Nanowire Field-Programmable Computing Platform

Santosh Khasanvis, Mostafizur Rahman, Prasad Shabadi, Pritish Narayanan, Hyung Suk Yu, Chi On Chui, and Csaba Andras Moritz

***Abstract*—A nanowire-based field-programmable computing platform is presented featuring intrinsic fine-grained device-level reconfiguration without emulation (i.e. no look-up tables involved) using programmable cross-nanowire transistors, and regular physical implementation with relaxed manufacturing requirements at nanoscale. This approach can potentially provide orders of magnitude benefits in terms of area, power and performance vs. scaled CMOS FPGA at lower cost.**

Keywords—reconfigurable architecture; nanowires; programmable FETs.

I. Introduction

CMOS-based FPGA is one of the fastest growing market segments in the semiconductor industry due to post-manufacturing reconfigurability, low design cost and fast time-to-market. However, they are fundamentally inefficient when compared to custom CMOS ASICs because FPGAs *emulate* reconfiguration at the circuit-level with SRAM look-up tables (10x-50x inefficiencies in terms of area, power and performance [1]). In addition, the underlying CMOS fabric requires high manufacturing precision with aggressive scaling ($3\sigma=\pm1.3$nm critical dimension control and $3\sigma=\pm3$nm overlay precision for 16nm CMOS [2]) to keep variability and defect rates within acceptable limits, which is difficult to achieve at nanoscale. This is because it uses arbitrary sizing and doping of nanoscale MOSFETs, arbitrary circuit layouts, and complex interconnectivity. FPGAs will continue to remain inefficient with limited applications unless these aspects are considered.

In this concept paper, we present a nanowire (NW) based field-programmable platform (called N3P) with (i) *intrinsic* device-level fine-grained reconfiguration (without emulation through look-up tables) using programmable cross-NW field-effect transistors (pxnwFETs); and (ii) a uniform physical layer architecture with regular structures that relaxes manufacturing-precision requirements. This comprehensive solution can potentially have a tremendous socio-economic impact with considerable area, performance and power benefits compared to scaled CMOS FPGAs, and a much simpler (cheaper) manufacturing.

II. Nanowire Field-Programmable Platform

The proposed fabric incorporates regular NW arrays patterned on an SOI substrate with orthogonal metal gates (Fig. 1). Thus it leads to grids where every crosspoint acts as a pxnwFET with no specific functionalization during manufacturing. Each of these grids, called a NW stage/tile, is provided with input gates and output vias to connect to the metal stack, and acts as a template onto which circuits can be programmed post-manufacturing.

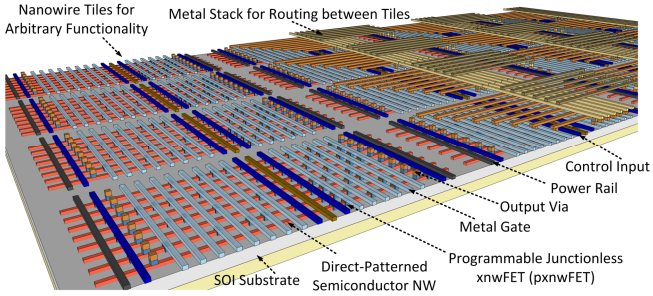

Fig. 1. Abstract 3D view of N3P with programmable NW stages at the bottom, and metal stack as routing layer.

A. Programmable Cross-NW Field-Effect Transistors

The programmability of cross-NW field-effect transistors (xnwFETs) is introduced by inserting functional engineered material (FEM) with memory function at the NW cross-point. While resistance-based FEMs have been a popular choice for crossbar memory, the appreciable leakage across the resistive cross-point poses an additional challenge besides the inherent lack of gain in its 2-terminal nature. Capacitance-based FEM, such as charge-trapping and ferroelectric gate dielectrics, might alternatively be incorporable into 3-terminal xnwFETs while fulfilling our stringent density, performance, and technology requirements. Capacitive charge-trapping stack [3] is among our top candidates yet its high operating voltage requirement and inability to be thinned down limits its versatility.

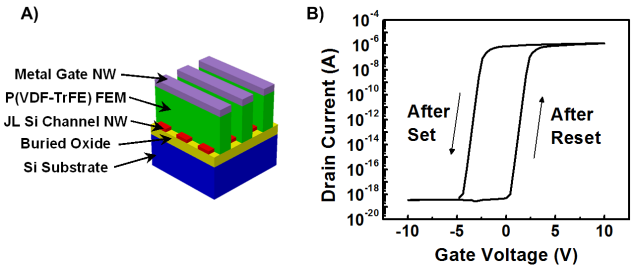

Fig. 2. A) An array of pxnwFETs with non-planar junctionless (JL) channel and organic ferroelectric dielectric formed above an SOI substrate; and B) Simulated hysteretic transfer characteristics of a pxnwFET showing wide memory window and steep subthreshold swing.

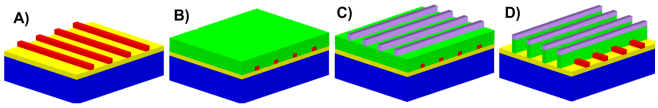

Fig. 3. Plausible low-thermal budget fabrication flow of pxnwFET array.

This work was supported by the Center for Hierarchical Manufacturing (CHM) at UMass Amherst.

H.S. Yu and C. O. Chui are with the Dept. of Electrical Engineering, University of California Los Angeles, CA.

The rest of the authors are with the Dept. of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003.

Email: khasanvis@ecs.umass.edu

Ferroelectric gate dielectrics can alternatively offer a wider reconfiguration window at lower voltages. Traditional ceramic ferroelectrics such as lead zirconate titanate (PZT) diffuse into the NW channel during crystallization annealing, causing instability. Organic ferroelectrics such as copolymer vinylidene fluoride and trifluoroethlene (P(VDF-TrFE)) can be reprogrammed at lower voltage yet have limited thermal stability during fabrication.

Given the needlessness of a junctionless channel NW to receive dopant implantation and high temperature defect annealing, its combination with P(VDF-TrFE) FEM constitutes a pxnwFET candidate (Fig. 2A) that offers wider memory window and excellent gate NW to channel NW control (Fig. 2B). As shown in Fig. 3, the pxnwFET structure can be implemented with junctionless Si channel NW arrays patterned on an SOI substrate. A conformal P(VDF-TrFE) layer is next spin-coated (and planarized). The metal gate NW arrays are then defined followed by self-aligned etching of P(VDF-TrFE) to minimize gate NW interference.

B. Field-Programmable Architecture

One possible organization for N3P is island-style architecture [4] with two main components – programmable logic and routing (Fig. 4A). The programmable logic blocks (LBs) incorporate NW tiles with pxnwFETs (Fig. 4B,C) that can be programmed to implement required logic directly (without emulation using look-up tables), using NW based dynamic circuit-style with single device type [5]. Initially, the NW tiles have all crosspoints acting as transistors. During programming phase certain crosspoints will be retained as transistors based on the targeted design, and other crosspoints will be de-functionalized such that they behave as if there are no transistors there.

The programmable routing is composed of switch blocks (SBs), connection blocks (CBs) and metal interconnects, where the SBs are used to create connections between different routing tracks (metal stack) and CBs provide access to the routing tracks for the LBs. The routing fabric will perform two functions – (i) serve as channels to program individual NW crosspoints in the N3P fabric during programming phase; and (ii) act as interconnections between NW tiles during operational mode for communication. An initial SB design using pxnwFETs is shown in Fig. 4D, which can route incoming signals from one direction to any one of three outgoing directions through switch-points (Fig. 4E). Each switch-point can be implemented with 6 pxnwFETs as pass transistors. As opposed to FPGAs which use SRAM (6 transistors or more per SRAM) in addition to pass transistors to store the configuration state, the SBs in N3P store configuration in the persistent state of pxnwFETs themselves and hence are very compact. These aspects allow N3P to be utilized in a similar manner to a conventional FPGA – but with a significantly improved efficiency due to no emulation.

C. Bootstrapped Programming/Reconfiguration Scheme for Functionality

Programming of N3P is envisioned to be completed in phases. The routing fabric (switch blocks, connection blocks

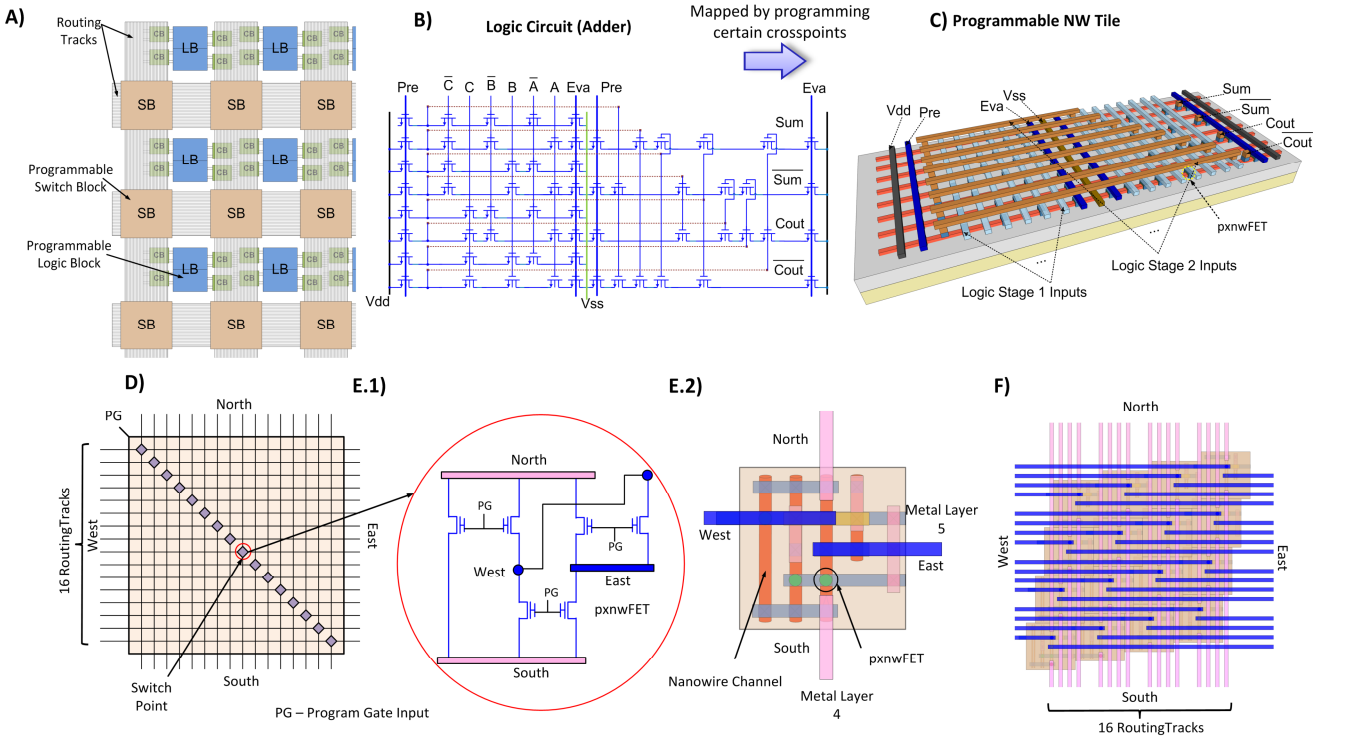

Fig. 4. A) Island-style architecture; B) Circuit example for logic (full-adder); C) LB with NW template to map logic circuits by programming certain crosspoints; D) 16x16 SB schematic: 16 horizontal and vertical routing tracks connected through 16 switch-points, each incoming track from one direction can connect to one of three outgoing directions through the switch-point; E.1) Switch-point circuit schematic with 6 pxnwFETs as pass transistors, where each pxnwFET can be programmed to be always ON or OFF based on connection; E.2) Abstracted view of a switch-point (SP) showing metal routing layers; F) Layout - Top view of 16x16 SB – the SPs are as in (E.2).

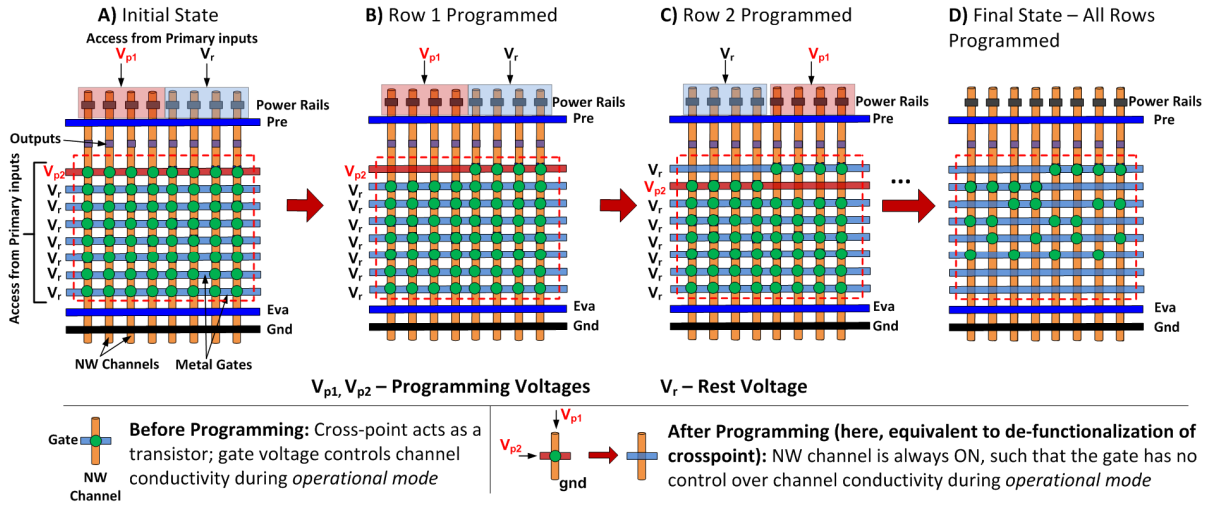

Fig. 5. Programming sequence example for the first stage of a 1-bit adder mapped to an N3P tile.

and metal interconnects) is used to access and program NW tiles and needs to be configured for reachability in the first phase. For a single programmable cross-nanowire FET (pxnwFET) this is straightforward. But in the presence of multiple NW tiles, we must program them in the correct order to ensure that we can reach every stage and every crosspoint. By programming NW tiles that are furthest from primary I/O ports, then the next furthest and so on, we ensure that we do not cut ourselves off from any NW grid. After all NW tiles have been programmed, the routing fabric (SBs and CBs) itself will be reconfigured to implement the required interconnections between NW stages as per desired circuits/system functionality. For a given NW tile, the programming of pxnwFETs can be done in a successive pattern starting from the first row. The pxnwFETs in a single row can be programmed in parallel when gated by a given wire simultaneously.

A programming example is shown in Fig. 5, where a template NW tile (logic block – LB) is programmed to implement the first stage of a 1-bit adder. Fig. 5A shows the initial condition where every crosspoint is a transistor. A crosspoint can be programmed (which is equivalent to de-functionalizing it in this case) by applying the correct combination of voltage bias across its terminals, such that after programming the channel is always ON and is not influenced by gate voltage under normal operational mode. The first row is programmed by de-functionalizing those crosspoints where the adder does not need a transistor. The remaining crosspoints are protected against unintentional de-functionalization by applying an intermediate voltage (V_r). Each row can then be programmed in succession, until all the rows are configured as shown in Fig. 5D.

D. Physical Layer – Mitigating Manufacturing Requirements

The key to addressing the nanomanufacturing challenge lies in designing with a mindset of simplifying the physical fabric itself. In N3P, a regular array structure is employed for the NW layer at the bottom. The a-priori assembly/direct-patterning of NW layer before any conventional lithographic step (e.g., for contacts/vias) means that (i) registration requirement is effectively minimized, allowing potentially cost-effective nanomanufacturing techniques to be used; (ii) overlay alignment requirements exist only between subsequent lithographic masks. Our prior work [6] shows this can be 3σ = 8nm for 16nm node and manufacturing techniques to achieve this are already optimized [2]. Junctionless channel transistors further simplify manufacturing and uniform connectivity at NW tile I/Os implies standard litho-design rules can apply between layers using standard vias [6].

III. Conclusion

A nanowire based field-programmable computing platform targeting CMOS FPGA replacement was presented. As part of our initial study, we mapped a small nanoscale processor WISP-0 [7] to N3P. Using scaling factors for FPGAs in [1], we estimate the density, power and performance benefits to be 54X, 50X, and 17X, respectively, with N3P following 16nm design rules vs. a 16nm scaled CMOS FPGA, with a relaxed overlay of 8nm vs. 3nm.

References

[1] H. Wong, et al., “Comparing FPGA vs. custom CMOS and the impact on processor microarchitecture”, Proceedings of the 19th ACM/SIGDA Inter. Symposium on Field Programmable Gate Arrays, pp. 5-14, 2011.

[2] “Lithography Technology Requirements,” International Technology Roadmap on Semiconductors, 2011. Available: http://itrs.org.

[3] S. Cui, et al., “High-quality Al_2O_3 for low-voltage high-speed high-temperature (up to 250 °C) nonvolatile memory technology,” IEEE Electron Device Letters, vol. 31, pp. 1443-1445, 2010.

[4] S. Brown, et al., "Field-Programmable Gate Arrays," Kluwer-Academic Publisher, Boston MA, 1992.

[5] P. Narayanan, et al., “CMOS Control Enabled Single-Type FET NASIC”, IEEE Computer Society Annual Symposium on VLSI, 2008.

[6] P. Panchapakeshan, et al., “3-D integration requirements for hybrid nanoscale-CMOS fabrics”, IEEE NANO, pp.849-853, 2011.

[7] T. Wang, et al., “Self-healing wire-streaming processors on 2-D semiconductor nanowire fabrics”, in Proceedings of Nanotech 2005, Nano Science and Technology Institute, Boston, MA, May 2006.