基于超几何分解的随机运算系统分析方法

马承光*[①] 仲顺安[①] David Lilja[②] 屈若媛[①]
[①](北京理工大学信息与电子学院 北京 100081)
[②](明尼苏达大学电子与计算机工程系 明尼阿波利斯 55455)

摘 要：基于 Bernoulli 分布的方差及期望传递方程一直是随机运算系统的数学基础，针对这种传统分析方法在实际应用中的不准确性和片面性，该文提出一种全新的数学方法：超几何分解(hypergeometic decomposition)，用来解决在更复杂情况下期望与方差在随机运算系统中的传播规律。基于超几何分解，提出 4 组更加精确的期望及方差传递方程，在数学上证明了随机运算体系更加广泛的适用性，并且通过随机运算系统在图像处理中的应用，提出了基于方差的系统评价方法，相比于传统按位仿真方法，基于方差的系统分析方法具有耗时短、准确和全面的优点。新的方差传递方程首次将随机信号源的类型引入性能分析，证明了具有特定码流长度的随机序列可以使系统性能达到最优。

关键词：随机运算系统；方差传递方程；超几何分解；系统评价
中图分类号：TP302.7 **文献标识码**： A **文章编号**：1009-5896(2013)02-0355-06
DOI: 10.3724/SP.J.1146.2012.00711

Analysis Method of Stochastic Computing System Based on Hypergeometric Decomposition

Ma Cheng-guang[①] Zhong Shun-an[①] David Lilja[②] Qu Ruo-yuan[①]
[①](*School of Information and Electronics*, *Beijing Institute of Technology*, *Beijing* 100081, *China*)
[②](*Department of Electrical and Computer Engineering*, *University of Minnesota*, *Minneapolis*, *MN* 55455, *USA*)

Abstract: As mathematical fundamental of stochastic computing system, transfer function of variance and expected value based on Bernoulli distribution is not accurate and general in system analysis. A novel mathematic method, hypergeometric decomposition is proposed to solve this problem; it offers a general way to calculate transfer function of expected value and variance under more complicated circumstance. There are four groups of transfer function proposed here, which proves the effectiveness of stochastic computing system in a more general way; also they offer a better way to evaluate stochastic system. Compared with traditional bit-level simulation, evaluation method based on variance is time saving, accurate and comprehensive. New variance transfer function includes type of input random stream into performance analysis for the first time, which proves that specific length of stochastic sequence can maximize system performance.

Key words: Stochastic computing system; Variance transfer function; Hypergeometric decomposition; System evaluation

1 引言

随着 CMOS 工艺按照摩尔定律的不断进步，纳米器件引起的种种问题渐渐凸显，微电子器件抵御外部和内部噪声的能力越来越低。与此同时，功耗密度不断升高、系统复杂度逐步增大、测试难度加大。面对这些问题，传统计算体系法已经无法提供根本性的解决方案，而基于其它工艺的设计方法由完全抛弃了 CMOS 工艺完善的设计理论体系，无法在短时间内得到应用。在这种背景中，随机运算(Stochastic Computing)体系作为一种全新的设计方法，在完全兼容现有 CMOS 工艺的前提下，成为一种在计算体系上解决纳米器件问题的有效方案。

不同于传统系统的编码方式，随机运算系统利用逻辑 1 在一个随机序列中的出现概率来表达数值大小，如随机序列“101000”代表 1/3。这从根本上改变了系统设计理论和方法，很多传统算法中的逻辑器件在随机逻辑体系中有了全新的运算行为，如一个与门在随机运算中可以完成乘法运算，一个选择器可以完成加法运算。这种电路形式上的简化同时意味着低功耗、高速、设计简单等优点[1]。另外，

2012-06-08 收到，2012-11-14 改回
*通信作者：马承光 web_cat@live.cn

由于随机运算的每一位的权值相同，因此个别位的翻转对整个码流的影响很小，因此随机运算架构具有很强的抗干扰能力。

目前，随机运算架构已经在神经网络[2]、数模转换器[3]、机器控制[4]、数字图像处理[5]、误差矫正译码[6,7]等领域得到广泛应用，这些研究大部分集中在如何将随机运算的优势发挥出来。在随机运算体系中，仍然存在一些十分突出的问题，由于随机序列的不准确性，如何高效分析和评估随机运算体系是制约随机运算发展的一个重要问题。传统的系统分析方法主要采用基于 Bernoulli 分布的概率论基本原理，由于在随机运算中的随机序列并不是严格意义上的 Bernoulli 分布，因此这种分析方法并不准确。按位仿真是另一种系统评价方法，这种方法较为准确，但是仿真时间较长，而且分析方法片面，并不具有设计指导作用。

本文通过采用一种全新的数学方法，在更加普适的条件下，提出了广义期望及方差的传递方程，从理论上证明了在不同随机信号发生源的情况下随机运算理论的完美性，同时证明了方差作为系统评价指标的可行性。通过利用随机运算体系实现数字图像处理算法，在实际应用中检验方差作为系统评价指标的有效性，并讨论随机信号源的类型对系统性能的影响。

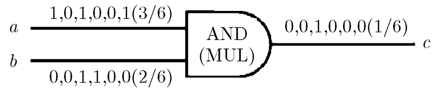
2 传统期望及方差传递理论

在随机运算体系中，系统性能是由两个指标决定的：期望和方差。如果一个系统运算的期望值与设计相符，同时方差又足够小，那么这个算法就是有效的。相反，则会产生较大的计算误差，导致算法失效。因此理解期望和方差的传播规律对设计和优化随机运算系统来说是十分重要的。

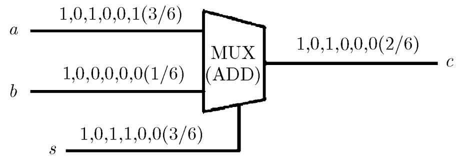
在经典的随机运算算法分析中，认为所有随机序列均为 Bernoulli 分布[8]，通过经典概率论相关原理，可以得到期望及方差的传递方程。下面通过对两个基本随机运算单元的分析，阐述经典的期望及方差传递规律。

图 1 为两个基本随机运算逻辑，与门可以用来完成乘法运算，或门可以用来完成带权值的乘法运算。在经典概率论理论中，假设两个逻辑门的所有输入均为 Bernoulli 分布，那么输出序列的期望服从式(1)和式(2)的概率学方程，方差服从式(3)的概率学方程[9]。其中 E_O代表逻辑门的输出期望值，n 代表随机序列的长度。

$$E_A(a,b) = E_a \cdot E_b \tag{1}$$



(a)随机乘法器



(b)随机权值加法器

图 1 两个基本随机运算单元

$$E_M(a,b,s) = E_a \cdot E_s + E_b \cdot (1 - E_s) \tag{2}$$

$$\sigma^2 = \frac{E_O \cdot (1 - E_O)}{n} \tag{3}$$

这组方程是随机运算的基本公式，其证明了与门和或门可以完成乘法和带权值加法运算。尽管这组方程可以描述期望及方差的传播规律，但由于随机运算系统中的随机序列并不都是 Bernoulli 分布(如 LFSR 产生的序列)，因此这组公式不够准确，它只在一个相对狭小的空间中，证明了随机运算的有效性及可靠性。

这种不准确性是无法将方差分析作为系统分析方法的主要原因，也是目前大多数随机运算系统仍采用按位仿真[10]进行系统验证的主要原因。按位仿真是一种实时仿真方法，虽然仿真结果准确，但仿真时间过长，特别是在随机序列长度过长的情况下。同时，它不具备系统级设计对评价方法的要求，无法指出算法的问题所在及改进方向。

3 超几何分解与广义期望及方差传递方程

针对随机运算具体实现电路的复杂性，为了能够在更广的条件下证明随机运算的有效性及可靠性，需要提高期望及方差传递方程的准确性和适用性。图 2 是一个更具有一般性的随机运算基本问题，假设输入序列的分布形式是未知的，已知条件只有输入序列的期望和方差，那么如何求解与门输出序列的期望和方差。此问题不具有 Bernoulli 分布的假设条件，问题的一般性增加了。

这里将采用一种全新的数学方法来解决此问题：超几何分解，这种数学方法的基本概念如图 3 所示。其基本思想是将未知的分布形式分解成一系

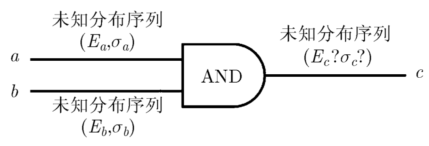


图 2 扩展的随机运算基本问题

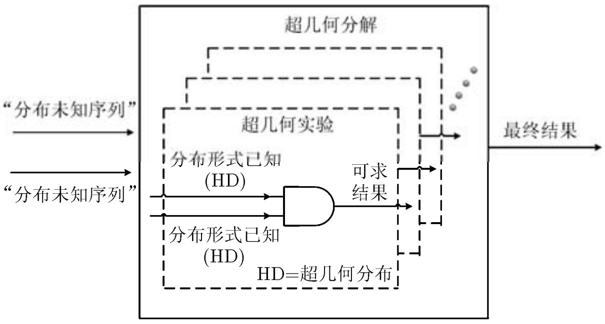


图 3 超几何分解原理示意图

列超几何分布，进而利用超几何分布的基本原理解决被拆分的单独问题，最终将结果组合成原问题的结果，其基本思想类似于 Taylor 分解。

首先对原来的问题进行分解。如式(4)所示，这里 b_1 代表输入序列 a 中逻辑 1 的个数，b_2 代表输入序列 b 中逻辑 1 的个数，E_{ij} 代表如果序列 a 中含 i 个逻辑 1，序列 b 中含 j 个逻辑 1，在此种情况下输出序列 c 中逻辑 1 个数的期望值。在此，成功地将一个未知分布形式的概率学问题转化成为多个更为“清晰”的子问题。

$$E(x)=\sum_{i=1,j=1}^{n,n}p(b_1=i,b_2=j)E_{ij}(x) \tag{4}$$

由于输入随机序列之间是相互独立的，因此有

$$E(x)=\sum_{i=1,j=1}^{n,n}P(b_1=i)p(b_2=j)E_{ij}(x) \tag{5}$$

求解 E_{ij} 的过程在本质上是求解一个超几何分布问题的过程，我们可以将此问题进行转化以帮助理解，转化方法如图 4 所示，在此我们将序列 b 中逻辑 1 的位置进行重新排序，即假设序列 b 中前 j 个数字为逻辑 1 其它未知为逻辑 0，从本质上来说，序列 b 中逻辑 1 的位置在求解问题中仅起到一个标签的作用。

因此求解 E_{ij} 就转变成为以下问题：在一个存在 i 个逻辑 1 的序列中(图中黑框所示序列)取出 j 个数字位，其中含有逻辑 1 的数字位的总和的期望是多少。这是一个典型的超几何分布问题，这也是我们将这种数学方法命名为超几何分解的原因，一个具有未知分布形式的问题被分解为一系列超几何分布

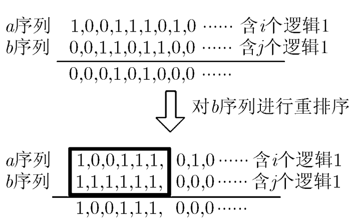


图 4 将原问题转化为超几何分布问题

形式的子问题。根据超几何分布基本理论有

$$E_{ij}(x)=ij/n \tag{6}$$

将式(6)代入式(5)，并作进一步处理，E_1 代表序列 a 中逻辑 1 总数的期望值，E_2 代表序列 b 中逻辑 1 总数的期望值。

$$E(x)=\sum_{i=1,j=1}^{n,n}p(b_1=i)p(b_2=j)\frac{ij}{n} \tag{7}$$

$$=\frac{\left(\sum_{i=1}^{n}p(b_1=i)\cdot i\right)\left(\sum_{j=1}^{n}p(b_2=j)\cdot j\right)}{n} \tag{8}$$

$$=\frac{E_1\cdot E_2}{n} \tag{9}$$

式(9)在进行单位化后与式(1)具有相同的形式，这就说明随机运算的基本公式与输入序列的具体分布形式没有必然联系，同时证明了随机运算在更加广义的条件下是可靠的，在数学本质上并不会出现偏差。采用同样的方法，可以求解方差的传播公式，

$$\begin{aligned}E(x^2)&=\sum_{i=1,j=1}^{n,n}p(b_1=i)p(b_2=j)E_{ij}(x^2)\\&=\sum_{i=1,j=1}^{n,n}p(b_1=i)p(b_2=j)(\sigma_{ij}(x)^2+E_{ij}(x)^2)\\&=\sum_{i=1,j=1}^{n,n}p(b_1=i)p(b_2=j)\\&\quad\cdot\left(\frac{ij(n-j)(n-i)}{n^2(n-1)}+\left(\frac{ij}{n}\right)^2\right)\\&=\frac{\left(\sum_{i=1}^{n}p(b_1=i)i(n-i)\right)\left(\sum_{i=1}^{n}p(b_2=j)j(n-j)\right)}{n^2(n-1)}\\&\quad+\frac{\left(\sum_{i=1}^{n}p(b_1=i)i^2\right)\left(\sum_{j=1}^{n}p(b_2=j)j^2\right)}{n^2}\end{aligned} \tag{10}$$

由于

$$\sum_{i=1}^{n}p(b_1=i)i=E_1 \tag{11}$$

$$\sum_{j=1}^{n}p(b_1=j)j=E_2 \tag{12}$$

$$\sum_{i=1}^{n}p(b_1=i)i^2=\sigma_1^2+E_1^2 \tag{13}$$

$$\sum_{j=1}^{n}p(b_2=j)j^2=\sigma_2^2+E_2^2 \tag{14}$$

其中 σ_1 代表序列 a 逻辑 1 总数的方差，σ_2 代表序列 b 逻辑 1 总数的方差，将式(11)-式(14)代入式(10)，

$$\begin{aligned}E(x^2)=&\frac{1}{n^2}\left(\frac{1}{n-1}(nE_1-\sigma_1^2-E_1^2)(nE_2-\sigma_2^2-E_2^2)\right.\\&\left.+\sigma_1^2E_2^2+\sigma_2^2E_1^2+\sigma_1^2\sigma_2^2+(E_1E_2)^2\right)\end{aligned} \tag{15}$$

因此

$$\sigma^2(x)=E(x^2)-E(x)^2 =\frac{1}{n^2}\left(\frac{1}{n-1}(nE_1-\sigma_1^2-E_1^2)(nE_2-\sigma_2^2-E_2^2) +\sigma_1^2E_2^2+\sigma_2^2E_1^2+\sigma_1^2\sigma_2^2\right) \tag{16}$$

式(16)与传统的方差传递式(3)有很大差别，如果定义式(16)中输入序列的期望和方差满足 Bernoulli 分布规律，即假设输入为 Bernoulli 序列，式(16)会简化为式(3)。这就说明传统方差传递方程仅是式(16)的一个特例，式(16)是更加普适，更加准确的方差传递方程。

在随机运算体系中，对于任何相互独立的输入信号序列，只要其方差与期望存在，就可以利用超几何分解的数学方法得到 4 种基本运算逻辑(AND, OR, XOR, MUX)的期望和方差的传播方程：

$$E_A(a,b)=E_a\cdot E_b \tag{17}$$

$$\sigma_A^2(a,b)=\frac{1}{n-1}(E_a-\sigma_a^2-E_a^2)(E_b-\sigma_b^2-E_b^2) +\sigma_a^2E_b^2+\sigma_b^2E_a^2+\sigma_a^2\sigma_b^2 \tag{18}$$

$$E_O(a,b)=E_a+E_b-E_a\cdot E_b \tag{19}$$

$$\sigma_O^2(a,b)=\frac{1}{n}(E_a'-\sigma_a^2-E_a'^2)(E_b'-\sigma_b^2-E_b'^2) +\sigma_a^2E_b'^2+\sigma_b^2E_a'^2+\sigma_a^2\sigma_b^2 \tag{20}$$

$$E_M(a,b,s)=E_a\cdot E_s+E_b\cdot E_s' \tag{21}$$

$$\sigma_M^2(a,b,s)=\sigma_A^2(a,s)+\sigma_A^2(b,s')-2E_aE_b\sigma_s^2 \tag{22}$$

$$E_X(a,b)=E_a\cdot E_b+E_a'\cdot E_b' \tag{23}$$

$$\sigma_X^2(a,b)=\sigma_A^2(a,b)+\sigma_A^2(a',b')+2\left(\frac{n}{n-1}(E_a-\sigma_a^2 -E_a^2)(E_b-\sigma_b^2-E_b^2)-E_aE_bE_a'E_b'\right) \tag{24}$$

其中 E_a' 代表 1-E_a，E_b'，E_s'，a', b', s' 具有相同的含义，注意式(17)-式(24)中的 E 和 σ 与式(6)-式(16)中具有不同含义，此处的期望与方差为单位化后的方差。

根据这组广义的期望与方差传播方程，不管在随机运算架构中输入的序列为何种分布形式，只要输入序列的期望与方差存在，就有以下结论：

(1)传统理论中假设所有输入序列为 Bernoulli 分布的条件是不必要的，随机逻辑的运算单元完全可以在更加复杂的条件下正确地完成运算。

(2)输出序列的期望值与输入序列的分布形式无关，其值仅由输入序列的方差和运算单元的类型决定。

(3)输出序列的方差是一个稳定的，可求解的变量，其值由输入序列的方差、期望和运算单元的类型决定。

4 利用广义方差传递方程分析和优化随机运算系统

通过这组广义期望及方差传递方程的推导，证明了方差完全可以作为评判性能的重要指标。只要系统结构确定，那么方差就可以在设计之初通过数学公式求解，相比于传统的按位仿真方法来说，使用方差可以更加准确地描述一个随机运算系统的数学原理。

在一个随机运算系统中，方差的产生、传播及缩放原理如图 5 所示。其中深色的部分是与方差变化有关的部分，包括方差产生、方差传播及方差缩放 3 个部分。方差的最初来源是随机序列生成器，目前较为常用的随机序列发生器包括两种：噪声二极管和 LFSR[11-13]。最初产生的方差在随机系统中传播，被不同类型的随机逻辑改变。由于在算法的开始阶段进行了数据标准化的操作，在算法的最终阶段方差会被缩放。

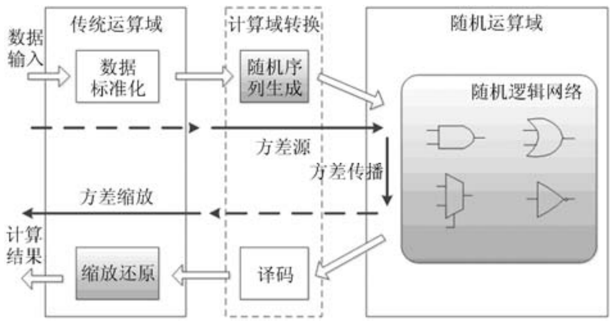


图 5 随机运算系统基本结构与方差的产生、传播及缩放

下面利用随机运算架构实现一个常见的数字图像处理算法：图像平均算法。此处略去具体设计过程，主要利用设计结果对比不同系统分析方法。图 6 是在不同随机码流长度下，使用随机运算架构的图像平均算法与传统算法的行为级仿真结果对比，图中数字代表随机码流的长度，TRA 代表传统算法。通过采用行为仿真的方法，可以直观地对设计结果进行对比，但很难准确全面地分析随机运算结构，同时无法提出有效的设计优化方案。

表 1 是通过采用方差传播原理对随机图像平均算法的分析。可以看出，基于方差的系统分析方法不仅正确反映了行为仿真结果，同时更加全面地分析了系统行为，而且耗时更短，可以作为设计优化的高效衡量标准。由于按位仿真求出的等效方差并不是所有像素点的平均方差，且其值并不稳定，因此结果较方差平均值有一定偏差。

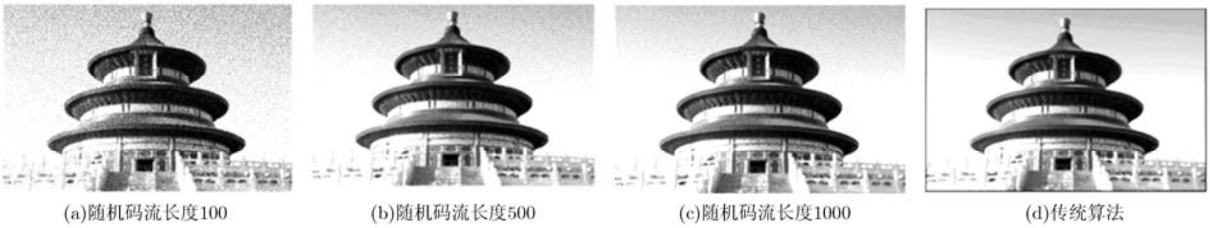


图 6 随机算法与传统算法按位仿真结果对比

表 1 系统分析方法对比

	基于方差传递方程(%)				基于按位仿真的等效方差(%)
	平均值	最小值	最大值	上限	
SC100	13.02	0.89	16.09	16.11	10.12
SC500	5.82	0.40	7.19	7.21	4.64
SC1000	4.12	0.28	5.09	5.10	3.09
时间消耗	<5 s				>1 h

广义方差传递方程不仅为分析随机运算系统提供了理论基础，同时首次在理论上将输入源的类型对系统性能的影响引入系统分析理论。在目前随机逻辑设计中，一般要求尽量保持输入随机序列的Bernoulli分布性质，但通过广义方差传递程的提出，证明了各种随机分布形式都能够驱动基本随机逻辑运算单元。因此，可以从随机信号源的类型入手，对系统进行理论分析与优化。

在利用 LFSR 作为随机序列发生器的系统中，随机序列长度 L 与寄存器个数 N 的选取一直是首先需要考虑的问题，大部分文献认为 L 需要小于等于 2^N-1，以保证序列的 Bernoulli 性质[9,11,14]。由于广义方差传递方程将输入源的类型引入性能分析，因此可以通过广义方差传递方程，进一步分析这个问题。图 7 是在 AND 门和 MUX 门中，不同的 L 与 N 的取值对最终方差的影响，当 L 与 2^N-1 相同时，两类门输出序列的方差最小。这证明了 Bernoulli 分布并不是最优的随机序列源，同时也为系统优化提出了新的方向。

通过采用特定长度的随机信号发生器，对上述数字平均算法进行优化设计，实验结果如图 8 所示。可以发现，相比于传统基于 Bernoulli 分布($L<<2^N-1$)的噪声二极管，采用 L 与 2^N-1 相等的 LFSR 信号源具有更好的系统性能。由于在图像平均算法中，主要用到乘法和加法运算，即主要随机逻辑单元为 AND 门和 MUX 门，因此实验结果符合图 7 有关随机序列类型与方差的分析。这不仅从侧面说明了广义方差传递方程理论的正确性，同时也从随机信号源方面入手，完成了随机运算系统的优化。

5 结束语

本文通过提出一种全新的数学方法：超几何分

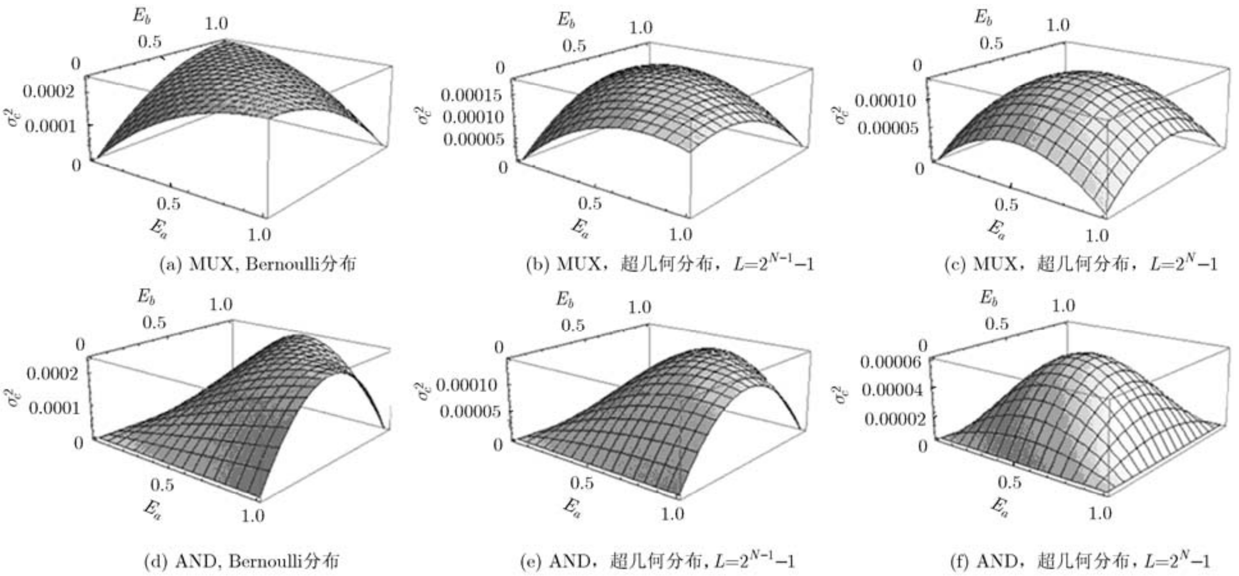


图 7 随机序列类型对输出方差的影响

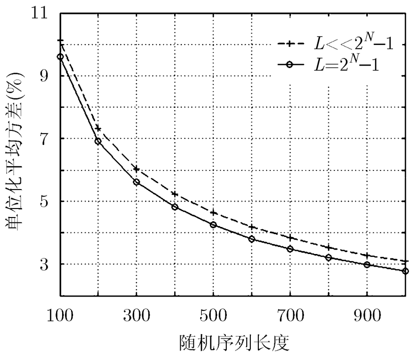


图 8 两种随机信号源对随机图像平均系统性能的影响

解，并将其应用在随机运算体系中，以解决随机运算误差分析和系统性能分析不准确的问题。证明并提出了一组适用于基本组合逻辑的广义期望及方差传递公式。这组方程不仅在更加普适的条件下证明了随机运算的有效性，同时为将方差作为随机运算系统的性能指标奠定了理论基础。

通过广义方差传递方程，可以更快、更准确、更全面地对随机运算系统进行评估，同时从随机信号源的类型入手，通过广义方差传递方程的理论指导，提出改进随机运算精度的办法，并通过随机图像处理的应用证明了系统分析及优化的有效性。未来的工作包括将基于方差的系统评估方法应用到更多的随机运算结构中，并考虑噪声、相关性等非理想因素对广义方差传递方程的影响。

参 考 文 献

[1] Brown B D and Card H C. Stochastic neural computation I: computational elements[J]. *IEEE Transactions on Computers*, 2001, 50(9): 891–905.

[2] Kondo Y and Sawada Y. Functional abilities of a stochastic logic neural network [J]. *IEEE Transactions on Neural Networks*, 1992, 3(3): 434–443.

[3] Ortega J, Janer C, Quero J, *et al*.. Analog to digital and digital to analog conversion based on stochastic logic [C]. Conference of the IEEE Industrial Electronics Society (IECON), Orlando, FL, USA, 1995: 995–999.

[4] Palem K. Energy aware computing through probabilistic switching: a study of limits[J]. *IEEE Transactions on Computers*, 2005, 54(9): 1123–1137.

[5] Li Peng and Lilja D J. Using stochastic computing to implement digital image processing algorithms[C]. 2011 IEEE 29th International Conference on Computer Design, Amherst, MA, USA, 2011: 154–161.

[6] Quang Trung-dong, Arzel M, and Jego C. Asynchronous stochastic decoder for spread spectrum digital watermarking[C]. 2011 IEEE 9th International New Circuits and Systems Conference (NEWCAS), Bordeiux, France, 2011: 21–24.

[7] Gaudet V C and Gross W J. Switching activity in stochastic decoders[C]. Proceedings of the 40th IEEE International Symposium on Multiple-Valued Logic (ISMVL), Barcelona, Spain, May 2010: 167–172.

[8] Gaines B R. Stochastic Computing Systems in Advances in Information Systems Science [M]. New York: Plenum Press, 1999, 2: 37–172.

[9] Qian Wei-kang, Li Xin, Riedel M D, *et al*.. An architecture for fault-tolerant computation with stochastic logic[J]. *IEEE Transactions on Computers*, 2011, 60(1): 93–105.

[10] Hori M and Ueda M. Fpga implementation of a blind source separation system based on stochastic computing [C]. 2008 IEEE Conference on Soft Computing in Industrial Applications (SMCia/08), Muroran, Japan, 2008: 25–27.

[11] Jeavons P, Cohen D A, and Shawe-Taylor J. Generating binary sequences for stochastic computing[J]. *IEEE Transactions on Information Theory*, 1994, 40(3): 716–720.

[12] Scholefield P H R. Shift register generating maximum-length sequences[J]. *Electronic Technology*, 1960, 37(10): 389–394.

[13] Qian Wei-kang, Riedel M D, Zhou Hong-chao, *et al*.. Transforming probabilities with combinational logic[J]. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2011, 30(9): 1279–1292.

[14] Toral S L, Quero J M, and Franquelo L G. SRC passivation controller implementation using stochastic computing[C]. The 2001 IEEE International Symposium on Circuits and Systems, Sydney, Australia, 2001: 723–726.

马承光： 男，1984 年生，博士生，研究方向为随机运算架构与 VLSI 实现.

仲顺安： 男，1957 年生，教授，博士生导师，研究方向为专用集成电路设计、硅基微波集成技术.

David Lilja： 男，教授，博士生导师，研究方向为计算机体系架构、高性能计算结构.

屈若媛： 女，1985 年生，博士生，研究方向为模拟集成电路设计.