A Reconfigurable Analog Substrate for Highly Efficient Maximum Flow Computation

Gai Liu and Zhiru Zhang

School of Electrical and Computer Engineering, Cornell University, Ithaca, NY

{gl387, zhiruz}@cornell.edu

Abstract

We present the design and analysis of a novel analog reconfigurable substrate that enables fast and efficient computation of maximum flow on directed graphs. The substrate is composed of memristors and standard analog circuit components, where the on/off states of the crossbar switches encode the graph topology. We show that upon convergence, the steady-state voltages in the circuit capture the solution to the maximum flow problem. We also provide techniques to minimize the impacts of variability and non-ideal circuit components on the solution quality, enabling practical implementation of the proposed substrate. Performance evaluation demonstrates orders of magnitude improvements in speed and energy efficiency compared to a standard CPU implementation.

1. Introduction

Overcoming the performance and efficiency limitations of traditional von Neumann architecture has been one of the major themes in the recent computing research. Exploring novel architectures and computing paradigms becomes even more imperative as power-limited technology scaling fails to offer the same performance improvement that we have seen in the past few decades [10]. This trend is motivating researchers to investigate new technologies and alternative models of computations to provide the next wave of major improvements in computing. An active direction of research seeks to improve performance and energy efficiency through domain-/application-specific hardware customization using specialized accelerators such as ASICs, FPGAs, and GPGPUs [6]. Parallel to this direction, more disruptive computing paradigms are being examined, including approximate computing [14, 22], neuromorphic computing [23], and quantum computing [19]. These disruptive computing paradigms are expected to achieve significant improvements in performance and energy efficiency for certain domains of applications compared to the traditional approaches.

In this work we investigate using a physics-based analog computing system as an unconventional approach to an important optimization problem — the maximum flow (max-flow) problem. The max-flow problem is one of the most pervasive optimization problems in applications such as transportation [27], Internet traffic scheduling [29], VLSI CAD [7], and many other emerging applications including computer vision [4] and data mining [11]. These emerging applications often deal with large-scale graphs, which need to be analyzed under ever-demanding performance and energy efficiency requirements. Needless to say, a new method that solves the max-flow problem efficiently in both time and energy consumption would be of great interest to the CAD community as well as the field of computing in general. However, the max-flow problem has been shown to be computationally demanding [13] and difficult to parallelize [12]. To this end, we propose a reconfigurable analog substrate composed of traditional analog devices including resistors, diodes and operational amplifiers (op-amps), as well as an emerging device called memristor. The substrate encodes graph topology using the on/off states of the memristor switches, and leverages Kirchhoff's circuit laws to enforce flow constraints, which are represented in the form of circuit node voltage values. The max-flow objective, defined as the net flow out of the source vertex, is then driven to its maximum value. Upon convergence, the steady-state node voltages represent the solution to the max-flow problem. We note that in practice, the solutions obtained from the substrate are often approximations to the original problems due to process variation and non-ideal circuit components. We address this problem by proposing techniques to mitigate the influence on the solution quality due to the non-ideal factors. We summarize our contributions as follows:

1. We present the first study of a reconfigurable analog substrate that efficiently solves max-flow problems.
2. We prove that the proposed substrate optimally solves the max-flow problem under ideal assumptions.
3. We analyze the influence of non-ideal circuit components and process variation on the solution quality, and propose techniques to mitigate the impact of these non-ideal factors.

The rest of the paper is structured as follows: Section 2 describes the circuit mechanism of our substrate; Section 3 introduces the reconfigurable crossbar architecture; Section 4 discusses the implementation details of the substrate; Section 5 provides the performance evaluation of our proposal; Section 6 reviews the related work and the paper concludes in Section 7.

2. Computing Max-flow using Analog Circuit

The max-flow problem we are interested in solving is stated as follows: Given a directed graph[1] $G = (V, E)$ with $n = |V|$ vertices and $m = |E|$ edges, we distinguish two vertices, a source vertex s and sink vertex t, and assign each edge e a nonzero integral capacity c_e. A $s - t$ flow is a function $f : E \rightarrow \mathbb{R}$ such that for every vertex n_i other than s and t, the flow coming into n_i is equal to the flow going out of n_i, and $0 \leq f(e) \leq c_e$ for all $e \in E$. The value $|f|$ of a $s - t$ flow is defined to be the net flow out of the source vertex. The max-flow problem asks for finding a feasible $s - t$ flow in G of the maximum value.

The circuit used to compute max-flow is constructed in a direct mapping manner, i.e., for every edge, there is a circuit component to ensure edge capacity constraint; and for every node, there is another circuit component to enforce flow conservation constraint; there is also a circuit component for realizing the functionality of max-flow objective function.

2.1 Edge Capacity Constraint

For each edge capacity constraints on edge x_i, we create a circuit component composed of two diodes and one voltage source, as

[1] Max-flow problem on an undirected graph can be converted to an equivalent problem on a directed graph by allocating two opposite edges with the same capacity as the edge in the original undirected graph.

DAC '15, June 07 - 11, 2015, San Francisco, CA, USA

http://dx.doi.org/10.1145/2744769.2744781

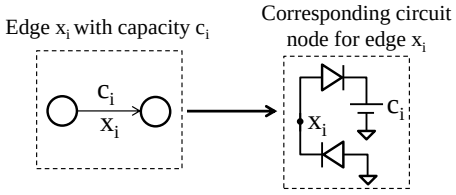


Figure 1: Circuit structure for edge capacity constraint.

shown in Figure 1, where the capacity on edge x_i is c_i. One of the two diodes is directly connected to the circuit node x_i, with its anode connected to ground. Ideally, the diode will turn on whenever the voltage on circuit node x_i is below zero[2]. Since the ideal diode in on-state behaves like an ideal wire, the voltage on circuit node x_i will always be non-negative. Similarly, the other diode and a voltage source of value c_i enforce that the circuit node voltage at x_i never exceeds c_i. Thus we have the following relation

$$0 \leq V_{x_i} \leq c_i \tag{1}$$

for every node x_i that corresponds to an edge x_i in the original graph, where V_{x_i} is the voltage value on circuit node x_i, and c_i is the corresponding edge capacity on edge x_i.

2.2 Flow Conservation Constraint

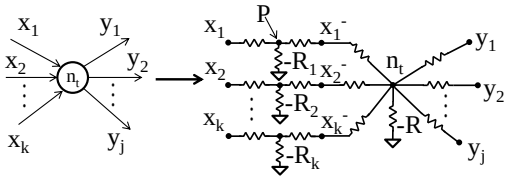


Figure 2: Circuit structure for flow conservation constraint.

Flow conservation constraint is enforced using the circuit in Figure 2. For each outgoing edge $y_i (i = 1, 2, \ldots, j)$ of a vertex n_t in the original graph, the corresponding circuit node y_i with voltage value V_{y_i} is connected to node n_t via a resistor; while for each incoming edge $x_i (i = 1, 2, \ldots, k)$ of the vertex n_t, the corresponding node of negated voltage value $V_{x_i^-}$ is connected to node n_t. All positive resistors have the same resistance value r. Negative resistors $-R_1, \ldots, -R_k$ all have the same resistance value of $-r/2$, and the negative resistor connected to n_t has resistance value $R = \frac{r}{N}$, where $N = j + k$. We can show that the circuit in Figure 2 indeed enforces flow conservation constraint on node n_t using Kirchhoff's current law (KCL):

Denote the voltage value on node P in Figure 2 as V_P, we have the following relation using KCL:

$$\frac{V_{x_1} - V_P}{r} + \frac{V_{x_1^-} - V_P}{r} = \frac{V_P}{-R_1} \tag{2}$$

Since $-R_1 = -r/2$, we have

$$V_{x_1} = -V_{x_1^-} \tag{3}$$

Similarly, we have $V_{x_i} = -V_{x_i^-}$ for all $i = 1, 2, \ldots, k$.

Based on KCL we also have the following relation on node n_t:

$$\sum_{i=1}^{k} \frac{V_{x_i^-} - V_{n_t}}{r} + \sum_{i=1}^{j} \frac{V_{y_i} - V_{n_t}}{r} = \frac{V_{n_t}}{-R} \tag{4}$$

Plug $R = \frac{r}{k+j}$ into Equation (4):

$$\sum_{i=1}^{k} V_{x_i^-} + \sum_{i=1}^{j} V_{y_i} = 0 \tag{5}$$

[2] Note that in practice we need to account for the turn-on voltage of the diode by adjusting the values of the voltage sources.

Using the result from Equation (3), we obtain the following relation that satisfies flow conservation constraint: $\sum_{i=1}^{k} V_{x_i} = \sum_{i=1}^{j} V_{y_i}$.

2.3 Max-flow Objective

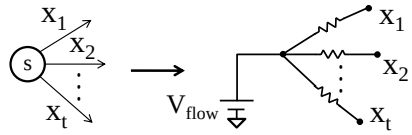


Figure 3: Circuit structure for implementing max-flow objective.

The objective function is implemented using the circuit in Figure 3, where all the positive resistors have the same resistance value of r. Intuitively, V_{flow} will drive the node voltages as large as possible, subject to the capacity and conservation constraints. As a result, the steady-state voltages of the circuit nodes reflect the maximum possible flow through the network that does not violate the requirements of the constraints. In the next section, we derive the steady-state solution of the circuit using basic circuit theorems, and show that the steady-state node voltages represent the solution for the max-flow problem. We will first prove that the value of $s - t$ flow will strictly increase as V_{flow} increases by showing that the constructed resistor network is passive. Combining this fact with the flow conservation constraint and edge capacity constraint derived in Sections 2.1 and 2.2, we can show the optimality of the circuit solution.

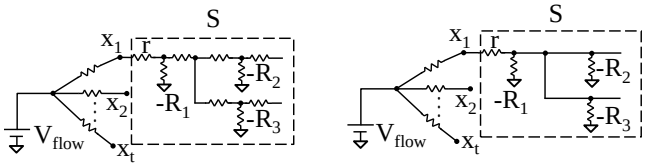
2.4 Optimality of Circuit Solution

We first show that the resistor network in the max-flow circuit is passive, i.e., the equivalent resistance of the resistor network is positive. Figure 4a shows a generic circuit for solving the max-flow problem, where nodes x_1 to x_t are the subset of nodes that are directly connected to V_{flow}. In Figure 4a, we mark the resistor that directly connects to node x_1 with its resistance value r. We also list all the negative resistors in the network as $-R_1, -R_2, \ldots, -R_k$. Note that according to the construction of the circuit, we have $|-R_i| < r$ for all the negative resistors. Now we find the equivalent resistance of the resistor subnetwork S to the right of node x_1, shown in the dashed box in Figure 4a. Based on the fact that reducing the resistance value of positive resistors to zero will only decrease the equivalent resistance of the network, we set all the resistance of the positive resistors except r in S to be zero, which only decreases the equivalent resistance of S. The resulting resistor network consists only of the resistors shown in Figure 4b. Since each of the resistance $|R_i|$ is strictly less than r, all the parallel negative resistors together also have an equivalent resistance that is less than r. As a result, the equivalent resistance of S, which is the sum of r and the equivalent resistance of these negative resistors, is always positive. Figure 4c shows the equivalent circuit of Figure 4a, where $R_{eq(i)} > 0$ for all $i = 1, 2, \ldots, t$.

Each circuit branch in Figure 4c is a basic voltage divider. Since all the remaining positive resistors have a resistance value of r, we have $V_{x_i} = V_{flow} \frac{R_{eq(i)}}{R_{eq(i)} + r}$ for all $i = 1, 2, \ldots, t$. Note that we have shown $R_{eq(i)} > 0$ for all $i = 1, 2, \ldots, t$, thus the steady-state solution V_{x_i} will strictly increase as V_{flow} increases, as long as edge capacity constraints and flow conservation constraints are satisfied. Summing up all the node voltages, we have

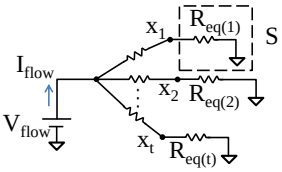
$$\begin{aligned} \sum_{i=1}^{t} V_{x_i} &= V_{flow} \sum_{i=1}^{t} \frac{R_{eq(i)}}{R_{eq(i)} + r} = V_{flow} \sum_{i=1}^{t} (1 - \frac{r}{R_{eq(i)} + r}) \\ &= tV_{flow} - rI_{flow} \triangleq J \end{aligned} \tag{6}$$

J in Equation (6) plays the role of objective function in the max-flow problem. V_{flow} in the circuit is set to a high voltage value,



(a) Original resistor network.

(b) Resistor network after reducing the resistance of certain positive resistors to zero.



(c) Lumped resistor network with equivalent resistance.

Figure 4: Replacing resistor network with equivalent resistance.

so that V_{flow} will try to maximize the node voltages in the circuit. Combining this result with the derived edge capacity constraint and flow conservation constraint, we have:

$$J = tV_{flow} - rI_{flow} = \sum_{i=1}^{t} V_{x_i} \tag{7a}$$

$$\sum_{i=1}^{k} V_{x_i} = \sum_{i=1}^{j} V_{y_i} \tag{7b}$$

$$0 \leq V_{x_i} \leq c_i \tag{7c}$$

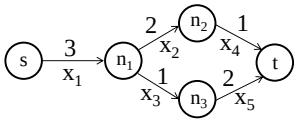
where Equation (7b) holds for all vertices with incoming edges x_i and outgoing edges y_i, and Equation (7c) holds for all edges in the graph. Equation (7) is identical to the original max-flow problem. Thus we conclude that the circuit solution is equivalent to the solution of the corresponding max-flow problem.

2.5 An Example

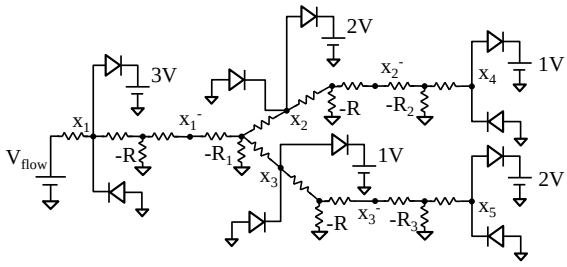
Figure 5b shows an example circuit that solves a max-flow problem instance in Figure 5a. A step function is first applied to V_{flow} that drives V_{x_1} to its maximum value $3V$. Then the circuit components implementing flow conservation constraints start working together to bring the other nodes to steady state. V_{x_3} and V_{x_4} will reach their maximum values of $1V$, corresponding to the case that flows on edge e_{x_3} and e_{x_4} reach their maximum value of 1. Then V_{x_1} and V_{x_2} in return bring V_{x_1} to its final value of $2V$. Figure 5c shows the waveform of the node voltages as a function of time.

3. Reconfigurable Crossbar Architecture

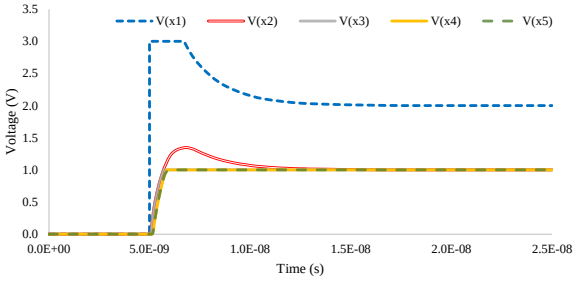
Constructing a custom circuit for each graph instance is generally impractical as fabricating a custom circuit is both expensive and time consuming. We propose to construct a reconfigurable substrate that can be configured to encode different graph topologies, enabling the solution of various problem instances using a single substrate. We propose to use memristors as the reconfigurable switches in our substrate. The memristor [16] is a two terminal device with a metal-insulator-metal sandwiched structure whose resistance (called memristance) can be modulated by external stimulus. Memristor can be switched between the high-resistance state (HRS) and the low-resistance state (LRS) by external voltage stimulus, and retains its resistance value when external stimulus is removed or below a certain threshold. Our reason for choosing memristors as switches is twofold: (1) memristor-based switches are more area-efficient than the traditional SRAM-based switches; (2) since resistors are widely used on our substrate, we can directly replace the resistors with memristors so that one single memristor acts both as a switch and as a resistor. Namely, A memristor in HRS



(a) An example max-flow problem.



(b) The corresponding circuit.



(c) Waveform of the node voltages.

Figure 5: An example of solving max-flow using analog circuit.

represents a disconnected switch; A memristor in LRS represents both a connected switch and a resistor whose resistance is equal to the LRS memristance. Additionally, using LRS memristors as normal resistors enables the fine-tuning of memristance in LRS, which helps mitigate the impact of process variation and parasitic resistance, as will be discussed in Section 4.3.2.

Our substrate is composed of an $n \times n$ crossbar as shown in Figure 6. At each intersection (n_i, n_j), there is a small circuit widget connecting into the crossbar through a memristor switch. This circuit widget represents an edge x_{ij}. Essentially, we can view the crossbar as a physical representation of the adjacency matrix of the graph. If edge (i, j) is present in the graph, the memristor switch at position (n_i, n_j) will be set to LRS. Otherwise, the memristor switch at position (n_i, n_j) will be set to HRS. When properly configured, the crossbar architecture implements the circuit described in Section 2. We implement the max-flow objective using the first row of the crossbar, where the memristor switch at position (s, n_i) will be turned on if and only if edge (s, i) is present in the graph. We enforce the edge capacity constraint using two diodes at each intersection, and honor the flow conservation constraint by connecting column n_i with all the edges incident to n_i. Figure 7 shows the mapping of the example circuit to the crossbar. By examining the circuit widgets at the intersections, we can verify that the crossbar correctly implement the circuit in Figure 5b.[3]

3.1 Configuring the Crossbar

In the configuration stage, we program the memristors in the crossbar to the desired resistance state using voltage pulses. V_{flow} is set to zero in this stage. The programming stage takes n cycles to complete, one cycle for each row. At the i^{th} programming cycle, row i is activated by setting the row wire to a low voltage V_{low}, while keeping all the other rows at 0V. For every memristor that needs to be set to LRS, the corresponding column will be set to a high volt-

[3] Since there is no outgoing edge from sink t, the negative resistors in the circuit widgets of column t can be omitted without affecting the correctness of the circuit.

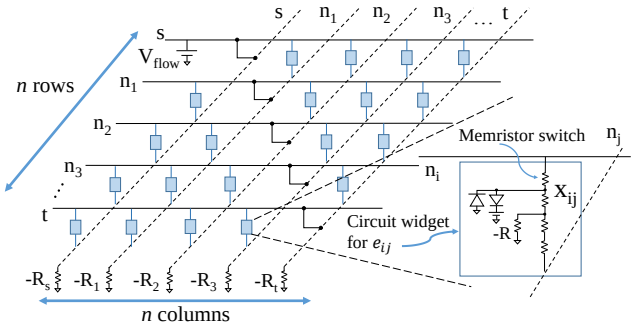


Figure 6: Reconfigurable crossbar architecture.

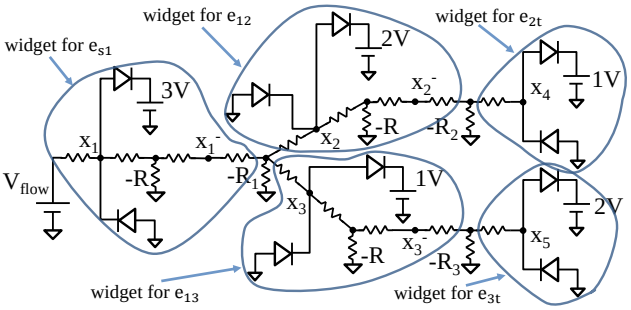


Figure 7: Mapping the example circuit to the crossbar.

age V_{high}, while keeping all the other columns at 0V. By setting the voltage difference $(V_{high} - V_{low})$ greater than the memristor threshold voltage, the desired memristors will be set to LRS, while all the other memristors stay in HRS. After n cycles of configuration, all the memristor switches are set to the desired state.

3.2 Computing Max-flow on the Crossbar

At the beginning of the computing stage, a step function is applied to V_{flow}, and the circuit starts converging to steady state. To avoid the complexity of detecting steady-state condition, we run the circuit for a period of time t. t should be long enough to ensure that the circuit converges, which can be determined by profiling the worst-case execution time of the substrate. After convergence, we measure the current through the voltage source V_{flow} and calculate the objective value (i.e. value of max-flow) based on Equation (7a).

4. Implementation

We have shown that the proposed substrate generates optimal solution for max-flow problems under several assumptions: (1) one voltage source for each edge with exact voltage level; (2) ideal negative resistors and diodes; (3) no process variations or parasitic resistance. In this section, we analyze the impact of these assumptions on the solution quality, and propose techniques that enable a practical implementation of the proposed substrate.

4.1 Voltage Level Quantization

In practice, it is too expensive to use one distinct voltage source on the substrate for each edge. Also, the output voltage level of a voltage source cannot be arbitrarily large. To address these two issues, we propose the following quantization scheme that maps the edge capacities to a set of discrete voltage levels. Given a supply voltage of value V_{dd}, we uniformly divide the voltage interval $[0, V_{dd}]$ into N voltage levels: $\mathbf{V} = \{\frac{1}{N}V_{dd}, \frac{2}{N}V_{dd}, \ldots, \frac{N-1}{N}V_{dd}, V_{dd}\}$. For each voltage level in $\mathbf{V}$, we construct one voltage source with the corresponding voltage value. One voltage source will be used for multiple edges if these edges share the same voltage level after quantization. The quantization function $\mathbf{Q}$ uniformly maps the edge capacities of a max-flow instance into the voltage levels in set $\mathbf{V}$. Specifically, given an edge capacity x, $\mathbf{Q}$ maps x to one of the voltage levels in $\mathbf{V}$: $\mathbf{Q}(\mathbf{x}) = \frac{\lfloor (x/C)N \rfloor}{N} V_{dd}$, where C is the largest edge capacity of the max-flow instance. The solution $\mathbf{Y} = \{V_{x_1}, V_{x_2}, \ldots, V_{x_m}\}$ obtained from the circuit is then mapped back to the interval $[0, C]$: $\tilde{\mathbf{Y}} = \mathbf{Y} \times \frac{V_{dd}}{C}$. $\tilde{\mathbf{Y}}$ is an approximate solution of the original problem with quantization errors. The worst-case quantization error e of the flow on one edge is bounded by the quantization step: $e = \frac{C}{N}$ with C being the largest edge capacity and N being the number of voltage levels. Note that the choice of N provides a trade-off between accuracy and cost—increasing N reduces the worst-case quantization error, but also increases circuit complexity and cost. Figure 8 illustrates the application of voltage quantization on the same graph in Figure 5a with $N = 20$ and $V_{dd} = 1V$, the result shows a 5% deviation in the final flow value $|f|$.

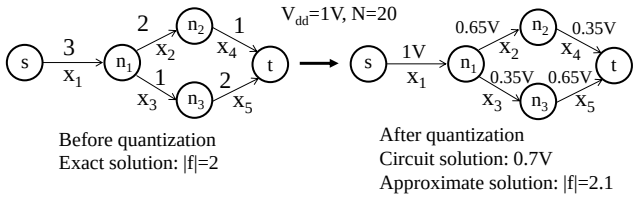


Figure 8: An example of voltage level quantization.

4.2 Negative Resistors

We can view a negative resistor as a current-controlled voltage source with the transresistance equal to $-R$, which can be implemented using an op-amp, as shown in Figure 9a. The precision of the negative resistor is determined by the open loop gain A of the op-amp. To derive the precision of the negative resistor, we calculate the effective resistance R_{eff} of the negative resistor as $R_{eff} = -(1 + \frac{1}{A}\frac{R_0}{R_{target}})R_{target}$, where $\frac{R_0}{R_{target}}$ is a constant that is close to 1. Thus the precision of the negative resistor is inversely proportional to the gain A of the op-amp. Since modern op-amps can easily achieve gain $A > 1000$, indicating a precision of $\pm 0.1\%$, we conclude that the implementation of op-amp has negligible impact on the precision of the negative resistor.

4.3 Process Variation and Parasitic Resistance

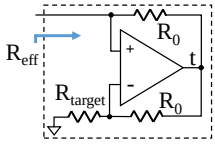
The actual resistance of the resistors deviates from their nominal values due to unavoidable process variation and parasitic effects, which degrade the solution quality. In this section, we first make an important observation that the solution quality depends only on the ratio of the resistance values instead of the absolution resistance values. This enables the use of layout matching techniques to minimize variations. We will also discuss post-fabrication resistance tuning to further reduce the impact of resistance variation.

4.3.1 Resistance Matching

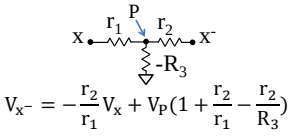
Although most integrated resistors have tolerances of ± 20 to 30%, the tolerance of the ratio between two resistors can be controlled to be better than $\pm 1\%$, and in many cases, to be within $\pm 0.1\%$ [15]. We have shown in Section 2.4 that the circuit node voltages are only functions of ratios of resistances. Since the circuit node voltages in steady state represent the flow value, we are encouraged to discover that the solution quality is also only a function of resistance ratios, independent of the absolute resistance values. Consequently, the substrate is expected to be highly insensitive to process variation. However, resistance compensation techniques are still needed to counter the other sources of variations such as parasitic resistance.

4.3.2 Resistance Tuning

The fact that all the resistances in the circuit are realized by memristors makes it possible to conduct post-fabrication fine-grained resistance tuning [21]. This is especially useful to minimize the



(a) Negative resistor using operational amplifier.



(b) Resistance tuning circuit.

Figure 9: Negative resistor and the resistance tuning circuit.

influence of parasitic resistance. We outline the tuning process as follows.

To begin with, we configure the substrate to implement the tuning circuit shown in Figure 9b, which is a simple circuit that enforces $V_x = -V_{x^-}$. For each of the tuning circuit, we first set $V_x = 0$ and modulate R_3 until $V_{x^-} = 0$. This step ensures $\frac{1}{R_3} = \frac{1}{r_1} + \frac{1}{r_2}$. We then set $V_x = 1V$, and collectively change the values of r_1 and r_2 until $V_{x^-} = -1V$. These two steps can be iterated for a few times for better tuning precision. We note that in principle the tuning process has to be done only once right after fabrication, thus no performance overhead is incurred afterwards. However, since the memristance may slowly drift over a long period of time, the tuning process can be repeated based on the endurance of the memristors.

5. Performance Evaluation

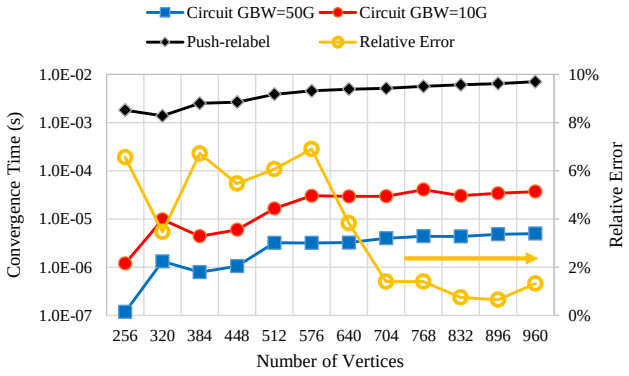
We construct a 1000×1000 substrate in circuit-level netlist and simulate the substrate in SPICE [25]. The detailed parameters used in the simulation are listed in Table 1. The parameters for memristors and op-amps are set to typical values from recent literature [3, 9]. We will first evaluate the execution time for various benchmarks, then estimate the power consumption of the substrate using an analytical method.

Table 1: Design parameters for the max-flow computing substrate.

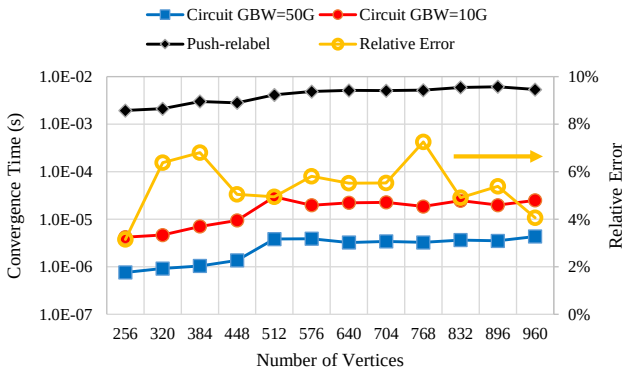
Memristor LRS resistance ($k\Omega$)	10
Memristor HRS resistance ($k\Omega$)	1000
Objective function voltage V_{flow} (V)	3
Open loop gain of op-amp	1×10^4
Gain-bandwidth product of op-amp (GHz)	10 to 50
Number of columns in the crossbar	1000
Number of rows in the crossbar	1000
Number of voltage levels	20

5.1 Convergence Time

The convergence time of the substrate is measured as the time interval between the rising edge of V_{flow} and the timestamp when the flow value is within 0.1% of the final value. To model the effect of parasitic capacitances, we add a parasitic capacitance of 20fF to each circuit net. We compare the convergence time against the execution time of the widely used push-relabel algorithm running on a server with 3GHz Intel Xeon processor and 16GB of RAM. The push-relabel algorithm is compiled using GCC 4.4.7 with -$O3$ optimization option. When measuring execution time on CPU, we exclude the time taken to read input file and the time taken to generate internal data structure for fair comparison. We use R-MAT synthetic graph generator [5] to generate both dense ($|E| \propto |V|^2$) and sparse ($|E| \propto |V|$) graphs. The number of nodes in the generated graph ranges from 200 to 1000, and the number of edges ranges from 500 to 8000. We plot the convergence time of the substrate and the execution time of the push-relabel algorithm in Figure 10. Using op-amps with 10G gain-bandwidth product (GBW), our substrate achieves a range of speedups between $150X$ and $1500X$ compared to the software implementation. If we further increase the GBW to 50G, we can achieve an additional $10X$ speedup. Such dramatic speedup is not too surprising as the circuit operates vastly different and more efficient than the CPU — instead of sequentially



(a) Dense graph: $|V| \propto |E|^2$.



(b) Sparse graph: $|V| \propto |E|$.

Figure 10: Convergence time of synthetic benchmarks.

executing millions of instructions, the circuit converges to the optimal solution in a massively concurrent fashion, with physics being the driving force, e.g., pulling up and pulling down of the node voltages. We also plot the relative error of the circuit solution compared to the optimal solution in Figure 10. The deviations from the optimal solutions are due to the quantized voltage levels and non-ideal circuit components. We can observe that in all cases the relative error is within 8%, with an average relative error of 3.7% for dense graphs and 5.4% relative error for sparse graphs.

5.2 Analysis of Power and Energy Consumption

We analytically model the power consumption of the substrate with the goal of showing the general trend of power consumption as a function of circuit size. The model provides insights on how a given power budget impacts the overall design. The majority of power is consumed by op-amps and resistors. For resistors, we observe that it is possible to proportionally scale up the resistor values without influencing the correctness of the solution (Section 4.3.1). Scaling up the resistance reduces the average current through the resistors, thus reducing the power consumption of the resistors. Consequently, the op-amps are the major source of power consumption. According to the circuit structure, one op-amp is needed for each edge that is present in the graph.[4] Additionally, one op-amp is needed for each node to enforce flow conservation constraint. Thus the power consumption for solving a graph of $|V|$ nodes and $|E|$ edges is roughly $(|E| + |V|)P_{amp}$, where P_{amp} is the average power consumption of an op-amp.

The maximum graph size that a substrate can support is limited by the power budget. Given a power budget of P_{tot} and assuming $|V| \ll |E|$, then the the substrate can support roughly up to P_{tot}/P_{amp} number of edges. Assuming a 1V supply voltage and an average current of 500μA for the op-amp, which is typical

[4] Note that if an edge is not present in the graph, the corresponding op-amp can be power gated, thus consuming no power.

for the 32nm technology node, P_{amp} evaluates to 500μW. Given a power budget of $P_{tot} = 5$W, which is typical for an embedded system, we can accommodate up to 10^4 active edges on the substrate. Increasing the power budget to $P_{tot} = 150$W, a typical value for high-end servers, the number of active edges that the substrate can support boosts to 3×10^5. This estimation shows that the substrate is indeed promising in solving large-scale max-flow problems. Although in this case power consumption of the substrate is comparable to CPUs, the energy efficiency is drastically higher because our substrate is around $150X$ to $1500X$ faster than the CPUs.

6. Related Work

There has been a large body of work seeking to break the limit of the traditional von Neumann architecture and to achieve higher performance and efficiency. Silicon implementation of artificial neural network is one of the most well-known efforts in this field. Recently, a digital neuromorphic chip with one million spiking neurons was built with a power consumption of less than 70mW [24]. Analog implementation of the neuron network is also studied to achieve speedup and power reduction for specific applications [1]. Novel devices such as memristors are utilized to construct unconventional computing substrates. Memristor-based substrates are demonstrated to solve maze [26], conduct matrix multiplication [17], as well as general approximate computation [20]. A generalized concept of universal memcomputing machine is also proposed as an alternative to the von Neumann architecture [8].

In an effort to achieve reconfigurability in analog circuit, field-programmable analog arrays (FPAAs) integrate dedicated analog components with floating-gate transistors and programmable resistor/capacitor networks, enabling reconfigurable computing in the analog domain [2]. More relevant to our work, Vichik and Borrelli design an analog circuit to solve linear and quadratic programming problem, where unknown variables are directly modeled by node voltages and each constraint is specified by a dedicated circuit component [30]. The node voltages representing unknown variables are driven towards their optimal value subjected to the circuit constraints. However, their approach is problem specific, i.e., different instances of the analog circuit need to be built for solving different linear or quadratic programs. The structure of circuit components in our paper is inspired from their work, but tailored to exploit the specific properties of max-flow problems.

Although an ASIC or FPGA implementation of max-flow accelerator can improve performance and efficiency on a pure software execution, we note that in practice the demonstrated performance benefit is typically a moderate 3-$5X$ speedup over CPUs using an FPGA-based accelerator [18]. In fact, it has been shown that the max-flow problem is inherently hard to parallelize [12], implying that ASIC or FPGA based hardware acceleration can only achieve limited performance improvements. This motivates us to investigate more disruptive approaches in this research.

7. Conclusions and Future Work

In this paper, we propose a reconfigurable analog substrate that enables fast computation of max-flow on directed graphs. Our approach maps graph topology to on/off states of the crossbar switches, and obtain the solution of max-flow problem using steady-state voltages in the circuit. Performance evaluation demonstrates promising results in speed and energy consumption compared to digital computers. Future directions include investigating the asymptotic time complexity behavior by analyzing the transient response of the circuit. We also plan to study techniques to handle even larger graph using graph decomposition [28].

8. Acknowledgements

This work was supported in part by NSF CAREER Award #1453378. The authors thank Steve Dai and Prof. Eilyan Bitar for fruitful discussions in the early phase of this research project.

References

[1] R. S. Amant, A. Yazdanbakhsh, J. Park, B. Thwaites, H. Esmaeilzadeh, A. Hassibi, L. Ceze, and D. Burger. General-Purpose Code Acceleration with Limited-Precision Analog Computation. *Int'l Symp. on Computer Architecture (ISCA)*, Jun 2014.

[2] A. Basu, S. Brink, C. Schlottmann, S. Ramakrishnan, C. Petre, S. Koziol, F. Baskaya, C. M. Twigg, and P. Hasler. A Floating-Gate-Based Field-Programmable Analog Array. *IEEE Journal of Solid-State Circuits (JSSC)*, 45(9):1781–1794, 2010.

[3] D. Biolek, M. Di Ventra, and Y. V. Pershin. Reliable SPICE Simulations of Memristors, Memcapacitors and Meminductors. *Radioengineering*, 22(4), 2013.

[4] Y. Boykov and V. Kolmogorov. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)*, 26(9):1124–1137, 2004.

[5] D. Chakrabarti, Y. Zhan, and C. Faloutsos. R-MAT: A Recursive Model for Graph Mining. *Int'l Conf. on Data Mining (ICDM)*, Dec 2004.

[6] E. S. Chung, P. A. Milder, J. C. Hoe, and K. Mai. Single-Chip Heterogeneous Computing: does the Future Include Custom Logic, FPGAs, and GPGPUs? *Int'l Symp. on Microarchitecture (MICRO)*, Dec 2010.

[7] J. Cong and Y. Ding. FlowMap: An Optimal Technology Mapping Algorithm For Delay Optimization in Lookup-Table Based FPGA Designs. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 13(1):1–12, 1994.

[8] M. Di Ventra and Y. V. Pershin. The Parallel Approach. *Nature Physics*, 9(4):200–202, 2013.

[9] A.-T. Do, Z.-H. Kong, K.-S. Yeo, and J. Y. S. Low. Design and Sensitivity Analysis of a New Current-Mode Sense Amplifier for Low-Power SRAM. *IEEE Trans. on Very Large-Scale Integration Systems (TVLSI)*, 19(2):196–204, 2011.

[10] H. Esmaeilzadeh, E. Blem, R. St Amant, K. Sankaralingam, and D. Burger. Dark Silicon and the End of Multicore Scaling. *Int'l Symp. on Computer Architecture (ISCA)*, Jun 2011.

[11] G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee. Self-Organization and Identification of Web Communities. *IEEE Computer*, 35(3):66–70, 2002.

[12] L. M. Goldschlager, R. A. Shaw, and J. Staples. The Maximum Flow Problem is Log Space Complete for P. *Theoretical Computer Science*, 21(1):105–111, 1982.

[13] F. Halim, R. H. Yap, and Y. Wu. A MapReduce-Based Maximum-Flow Algorithm for Large Small-World Network Graphs. *Int'l Conf. on Distributed Computing Systems (ICDCS)*, Jun 2011.

[14] J. Han and M. Orshansky. Approximate Computing: An Emerging Paradigm for Energy-Efficient Design. *IEEE European Test Symposium (ETS)*, 2013.

[15] A. Hastings. The Art of Analog Layout. 2006.

[16] Y. Ho, G. M. Huang, and P. Li. Nonvolatile Memristor Memory: Device Characteristics and Design Implications. *Int'l Conf. on ComputerAided Design (ICCAD)*, Nov 2009.

[17] M. Hu, H. Li, Q. Wu, and G. S. Rose. Hardware Realization of BSB Recall Function Using Memristor Crossbar Arrays. *Design Automation Conf. (DAC)*, Jun 2012.

[18] D. Kobori and T. Maruyama. An Acceleration of a Graph Cut Segmentation with FPGA. *Int'l Conf. on Field Programmable Logic and Applications (FPL)*, Sep 2012.

[19] M. N. Leuenberger and D. Loss. Quantum Computing in Molecular Magnets. *Nature*, 410(6830):789–793, 2001.

[20] B. Li, Y. Shan, M. Hu, Y. Wang, Y. Chen, and H. Yang. Memristor-Based Approximated Computation. *Int'l Symp. on Low-Power Electronics and Design (ISLPED)*, Jul 2013.

[21] B. Liu, M. Hu, H. Li, Z.-H. Mao, Y. Chen, T. Huang, and W. Zhang. Digital-Assisted Noise-Eliminating Training for Memristor Crossbar-Based Analog Neuromorphic Computing Engine. *Design Automation Conf. (DAC)*, Jun 2013.

[22] G. Liu, Y. Tao, M. Tan, and Z. Zhang. CASA: Correlation-Aware Speculative Adders. *Int'l Symp. on Low-Power Electronics and Design (ISLPED)*, Jul 2014.

[23] C. Mead. Neuromorphic Electronic Systems. *Proceedings of the IEEE*, 78(10):1629–1636, 1990.

[24] P. A. Merolla et al. A Million Spiking-Neuron Integrated Circuit with A Scalable Communication Network and Interface. *Science*, 345(6197):668–673, 2014.

[25] L. W. Nagel and D. O. Pederson. SPICE: Simulation Program with Integrated Circuit Emphasis. 1973.

[26] Y. V. Pershin and M. Di Ventra. Solving Mazes with Memristors: A Massively Parallel Approach. *Physical Review E*, 84(4):046703, 2011.

[27] A. Schrijver. On the History of the Transportation and Maximum Flow Problems. *Mathematical Programming*, 91(3):437–445, 2002.

[28] P. Strandmark and F. Kahl. Parallel and Distributed Graph Cuts by Dual Decomposition. *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun 2010.

[29] T. Szymanski. Max-Flow Min-Cost Routing in a Future-Internet with Improved QoS Guarantees. *IEEE Trans. on Communications*, 61(4):1485–1497, 2013.

[30] S. Vichik and F. Borrelli. Solving Linear and Quadratic Programs with an Analog Circuit. *Computers and Chemical Engineering*, 70(5):160–171, 2014.