terminal device structure in [1] and the dual-gate structure in [5]. These structures were designed specially for the XOR operation, and therefore would prove more effective. Under the weighted area logic scheme, the XOR gate (Fig. 5(c)) can be designed as four cantilevers operating in parallel and enabled by different gate inputs. This demonstrates a regular pattern in which a NEMS-CNT circuit can be designed and fabricated. It reduces the maximum delay of the function to two mechanical units.

![Fig. 5. NEMS-WAL designs of adder components](image)

D. Three input functions $Y=A+B.C$

In theory, the weighted area logic design can extend to any number of inputs. However, the practical limitations of an integrated weighted logic design are determined by the number of gates that can be fabricated for a single cantilever. In practice, an OR expression of the form $Y = A+B+C+D+...$ can be accomplished with multiple cantilevers in parallel, as in the alternate NAND design (Fig. 5(a)).

The AND expressions of the form $Y = A.B.C.D...$ are trickier. The logic function $Y = A + B.C$ offers good insights into implementing AND expressions. Fig. 5(d) shows a design that employs three gates driving a cantilever together with different weights. The truth table of the function $Y = A+B.C$ can be split into two parts, one part where maxterms $M_0$ to $M_2$ have the output 1 ($Y_1$), and the other part where maxterms $M_3$ through $M_7$ have an output 0 ($Y_0$). To implement $Y_0$, two conditions must be satisfied: (1) the input A must be capable of closing the switch individually; and (2) inputs B and C must be capable of pulling in the cantilever together. Thus, as shown in Fig. 5(d), the overlap area of A with the cantilever biased at GND is designed to be large enough to establish the appropriate electrostatic attraction, while the overlap area is large and equally divided between the inputs B and C. In the case of output of 1, the input A must not pull-in the cantilever by itself, but incorporate the force of B and/or C also. Therefore, the overlap area must be adjusted accordingly. If the overlap areas are specified in terms of weights, where a weight of 1.0 indicates that the input is capable of operating the cantilever by itself, then the weighted logic design of the function $Y = A+B.C$ can be represented as $Y_0 = \{A(1.0), B(0.5), C(0.5)\}$ for output of 0, and $Y_1 = \{A(0.75), B(0.25), C(0.25)\}$ for output of 1. The output of this design is then applied to a NOT gate to obtain the final output $Y$.

As the number of inputs in the AND term increases, the cantilever must be elongated for all inputs to be placed under the same cantilever. Design rules for a practical gate has limits of a minimum width, a minimum distance between two gates, and length of cantilever. These limits determine the number of inputs that can be implemented as an AND term. As fabrication challenges are addressed, it may become possible to have more than two gate inputs that actuate a single cantilever beam, thus forming more compact designs.

V. PUTTING IT TOGETHER - 32-BIT ADDER DESIGN

The advantages derived from using a custom NEMS circuit, when compared to a non-custom circuit, increase as the logic function becomes complex. The three circuits discussed above form the basic circuits in an n-bit adder circuit. By composing them into a 32-bit adder architecture, we can now quantify the speed-up obtained by using the weighted area logic scheme over the standard technique of design. Fig. 5 shows the designs for the base circuits of the adder. The AND gate (Fig. 5(b)) is designed as the previously described NAND gate followed by an inverter. The mechanical delay of this gate is two units – one unit due to the NAND gate and the second due to the NOT gate.
A. Speed-up and Energy

Amongst the various adder topologies proposed in the past, the Kogge-Stone adder has the lowest delay, and the Sklansky adder has the lowest energy-delay product [12]. By bringing logic functionality into the device structure, the WAL design strategy reduces the delay of a design, as well as power due to the reduction in number of devices. Fig. 6 shows the performance graphs. With the same simulation parameters as used for the NAND gate comparisons, the adder results show a 13% reduction in delay, 33% reduction in power, 42% in energy and about 50% reduction in the energy-delay product. Thus, the delay and power advantages combine to result in a significant energy-delay efficiency.

B. Area Reduction

The weighted area logic designs reduce the number of cantilevers required to implement the logic function. For example, in the 32-bit Kogge-Stone adder, the WAL design uses 40% fewer cantilevers than the standard design. This directly reduces the area of the designs.

The NEMS-CNT based designs also have the potential to reduce the area of logic functions compared to an equivalent CMOS implementation due to a number of reasons. The elimination of the requirements of different types of wells (P-type and N-type) is the primary reason for the area reduction. Secondly, the layout of multiple cantilevers in parallel can form regular grids, leading to an effective use of the chip area. Finally, the programmability of the circuits as NAND and NOR circuits can be expected to add to this reduction.

VI. CONCLUSION

Digital switches fabricated using the emerging technology of Nanoelectromechanical systems (NEMS) show promising characteristics for addressing the power issues in digital circuits. However their switching delay is considerably higher than electronic devices, limiting their use in very high-speed digital designs. In this paper, we addressed this issue at the circuit level, and proposed a novel design strategy called ‘weighted area logic (WAL) technique’ to perform logic operations in a purely electromechanical way, and to design combinational electromechanical circuits with cantilever-based NEMS switches. We showed that the design technique can lead to circuit implementations that are not only faster but also lower-power than implementations where NEMS devices replace CMOS devices directly. We presented the design and evaluation of logic functions that lead to a 32-bit adder circuit, and showed that the WAL-based adder has a lower delay, power consumption, energy and the energy-delay product when compared to a NEMS adder based on standard CMOS-like circuit design strategy.

The WAL design strategy for the NEMS technology incorporates the electromechanical device operation into logic computation and increases the functionality of a single device, transforming it from a simple switch into a logical switch. This innovative approach to the use of NEMS technology paves the way for electromechanical computation in a fast, integrated and low-power manner.

ACKNOWLEDGEMENTS
We thank Prof. T. Andrew Taton for his invaluable comments. This work was supported in part by the DARPA NEMS program.

REFERENCES