

引用格式:周游,李杰,贺光辉. 面向高精度随机计算单元的加法运算电路与 MAX 运算电路设计[J]. 微电子学与计算机, 2021,38(7):1-6. [ZHOU Y, LI J, HE G H. Design of adder and MAX unit for high accuracy stochastic computing [J]. Microelectronics & Computer, 2021,38(7):1-6.]

面向高精度随机计算单元的加法运算 电路与 MAX 运算电路设计

周 游¹, 李 杰², 贺光辉¹

(1 上海交通大学 微纳电子学系, 上海 200240;

2 上海航天测控通信研究所, 上海 200092)

摘 要: 随机计算(Stochastic Computing, SC)是一种以概率编码数值的新型技术,可以有效减小电路功耗和面积,但其所固有的随机特性会导致计算精度损失,从而限制了其应用范围. 为提高随机计算的准确性,提出了基于误差补偿原理的高精度随机计算加法运算单元,设计了一种计算随机序列最大值的 SC-MAX 单元. 仿真结果显示:与同类设计相比,两种计算单元的计算精度明显提高,其中同等 SC 序列长度下加法单元的相对误差降低了 80% 以上,SC-MAX 计算单元的相对误差减小了 90%,且 LUT 与 FF 开销减小了 10% 以上. 将两种计算单元运用于神经网络,在 MNIST 及 CIFAR10 数据集上的推理结果显示:精度损失跟浮点数相比分别可低至 0.04% 和 0.86%.

关键词: 随机计算;高精度;加法单元;SC-MAX 单元

中图分类号: TP929.1

文献标识码: A

文章编号: 1000-7180(2021)07-0001-06

Design of adder and MAX unit for high accuracy stochastic computing

ZHOU You¹, LI Jie², HE Guanghui¹

(1 Department of Micro Nano Electronics, Shanghai Jiao Tong University, Shanghai 200240, China;

2 Shanghai Aerospace Measurement and Control Communication Institute, Shanghai 200092, China)

Abstract: Stochastic Computing(SC) is a new technology that encodes numerical value by probability, which can effectively reduce the power consumption and area of the circuit. However, the loss of calculation accuracy caused by its inherent random characteristics limits its application. In order to improve the accuracy of Stochastic Computing, a high-precision Stochastic Computing addition unit based on error compensation principle is proposed, and a SC-MAX element is designed to calculate the maximum value of SC sequence. The simulation results show that, compared with the similar design, the calculation accuracy of the two calculation units is significantly improved. The relative error of the addition unit is reduced by more than 80% under the same SC length, and the relative error of the SC-MAX calculation unit is reduced by 90% compared with the similar design, and the LUT and FF overhead is significantly reduced by more than 10%. The two kinds of calculation units are applied to neural network. The reasoning results on MNIST and CIFAR10 data sets show that the precision loss is less than 0.04% and 0.86% respectively compared to float point results.

Key words: stochastic computing; high precision; SC adder unit; SC-MAX unit

收稿日期: 2020-10-28; 修回日期: 2020-11-15

基金项目: 国家重点研发计划项目(2020YFB2205500); 上海航天联合基金(USCAST2019-28); 国家自然科学基金项目(61674103, 62074097)

1 引言

随机计算是从数据编码角度提出的一种新型计算架构,它利用由 0 和 1 构成的随机比特序列中 1 的概率来编码数值,能用简单的门逻辑实现复杂的运算^[1]. 相比传统的二进制计算方法,随机计算运算电路结构简单,且具有低功耗和高容错的天然优势,故可以运用于多种应用.

但是,由于随机计算固有的随机特性,使得计算结果与二进制结果相比,会产生一定的精度损失,使其应用受到限制. 为了解决该问题,本文对随机计算单元进行设计. 最初的随机计算加法运算采用传统多路选择器实现^[2],但当输入个数增加时,该方法缩放系数增大,会引入较大误差;文献[3]提出了一种近似并行计数器加法单元(Approximate Parallel-Counter, APC),该设计基于与或门交替的近似单元与并行加法树结构,但该设计对随机序列的长度要求较高;对于计算最大值的随机计算单元,文献[4]提出一种基于序列分割的 Max Pooling Block 运算电路,该电路需要多个计数器实现,硬件开销较大且精度依赖于分段大小. 针对这些问题,本文提出两种高精度随机计算单元:利用随机计算序列的高容错率特性,提出了一种基于误差抵消原理的高精度随机计算加法运算单元,通过平衡运算中产生的正负误差以减小加法运算总误差;同时,本文提出一种计算随机序列最大值的 SC-MAX 单元,该单元基于反馈电路来实现最大值的输出,较大程度提高了计算精度,其相对误差较文献[3]中设计提高了 90%,且显著节省了硬件开销. 最后,为了验证本设计的有效性,将其运用于神经网络,在 MNIST 和 CIFAR10 数据集上进行推理运算,结果显示其分类准确度可达到 99.11%和 81.24%.

2 随机计算原理

SC 序列由 1 和 0 组成,与二进制数不同,序列中 1 的概率即该 SC 序列所表示的数值,但是序列长度和 1 的位置是不固定的,例如{0,1,0,0}, {1,0,0,0}, {1,0,1,0,0,0,0,0}根据定义都表示数 0.25. 随机计算主要有两个优点:一是可以用简单的电路实现复杂的运算,例如图 1(b)中为实现 SC 序列 A 与 B 的乘法,仅需逐比特通过与门,得到 Z 序列,其概率值即为乘积;二是随机计算具有高容错性,在长度为 m 的序列中单比特的翻转仅带来 $1/m$ 的误差.

随机计算系统由三个部分构成:SC 序列生成单

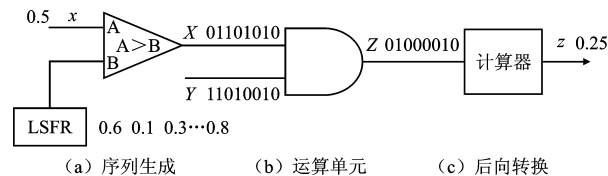


图 1 随机计算系统基本组成

元、SC 运算单元和 SC 后向转化单元,如图 1 所示. 其中 SC 序列生成单元实现二进制数转化为 SC 序列的功能,一般包含一个用于产生随机数的线性移位寄存器(Linear Feedback Shift Register, LFSR)和一个比较器;SC 后向转化单元用来将 SC 序列还原成二进制数,可以由计数器完成;SC 运算单元即随机计算乘法、加法等运算单元,例如图 1-b 中的乘法单元.

随机计算的局限性主要有两点:一是 SC 运算单元的精度不高;二是 SC 序列的长度随着所表示的数值精度的提升呈指数性增长^[5]. 因此,目前随机计算亟需设计在序列长度较短的情况下仍能保持高精度的随机计算单元.

3 高精度随机计算加法运算单元设计及优化

3.1 电路设计

现阶段的研究广泛使用或门、多路选择器 MUX 或并行全加器(Parallel Counter, PC)来实现随机计算中的加法运算. 如图 2 所示,其中或门结构开销最小,但多个 SC 序列累加会逐渐累积误差,降低了准确率;MUX 结构是一种缩放加法器,其精度相比或门有一定提升,但是随着输入个数增加,缩放因子增大,导致结果被压缩为一个较小的值,从而引入一定误差;PC 是精度较高的随机计算加法器,但开销和时延也较大. 文献[3]提出的 APC 基于与或门交替的近似单元与并行加法树结构,但该设计仅在随机序列长度较大的情况下才能保持较高精度.

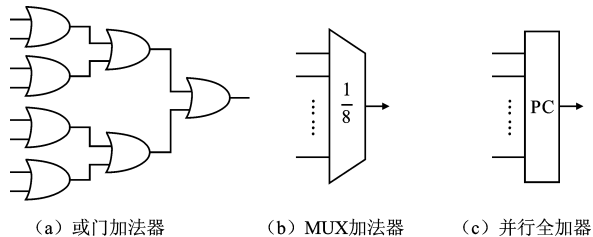


图 2 常用随机计算加法单元

基于上述随机计算加法器的优缺点,本文提出一种基于多路选择器的加法树结构(Mux Parallel Counter, MPC),该 MPC 加法器执行的功能是求得

输入之和的 1/2. 如图 3 所示,该加法器由两个部分构成,即并联 MUX 部分和并行加法器部分,第一部分采用多个并联的 MUX 结构,选通信号是值为 0.5 的随机序列,每一个 MUX 的结果是两输入之和的 1/2,所有 MUX 的结果输出到 PC;PC 部分为二进制加法器,无精度损失. 如图 3 所示,本设计中 PC 加法器部分的输入个数为总输入个数一半,相较于直接使用并行全加器可以节省面积和功耗.

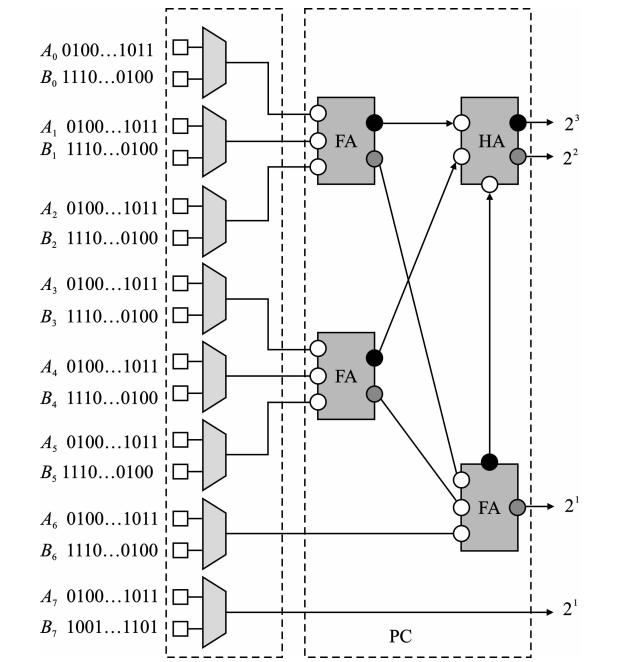


图 3 MPC 加法单元结构

3.2 原理与误差分析

对于 MPC 的原理和误差分析如图 4 所示,分为两种情况:

(1)当输入对应位置的比特相同时,若两个输入比特同为 1,则 MUX 输出为 1,求得两数之和的 1/2,故没有误差. 对于两个输入同为 0 时情况类似,不引入误差;

(2)当输入对应位置的比特不同时,这时 MUX 根据选通信号决定输出值,由于理论的输出值应该为两数之和的 1/2 即 0.5,因此 MUX 若输出 1 会引入 +0.5 的误差,输出 0 会引入 -0.5 的误差,由于选通信号值为 0.5,MUX 输出 1 和 0 的概率相同,故一个 MUX 产生的 +0.5 或 -0.5 的误差就可以内部进行相互补偿. 同时,SC 序列固有的随机性使得多个并行 MUX 之间的误差也能进行相互补偿,两次误差补偿将绝对误差控制在较小的范围,从而解决误差累积的问题,以此提高准确率.

为了进一步优化 SC 加法运算单元,提高其计

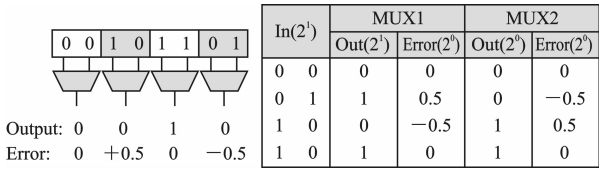
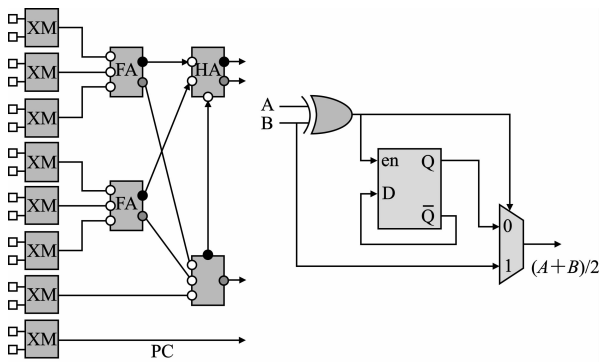


图 4 MPC 误差分析

算准确度,基于上文中所提的误差抵消的特性,对 MPC 结构进一步优化,提出了 XMPC. 如图 5 所示,该设计将上文中的 MUX 用如图 5(b)中的结构 XM 代替,该结构的功能是每当输入对应位置的比特不同时,记录当前输入比特,以保证下一次两个输入比特不同时,输出与上一次相异,从而保证 XM 的输出为交替的 0 和 1. 与 MPC 利用 0.5 为选通信号的随机误差补偿机制不同的是,该设计交替输出 0 和 1 能强制使输出产生的误差相互抵消,从而进一步提高了计算精度.



(a) XMPC 结果框图

(b) XM 结构图

图 5 XMPC 加法单元结构

3.3 性能分析

为了衡量上述 SC 加法运算单元的性能,本文对其进行了仿真验证. 采用 10 000 个范围在 [0,1] 之间的随机数作为输入数据,以二进制精确值为基准,用相对误差 (Relative Error, RE) 来衡量 SC 加法运算单元的误差. 如图 6 所示,RE 随输入个数增大而减小,仿真结果表明,在 SC 序列长度为 1 024 比特下,MPC 与 XMPC 加法单元精度远高于 APC^[3],XMPC 精度略高于 MPC. 同时,MPC 与 XMPC 加法器能在短随机序列下达到较高的精度,例如在 SC 序列长度为 64 比特下,精度高于采用序列长度为 1 024 比特的 APC^[3].

为了衡量该加法器的硬件开销,对 SC 加法器进行了硬件仿真,在 FPGA 上的综合结果如图 7 所示,其中 MPC 与 APC 加法器由于节省了一半的全并行加法器输入,其 LUT 与寄存器开销明显低于

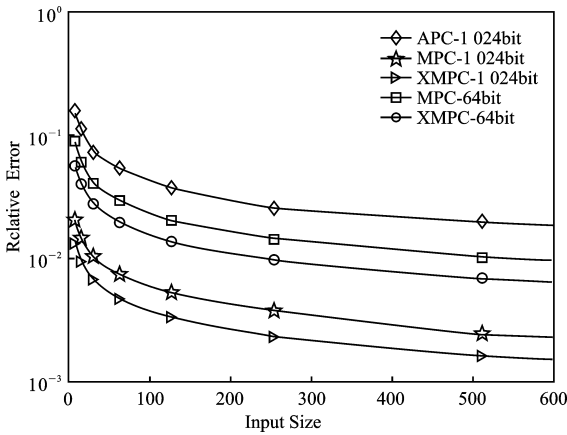


图6 APC, MPC, XMPC 性能比较

全并行加法器,且 MPC 的开销仅略高于 APC 加法器.由此可得,MPC 加法器在保持较高精度情况下不增加过多的额外资源开销.

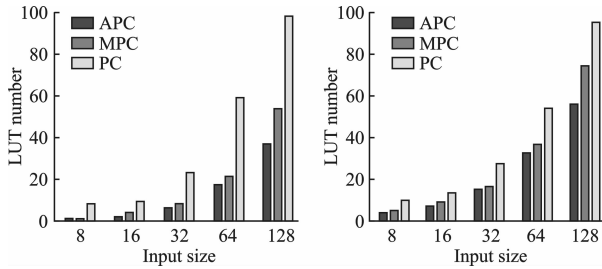


图7 APC, MPC, PC 的硬件资源开销

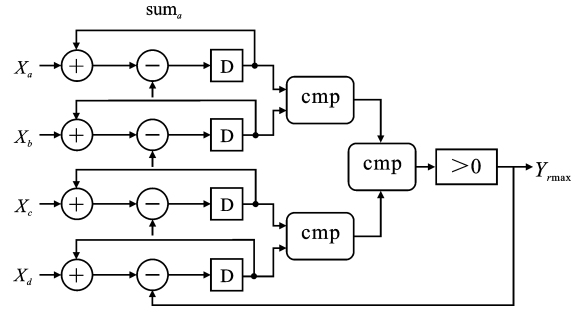
4 高精度随机计算 MAX 运算单元设计

4.1 电路设计

最大值计算单元在神经网络中是一种常用的池化方法,在二进制域中可以直接采用比较器获得多个数的最大值.然而在 SC 运算中,只有通过计算整个 SC 序列中的 1 的总数,才能在 n 个 SC 序列中找到最大值,从而带来了较长的延迟和较大的能耗.为了解决这个问题,文献[4]提出一种基于序列分割的电路,该电路将整个 SC 序列分段,当前序列段的输出由上一序列段的计数结果决定.但是该设计需要多个计数器,占用较大的硬件开销,计算精度较低且依赖于分段规则,限制了该电路的应用.

为了解决这个问题,本文提出了一种最大值计算电路 SC-MAX.如图 8 所示,以 $n=4$ 为例,图中输入为 4 个序列,该输入序列既可以是 SC 序列也可以是整数序列,输出序列为 SC 序列.图中每个序列都分别经过一个累加器和减法器,计算输入序列 X 与输出序列 Y 的差值.由于输出序列 Y 只包含 0 和 1,故减法器仅执行减 1 或 0 的功能,图中 D 单元

表示寄存器,用于存储当前差值,反馈到累加器中与下一个输入进行累加. CMP 为比较器,级联后输出四个差值中的最大值,将该最大值与 0 比较,若大于 0 则输出 1,反之则输出 0.

图8 SC-MAX 电路图,以 $n=4$ 为例

4.2 原理与理论证明

假设输入序列 X_a 趋近于均匀分布,根据切比雪夫定律可以得出如下推论:整数序列 X_a 任意前缀子序列 X_{a_k} 的均值等于 X_a 的均值,即

$$\sum_{i=1}^k X_{a_i} = P(X_a) \cdot k \quad (1)$$

同理对于 X_b, X_c, X_d 序列有:

$$\sum_{i=1}^k X_{b_i} = P(X_b) \cdot k \quad (2)$$

$$\sum_{i=1}^k X_{c_i} = P(X_c) \cdot k \quad (3)$$

$$\sum_{i=1}^k X_{d_i} = P(X_d) \cdot k \quad (4)$$

对输出 Y 序列做出相同假设,故有:

$$\sum_{i=1}^k Y_i = P(Y) \cdot k \quad (5)$$

图中第 k 次输入时,对于 X_a, X_b, X_c, X_d 序列各个累加器中 sum 的值分别为:

$$\text{sum}_{a_k} = \sum_{i=1}^k (X_{a_i} - Y_i) \quad (6)$$

$$\text{sum}_{b_k} = \sum_{i=1}^k (X_{b_i} - Y_i) \quad (7)$$

$$\text{sum}_{c_k} = \sum_{i=1}^k (X_{c_i} - Y_i) \quad (8)$$

$$\text{sum}_{d_k} = \sum_{i=1}^k (X_{d_i} - Y_i) \quad (9)$$

因此 CMP 单元后的最大值为:

$$\begin{aligned} \text{sum}_k &= \max(\text{sum}_{a_k}, \text{sum}_{b_k}, \text{sum}_{c_k}, \text{sum}_{d_k}) \\ &= \max\left(\sum_{i=1}^k (X_{a_i}), \dots, \sum_{i=1}^k (X_{d_i})\right) - \sum_{i=1}^k (Y_i) \\ &= \max(P(X_a), P(X_b), P(X_c), P(X_d)) - P(Y) \end{aligned} \quad (10)$$

因此,下一比特的输出为:

$$Y_{k+1} = \text{sign}[\max (P(X_a), \cdots, P(X_d)) - P(Y)] \tag{11}$$

下面根据不同的 X 输入分析 Y 的结果:

记 $P(X_{\max}) = \max (P(X_a), P(X_b), P(X_c), P(X_d))$,

(1) 当 $P(X_{\max}) \geq 1$ 时,由于 Y 是 SC 序列,所以 $P(Y) \leq 1$ 恒成立,因此 $Y_{k+1} = \text{sign}[P(X_{\max}) - P(Y)] = 1$,即输出序列 Y 为全 1,即 $P(Y) = 1$;

(2) 当 $P(X_{\max}) \leq 0$ 时,由于 Y 是随机序列,所以 $P(Y) \geq 0$ 恒成立,因此 $Y_{k+1} = \text{sign}[P(X_{\max}) - P(Y)] = 0$,即输出序列 Y 为全 0,即 $P(Y) = 0$;

(3) 当 $0 < P(X_{\max}) < 1$ 时,假设在第 k 次输入时, $P(Y) < P(X_{\max})$, 则输出为 $Y_{k+1} = 1$. 因此 $P(Y)$ 会增大,变得与 $P(X)$ 接近. 反之,如果第 k 次输入时, $P(Y) > P(X_{\max})$, 则输出为 $Y_{k+1} = 0$. 因此 $P(Y)$ 会减小,与 $P(X)$ 接近. 因此,随着序列的不断输入, $P(Y)$ 会不断波动,最终逼近 $P(X_{\max})$. 若在稳定状态, $P(Y)$ 只会比 $P(X_{\max})$ 大或小 1 bit, 因此误差在 $1/n$.

综上, $P(Y)$ 可表示为:

$$\begin{cases} 0, & P(X) \leq 0 \\ P(X_{\max}), & 0 < P(X_{\max}) < 1 \\ 1, & P(X) \geq 1 \end{cases} \tag{12}$$

由于该证明已经假设前缀序列 X_i 的均值等于序列 X 的均值,故对于四个 SC 序列中值最大的序列,它的前缀序列的值也是四个前缀序列中最大的.

4.3 性能分析

为了衡量本文提出的 SC-MAX 单元性能,对该设计进行仿真验证,同时与文献[4]中的 Max Pooling Block 运算单元进行对比,仿真结果显示,本文提出的 SC-MAX 单元能在较短的序列下达到较高的精度,能达到与二进制结果几乎无误差的结果,其相对误差如表 1 所示. 如图 9 所示,图中显示了对于输入个数分别为 4, 9 和 16 的情况下的相对误差,可以看出,该设计在 SC 序列长度为 64 比特的情况下精度仍远高于文献[4]中使用 512 比特序列长度的精度.

表 1 SC-MAX 单元相对误差/%

输入个数	SC 序列长度				
	64	128	256	384	512
4	1.23	0.839	0.517	0.369	0.290
9	1.43	1.250	0.840	0.587	0.439
16	1.86	1.310	0.939	0.677	0.498

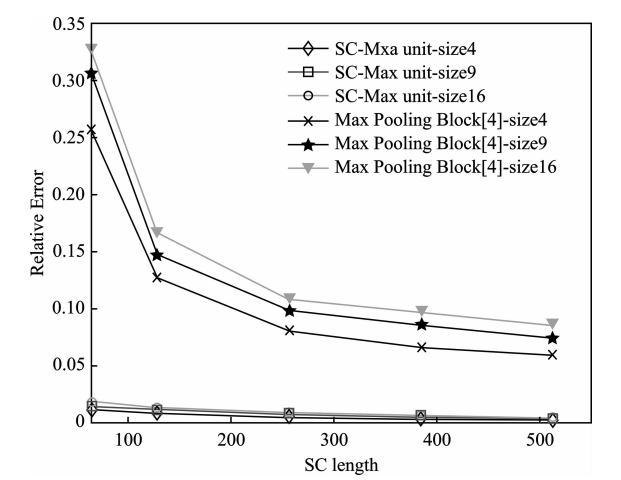


图 9 SC-MAX 设计性能

为衡量该设计的硬件开销,对文献[4]中的结构和本文的结构进行了硬件仿真. 在 FPGA 上的综合结果如表 2 所示,可以看出,SC-MAX 结构在保持较高精度情况下能节省一定资源开销.

表 2 SC-MAX, 文献[4]各硬件资源消耗

		输入个数	4	9	16
SC-MAX	LUT 数		52	112	194
	FF 数		31	61	103
文献[4]	LUT 数		58	132	221
	FF 数		47	78	118

5 在神经网络中的应用

5.1 实验环境

为评估本文提出的随机计算单元的性能,将该随机计算单元运用于神经网络中,用 XMPC 替代神经网络中的加法运算单元,用 SC-MAX 实现池化计算单元,网络中乘法计算采用文献[6]中的结构实现. 选用 MNIST, Cifar10 数据集进行推理测试,其中 MNIST 采用 Lenet-5 网络结构,包含 3 层卷积层,2 层池化层和 2 层全连接层;Cifar10 网络结构与文献[7]一致,即 2 卷积-1 池化-2 卷积-1 池化-2 全连接的结构. 其中,卷积核的大小为 3×3 ,步长为 1;池化采用的是最大值池化,滑动窗的大小为 2×2 ,步长为 2. 卷积层分别有 32, 32, 64, 64 个滤波器,全连接层分别有 512 和 10 个神经元.

5.2 实验结果与分析

实验仿真结果如表 3 所示,通过表格可以看出,运用本文提出的 SC 运算单元,在采用 64 比特或 128 比特这样较短 SC 序列长度的情况下即可达到较好的精度,同类设计中一般需要采用远高于本设

计的 SC 序列长度. 在 SC 序列长度为 128 比特下, 与浮点数结果相比, 对于 MNIST 数据集, 本设计仅有 0.04% 的精度损失; 对于 Cifar10 数据集, 本设计在 256 比特下仅有 0.86% 的误差.

表 3 神经网络推理实验结果

网络	文献	序列长度	准确率/%	FLOAT-SC/%
MNIST	文献[4]	256	98.26	0.22
MNIST	本文	64	99.04	0.11
MNST	本文	128	99.11	0.04
Cifar10	文献[7]	1 024	82.83	1.03
Cifar10	本文	64	80.40	1.70
Cifar10	本文	256	81.24	0.86

6 结束语

本文为了提高随机计算运算单元的准确性, 针对以下两种随机计算单元进行了设计: (1) 提出了基于误差补偿原理的高精度随机计算加法单元; (2) 设计了一种计算随机序列最大值的 SC-MAX 单元. 仿真结果显示, 与同类设计相比, 两种计算单元的计算精度明显提高, 其中加法单元的相对误差降低了 80%, SC-MAX 计算单元的相对误差降低了 90%, 且 LUT 与 FF 开销减小. 同时, 将提出的两种计算单元运用于神经网络, 在 MNIST 及 CIFAR10 数据集上进行推理, 可以达到较好的准确率. 因此, 提出的两种结构为提高 SC 的精度提供了新思路.

参考文献:

[1] ALAGHI A, HAYES J P. Survey of stochastic computing [J]. ACM Transactions on Embedded Computing Systems, 2013, 12(2S): 92. DOI:10.1145/2465787.2465794.

[2] AKHTAR R, KHANDAY F A. Stochastic computing: systems, applications, challenges and solutions[C]//2018 3rd International Conference on Communication and Electronics Systems (ICCES). Coimbatore, India: IEEE, 2018: 722-727. DOI: 10.1109/CESYS.2018.8723960.

[3] KIM K, LEE J, CHOI K. Approximate de-randomizer for stochastic circuits[C]//2015 International SoC De-

sign Conference (ISOCC). Gyeongju, Korea (South): IEEE, 2015: 123-124. DOI: 10.1109/ISOCC.2015.7401667.

[4] REN A, LI Z, DING C W, et al. SC-DCNN: highly-scalable deep convolutional neural network using stochastic computing[C]//Proceedings of the Twenty-Second International Conference on Architectural Support for Programming Languages and Operating Systems. Virtual, MI, USA: Association for Computing Machinery, 2017: 405-418. DOI: 10.1145/3037697.3037746.

[5] 熊海翥, 贺光辉. 基于随机计算的大规模 MIMO 检测算法研究与硬件实现[J]. 微电子学与计算机, 2020, 37(7): 36-41. DOI: 10.19304/j.chki.issn1000-7180.2020.07.008.

XIONG H F, HE G H. Design and implementation of stochastic computing based massive MIMO detector [J]. Microelectronics & Computer, 2020, 37(7): 36-41. DOI: 10.19304/j.chki.issn1000-7180.2020.07.008.

[6] XIONG H F, ABU BAKAR M, HE G H. Hardware implementation of an improved stochastic computing based deep neural network using short sequence length [J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2020, 67(11): 2667-2671. DOI: 10.1109/TCSII.2020.2969691.

[7] YU J, KIM K, LEE J, et al. Accurate and efficient stochastic computing hardware for convolutional neural networks[C]//2017 IEEE International Conference on Computer Design (ICCD). Boston, MA: IEEE, 2017: 105-112. DOI: 10.1109/ICCD.2017.24.

作者简介:

周 游 女, (1996-), 硕士研究生. 研究方向为基于随机计算的神经网络电路设计与优化.
E-mail: zhou-you14hnu@163.com.

李 杰 男, (1989-), 博士, 工程师. 研究方向为计算机视觉、人工智能及其嵌入式优化.

贺光辉 男, (1980-), 博士, 副教授. 研究方向为无线通信系统算法、神经网络硬件加速器设计.