Nonvolatile Processors: Why is it Trending?

Fang Su\textsuperscript{1}, Kaisheng Ma\textsuperscript{2}, Xueqing Li\textsuperscript{2}, Tongda Wu\textsuperscript{1}, Yongpan Liu\textsuperscript{1}, and Vijaykrishnan Narayanan\textsuperscript{2}
\textsuperscript{1}Tsinghua National Laboratory for Information Science and Technology (TNList)
\textsuperscript{2}The Pennsylvania State University
Email: ypliu@tsinghua.edu.cn, vijay@cse.psu.edu

Abstract—Energy harvesting has become a promising solution to power up Internet-of-Things (IoT) devices. In this scenario, the constrained power budget and frequent absence of ambient energy cause severe reliability issues and performance degradation on conventional CMOS computing circuits. Fortunately, the advent of nonvolatile processor (NVP) opens the possibility to compute continuously using an intermittent power supply. It is considered as a key component of the next generation IoT edge devices. In this work, we provide insights to the evolution of the NVP and its application in real world scenarios. Efforts on improving the performance of NVP and future research prospects are also discussed in this paper.

Index Terms—Internet of Things (IoT), energy harvesting, nonvolatile processor (NVP)

I. INTRODUCTION

The notion of interconnected world has already seen a significant transition since the advent of the internet. The next major transition is happening currently with almost every object that we as humans interact with joining this interconnected world creating a boom in Internet-of-Things (IoT) [1]. More than 200 billion devices are projected to become part of the IoT by the year 2020 [2]. While they offer unprecedented opportunities to monitor, analyze and control the physical world with which we interact, powering up all these devices is a critical barrier for IoT deployment [3].

Batteries that have been adopted by most of today’s mobile devices is a potential candidate for this large market. However, batteries pose dimension, maintenance and pollution issues in IoT application spaces that would be much more ubiquitous than current day mobile systems [4]. Imagine the challenges that we already face in recycling batteries from our old mobile phones. Energy harvesting [5]–[9] has been widely investigated as a promising substitute for batteries. Energy scavenged from ambient environment such as vibrations, thermal differences, RF energy, solar power and delivered directly to the device. In such a scenario, computing devices have to operate sporadically rather than continuously due to the frequent absence of ambient energy [10].

Conventional CMOS circuits “forget everything” if power supply disappears [11], while retaining computational data in a sleep mode still incurs leakage power, resulting in waste of the precious harvested energy. Other remedies such as checkpointing system state into a remote nonvolatile memory (NVM) [12] suffers from low speed and large energy penalty. Fortunately, nonvolatile processor (NVP) [13] simultaneously meets the requirements of zero leakage and nonvolatility. By incorporating emerging nonvolatile technologies, NVP maintains temporary states within embedded nonvolatile flip-flops (NVFFs) during a power failure, and resumes back to computational tasks once power supply is recovered. It shows 1000× higher backup and recovery speeds when compared with conventional processors. Such features make NVP the key component of a battery-less energy harvesting IoT system, as shown in Fig. 1.

In this paper, we review the history and discuss the research trends and future prospects of NVP. The objective of this paper is to help researchers understand what has been done and what still remains to be addressed in the field of NVP.

The remainder of the paper is organized as follows. In Section II, we first revisit how a processor evolved into an NVP. Emerging applications of NVP, as well as their features and characteristics, are discussed in Section III. Ongoing research efforts from different aspects are presented in Section IV. Finally, Section V discusses the future research prospects and Section VI concludes the paper.

II. EVOLUTION OF NVP

The capability to maintain data and system states in the absence of input power is essential for an energy harvesting powered computing device. Due to the volatile nature of CMOS circuits, conventional processor [17] has no choice but to adopt an off-chip memory (such as Flash) for data backup, as shown in Fig. 2(a). Upon a power failure (resumption), the write (read) operations of the remote memory are slow and energy-consuming. In addition, the low endurance of Flash
memory (< 10⁶ write/erase cycles) precludes the processor from long-term autonomous service.

Advances in VLSI have made it feasible to integrate CMOS and emerging NVM [18] onto one die, as shown in Fig. 2(b). These NVM technologies, including ferroelectric random access memory (FRAM) [19], magnetic random access memory (MRAM) [20], phase-change memory (PCM) [21], resistive random access memory (RRAM) [22], offer a full range of benefits such as high density, low read/write energy, long endurance and 3-D integration compatibility. For example, the processor in [23] integrates an FRAM macro to copy system states in the event of a power failure. Nevertheless, this is not a fundamental solution since the bits held in flip-flops still have to be transferred into and read out from the centralized NVM in a sequential manner, resulting in significant energy and timing overheads.

The heavy burden of data movement severely limits the processor’s efficiency, especially given a tight energy budget in energy harvesting scenarios. To address this issue, the first NVP [13] is invented and fabricated through a CMOS/ferroelectric hybrid process, achieving 3 μs recovery time from power off. As Fig. 2(c) shows, an NVM element is attached to the standard flip-flop – the component which holds data in a processor – to form an NVFF and realize in-place data backup and restore. The main idea behind NVP is to replace the time-consuming and energy-inefficient byte-by-byte global data migration with a localized full-parallel bit-to-bit transfer. Compared with conventional (volatile) processors, NVP offers 10³× higher backup/restore speed and 10⁴× energy savings [13].

Along this line of thought, NVPs have been widely investigated with regard to various NVM technologies. In [24] and [25], ferroelectric NVPs are implemented at lower area cost and higher on/off switching speed than [13]. Nanoseconds backup and restore time is achieved by integrating MRAM in NVP [26]. Recently, RRAM based NVP [27] is reported, providing higher energy efficiency by adopting an adaptive retention scheme.

Aside from the emerging nonvolatile memory technologies, the advent of emerging beyond-CMOS logic transistors, such as tunneling field-effect transistors (TFETs) and negative capacitance field-effect transistors (NCFETs), has also brought great opportunities towards a new paradigm of future low-power nonvolatile computing. Those emerging devices, either showing enhanced Boolean logic operation with higher energy efficiency, or could actually be harnessed to redesign existing computing methods by introducing features beyond CMOS Boolean logic.

One example has already been shown in the region of NVP design in [31], where TFET transistors are used to replace CMOS transistors to operate at a lower voltage for higher computing energy efficiency, as well as for higher power conversion efficiency from ambient energy sources. It is shown in Fig. 3, that the adoption of TFET provides an average of 2.7× computation forward progress improvement over the baseline NVP design using LP CMOS. Similar advantage is also possible for other steep-slope devices [28], [33], [34].

One more exciting technology for NVP is the idea of embedding nonvolatility into the computing logic with NCFETs (Fig. 2(d)), in a means that logic gates could also store their states in a nonvolatile fashion: the output will be restored when supply is back after a power outage. Such a feature is a co-design of device and circuit, and have great potential of
further optimization at the architecture level. Fig. 4(a) shows an NCFET device structure. Fig. 4(b) shows a typical NCFET I-V curve with hysteresis around zero gate-source voltage. Fig. 4(c) shows the concept of a nonvolatile logic gate (an inverter here), in which the inverter is powered by GND and VDD. When the input D ranges from $V_{\text{LOW}}$ to $V_{\text{HIGH}}$, i.e. the gate-source voltage at the rising and falling edges of the NCFET I-V hysteresis, respectively, the output Q will not be changed, just like a memory. Q will only be updated when the input D is beyond the range of $[V_{\text{LOW}}, V_{\text{HIGH}}]$. Such a feature enables a dynamic operation of logic-memory synergy. More importantly, this memory is nonvolatile and will be capable of restoring the output when the power supply recovers after a power outage. Meanwhile, as is shown, it is different from conventional DFF-based designs in that it does not need a clock signal. Further exploration of NCFET nonvolatile computing would be very promising.

Table I lists the performance of silicon-verified NVPs as well as volatile processors. Here we highlight the active power, backup/restore time and energy as the most important performance indicators.

III. APPLICATIONS OF NVP

A. Why NVP: A Case Study

Before we dive into the discussion about applications, one question should be clarified – why NVP outperforms volatile processor? The following case study will answer the question and reveal the scientific reason behind NVP.

Suppose a body-heat powered smart patch wants to measure the exposure to ultraviolet (UV) radiation of human skin. Since the power income is extremely low (20 $\mu$W/cm²), the patch runs sporadically, not continuously, with the help of a capacitor $C_{\text{bulk}}$ as energy buffer (Fig. 5(a)). The system stays OFF until $C_{\text{bulk}}$ reserves enough energy for a sensing operation and a successful restore-backup pair.

For comparison, a volatile processor [17] and an NVP [13] are adopted, individually, in such system, and their parameters are listed in Fig. 5(b). Since NVP exhibits much lower backup/restore energy than volatile processor, it requires a much smaller $C_{\text{bulk}}$ (106 nF vs. 47 $\mu$F) and therefore a shorter charging time (52 ms vs. 22.3 s) for a single measurement, as shown in Fig. 2(c). As a consequence, NVP provides 400× data throughput boost than its counterpart. In other words, NVP converts the precious harvested energy into more valuable information, and that is the fundamental reason why we build NVP.

B. Application Domains

Self-powered IoT applications, which tend to be hard to reach [36], or limited in size [37], or require autonomous long-term service [38], can take full advantage of NVP. Also, the fast power-on response makes NVP useful in timing-critical systems such as video surveillance [39]. It’s noteworthy that

---

**Table I**

<table>
<thead>
<tr>
<th>Publication</th>
<th>Volatile Processor</th>
<th>NVP</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMOS Technology</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>NVM</td>
<td>Flash</td>
<td>FRAM</td>
</tr>
<tr>
<td>Active Power</td>
<td>450 $\mu$W/MHz</td>
<td>200 $\mu$W/MHz</td>
</tr>
<tr>
<td>Backup Energy</td>
<td>0.28 $\mu$J/bit</td>
<td>N/A</td>
</tr>
<tr>
<td>Restore Energy</td>
<td>0.37 $\mu$J/bit</td>
<td>N/A</td>
</tr>
<tr>
<td>Backup Time</td>
<td>6 ms</td>
<td>212 $\mu$s</td>
</tr>
<tr>
<td>Restore Time</td>
<td>3 ms</td>
<td>310 $\mu$s</td>
</tr>
</tbody>
</table>

---

**Fig. 5.** Case study: a body-heat powered UV smart patch.

---

**Fig. 6.** Application domains of NVP.
NVP is not restricted to low-power applications. The literally-zero standby power and fast on/off switching capability are also promising for fine-grained power management in high performance computers.

Fig. 6 presents the features and characteristics of three selected application domains of NVP: 1) structural health monitoring (SHM); 2) building security and 3) personal healthcare. One interesting observation is that the energy breakdown varies a lot among applications due to different task pattern and computation complexity. Therefore, energy consumption on 1) data backup/restore; 2) computation and 3) sensing and transceiving (or operations) should be simultaneously optimized. In addition, the energy harvester and power management circuits should also be carefully designed to increase the amount of energy scavenged from ambient environment and minimize conversion loss. Section IV will present the ongoing efforts to improve NVP from these aspects.

IV. ONGOING EFFORTS

A. Data Backup/Restore

Efforts on reducing backup/restore overheads of NVP can be categorized into three ways.

1) Minimize the frequency of backup/restore operations: A time-domain adaptive NVP is proposed in [27], wherein data retention is preferred rather than backup/restore to survive “short” power interruptions. The failure times can also be reduced by scaling the supply voltage or operating frequency [40], [41] in accordance with input power, which will be elaborated in the next subsection.

2) Reduce the amount of data to be stored: Techniques exploiting data pattern have been reported to eliminate redundant backup/restore of unused bits [27], [42]. Compression based approaches [43], [44] also helps to save the read/write energy.

3) Optimize bit-level backup/restore cost: Per-bit backup and restore cost can be reduced with advanced nonvolatile materials and more efficient read/write circuits [45]–[49]. Among those approaches, self-write-termination (SWT) [27] is a promising one to save write energy as well as prolong the lifetime of RRAM based NVFFs.

The motivation for SWT is to handle the large switching time variation of RRAM devices and to eliminate unnecessary store operations (e.g. the original state of RRAM matches the data to be stored). As Fig. 7(a) shows, the fast and slow devices show more than 100× difference in switching time. A fixed SET/RESET pulse will induce 1) energy waste and 2) degraded lifetime (Fig. 7(b)). The SWT scheme pre-senses the device state through a feedback path and terminates the SET/RESET process if the RRAM device is already in the target state. Fig. 7(c) shows the schematic of a SWT-NVFF, which achieves up to 172× reduction in write energy.

B. Computation

We define forward progress [15] as instructions committed within NVP, which roughly equals to the energy used for computation divided by instruction (EPI). In order to transform as much harvested energy as possible, we deploy dynamic computation in a NVP system as shown in Fig. 8. It consists of a bottleneck resource predictor, a proper frequency predictor, and a modified NVP with support to adjustable resources and frequency. In a traditional NVP [34], [50], [51], we have two observations. On one side, dissipating energy slowly may cause the storage capacitor to have a large potential to be full, reducing the space for future energy boost. On the other side, consuming the energy too fast results in frequent energy emergency and eventually frequent backup operations.

Frequency scaling is one of the possible alternatives to aggrandize the partition of energy transformed into computational energy by dynamically adjusting the energy dissipating speed. The frequency maintains at a very low level to make the NVP run over most of the power outages. While the frequency boosts significantly to convert more energy boost into forward progress when abundant energy is monitored. The concern may lay in the policy for proper frequency proposal to change along with the harvested power and stored energy.

Another possible solution for better forward progress is to reduce EPI by powering up some essential resources so as to boost the instruction per cycle in an Out-of-order processor. We observe that different testbenches has different sensitivity to certain resources. And providing more resources may significantly boost the instruction per cycle in an Out-of-order processor. How to identify the key resources for the specific testbench and when to power on these resources remain an interesting topic. The potential to merge the frequency scaling and resource allocation is challenging but remains possible.

C. I/O Operations

I/O operations refer to the interaction between NVP and peripheral devices, which consist of three phases: 1) I/O bus initialization; 2) peripheral configuration and 3) data exchange. These processes have to be repeated if attacked by a power failure. Various approaches have been employed to maintain or restore I/O bus status [12], [52]–[54], but the recovery of the latter two still remains to be addressed.
Consequently, a higher current is drawn from the harvester frequency controller (CFC) is proposed in [58]. As Fig. 10(b) shows, when sufficient energy is available and the supply voltage rises, the CFC increases the clock frequency of NVP. Consequently, a higher current is drawn from the harvester until the system reaches equilibrium. 162% overall efficiency enhancement is observed at negligible area penalty.

V. FUTURE WORKS

Continuous performance improvement on NVP can be expected with advances in beyond-CMOS devices and NVM technologies. Meanwhile, the energy consumption on sensors and analog processing circuits – as the bridge connecting physical and digital world – will become the bottleneck and necessitate cautious redesign towards the IoT era. High-level optimization like architectural exploration, software support and system-on-chip integration are also necessary to take full advantage of NVP. Last but not least, new computing paradigms enabled by emerging devices and NVMs, such as non-Boolean logic [59], [60], in-memory processing [61] and neuro-inspired computing [62], may bring about a fundamental change on how NVP works and thus require a device-circuit-application co-design.

VI. CONCLUSION

NVP is a promising technology for energy harvesting powered IoT edge devices, as it enables continuous and reliable operation in spite of the unstable and intermittent power supply. This paper provides an overview of the research in NVP and reveals the fundamental road map from conventional processor to NVP. In this paper, we also discuss emerging application domains of NVP, and present ongoing efforts to enhance its performance. We believe this paper will help researchers better understand this new born area, and motivate further researches towards optimization of NVPs and their applications.

REFERENCES


X. Li et al., “RF-powered systems using steep-slope devices,” in *New Circuits and Systems Conference (NEWCAS)*, 2014, pp. 73–76.


S. Baldauf et al., “A 32-bit CPU with zero standby power and 1.5-clock sleep/2.5-clock wake-up achieved by utilizing a 180nm c-axis aligned crystalline In-Ga-Zn oxide transistor,” in *Symposium on VLSI Circuits Digest of Technical Papers*. IEEE, 2014, pp. 1–2.


