

DOI:10.3969/j.issn.1003-5060.2024.05.012

基于 SoC 的非对称数字系统算法设计与实现

姜智, 肖昊

(合肥工业大学 微电子学院, 安徽 合肥 230601)

摘要:文章提出一种在片上系统(System on Chip, SoC)实现高吞吐率的有限状态熵编码(finite state entropy, FSE)算法。通过压缩率、速度、资源消耗、功耗 4 个方面对所提出的编码器和解码器与典型的硬件哈夫曼编码(Huffman coding, HC)进行性能比较,结果表明,所提出的硬件 FSE 编码器和解码器具有显著优势。硬件 FSE(hFSE)架构实现在 SoC 的处理系统和可编程逻辑块(programmable logic, PL)上,通过高级可扩展接口(Advanced eXtensible Interface 4, AXI4)总线连接 SoC 的处理系统和可编程逻辑块。算法测试显示, FSE 算法在非均匀数据分布和大数据量情况下,具有更好的压缩率。该文设计的编码器和解码器已在可编程逻辑块上实现,其中包括 1 个可配置的缓冲模块,将比特流作为单字节或双字节配置输出到 8 bit 位宽 4 096 深度或 16 bit 位宽 2 048 深度的块随机访问存储器(block random access memory, BRAM)中。所提出的 FSE 硬件架构为实时压缩应用提供了高吞吐率、低功耗和低资源消耗的硬件实现。

关键词:有限状态熵编码(FSE);哈夫曼编码(HC);片上系统(SoC);高吞吐率;块随机访问存储器(BRAM)

中图分类号:TN762;TN764

文献标志码:A

文章编号:1003-5060(2024)05-0655-06

Design and implementation of asymmetric numerical system algorithm based on SoC

JIANG Zhi, XIAO Hao

(School of Microelectronics, Hefei University of Technology, Hefei 230601, China)

Abstract: An algorithm for achieving high-throughput finite state entropy (FSE) in System on Chip (SoC) is proposed. The performance of the proposed FSE encoder and decoder is compared to that of typical hardware Huffman coding (HC) in terms of compression ratio, speed, resource consumption, and power consumption. The results show that the proposed hardware FSE (hFSE) encoder and decoder have significant advantages over HC. The hFSE architecture is implemented on the processing system and programmable logic (PL) of an SoC, connected via the Advanced eXtensible Interface 4 (AXI4) bus. Algorithm tests demonstrate that FSE algorithm has better compression ratios for non-uniform data distributions and large data volumes. The encoder and decoder, which includes a configurable buffer module that outputs bit streams as single or double bytes to $4\ 096 \times 8$ -bit or $2\ 048 \times 16$ -bit block random access memory (BRAM), have been implemented on PL. The proposed FSE hardware architecture provides a low-power-consumption, low-resource-consumption and high-throughput hardware implementation for real-time compression applications.

Key words: finite state entropy (FSE); Huffman coding (HC); System on Chip (SoC); high throughput; block random access memory (BRAM)

收稿日期:2023-02-22;修回日期:2023-03-24

基金项目:国家自然科学基金资助项目(61974039)

作者简介:姜智(1996—),男,安徽阜南人,合肥工业大学硕士生;

肖昊(1982—),男,安徽合肥人,博士,合肥工业大学教授,博士生导师,通信作者, E-mail: xiaohao@hfut.edu.cn.

0 引 言

数据传输的带宽对于数字计算系统至关重要,特别是在当前的通用计算系统中,数据带宽已经成为中央处理器(central processing unit, CPU)和外部输入输出设备通信的主要瓶颈。为了提高数据传输带宽,实时数据压缩和解压是最有效的方法,而熵编码技术则是实现实时数据压缩和解压的核心。

传统的熵编码技术,如哈夫曼编码(Huffman coding, HC)和算术编码(arithmetic coding, AC)在实时系统、汽车系统、视频会议、图像处理、音频和视频录制等领域得到广泛应用,这些领域都需要高效的数据压缩算法来提高带宽和降低存储成本。通常认为 HC 在速度方面具有优势而 AC 在压缩率方面具有优势。

虽然 HC 速度相对较快,但是在实时压缩领域 HC 的效率依然较低^[1],这是由于在哈夫曼算法中构建二叉树是极其昂贵的^[2],编码过程中,不仅要存储每个节点的深度,还要不断更新二叉树,这是造成 HC 效率低的主要原因,尽管可以通过贪心算法优化二叉搜索树^[3],或者通过随机模拟算法(stochastic simulation algorithm, SSA)优化哈夫曼树,哈夫曼算法依然难以满足实时要求高的应用场景。

通过硬件加速可以优化 HC 的效率问题,这方面研究较多。在解码和编码过程中,采用并行化的方法可以极大地提高吞吐率^[4],但同时也会消耗更多的硬件资源;在硬件设计中优化树的结构也可以提高数据吞吐量并降低资源消耗^[5],不过性能提升却相对有限。文献^[6]提出字典式编码(Lempel-Ziv-1977, LZ77)和 HC 的硬件设计,并设计了在块内存中构建和存储哈夫曼树的方法,这构成压缩文件存档(compressed file archive, GZIP)在现场可编程门阵列(field programmable gate array, FPGA)上实现的基础。然而,哈夫曼解码器的效率依然较低,这是由于哈夫曼解码器需要处理变长代码中的最大比特数^[7]造成的。

非对称数字系统(asymmetric numeral system, ANS)是一种用于数据压缩的无损熵编码算法,最初由文献^[8]提出,并被纳入多项技术的专利中。如谷歌的专利文件中有多个与 ANS 相关的专利,包括对 ANS 编码的优化和 ANS 的视频编码等。而微软则在其专利中使用 ANS 算法进

行语音和图像数据的压缩,并将 ANS 算法与其他压缩算法进行结合,提高数据压缩的效率。此外,ANS 也在许多开源项目中得到了广泛应用,如 Facebook 的 zstandard 压缩库和 Tukaani 的马尔可夫链(Lempel-Ziv-Markov chain algorithm, LZMA)压缩工具,这些工具都使用了 ANS 作为其压缩算法之一。ANS 算法还被用于实现哈希函数^[8]、随机数生成器^[8]和加密算法等领域^[9],具有高度的可扩展性和灵活性,可以适应不同的应用场景和需求。ANS 算法的主要思想是将数据转换为一系列数字,这些数字可以被编码为不同数量的比特位,以便实现高效的压缩和解压缩。研究者提出了 2 种 ANS 算法,分别是 rANS(range asymmetric numeral system)、tANS(table asymmetric numeral system)。文献^[10]提出硬件实现 tANS 的解码器,该硬件架构在吞吐率方面有优势,但是存在消耗存储资源过大的问题。研究者在 ANS 算法基础上,提出了有限状态熵(finite state entropy, FSE)算法,并已经在工程方面被广泛应用,但是该算法在硬件方面的研究较少。

本文提出一种基于片上系统(System on Chip, SoC)实现高吞吐率的有限状态熵编码算法,简称为 hFSE(hardware finite state entropy),并且与传统硬件哈夫曼算法的编码器、解码器进行性能对比,本文提出的编码器和解码器在压缩率、速度、资源消耗、功耗 4 个方面都比硬件哈夫曼编码器和解码器有显著优势。

1 算法设计

1.1 hFSE 算法架构

hFSE 算法数据处理流程可以分为数据预处理、符号编码、符号解码 3 个步骤。

数据预处理的主要目的是生成 6 个固定的查找表,即 1 个编码表、2 个符号转移表、3 个解码表进而为后续编码和解码使用。

符号编码的过程是利用编码表和符号转移表不断更新当前状态值,并从当前状态值取出相应位数构成二进制码流。

符号解码的过程则是利用解码表从二进制码流还原符号。

为了优化算法的硬件速度,将查找表大小设置为 2 次整数次方,这样乘法可以转换成左移,除法可以转换成右移,取模运算可以转换成逻辑与运算,进而极大地提高运算速度,考虑到在实际应

用过程中数据预处理只需要计算 1 次,而符号编码和解码需要计算多次,数据预处理过程通过 CPU 计算,符号编码和解码过程通过 FPGA 加速,即使用 SoC 的处理器系统 (processing system, PS) 端实现数据预处理,使用 SoC 的可编程逻辑块 (programmable logic, PL) 端实现符号编码和解码。

本文所提的 hFSE 非对称数字系统架构设计如图 1 所示,用 Vivado 建立模块设计 (block design, BD) 工程,按照从左到右、从上到下的顺序,分

别为 Zynq-7000 型号芯片的片上处理芯片模块、高级可扩展接口 (Advanced eXtensible Interface, AXI) 互连模块、AXI 的随机访问存储器 BRAM (block random access memory) 控制器模块、BRAM 生成器模块、自定义打包的 FSE 编码器模块、BRAM 生成器模块、自定义打包的 FSE 解码器模块和 BRAM 生成器模块。

其中上侧 4 个模块为 PS 端模块,下侧 4 个模块为 PL 端模块,通过高级可扩展接口 AXI4 总线实现 SoC 的 PS 端和 PL 端的通信。

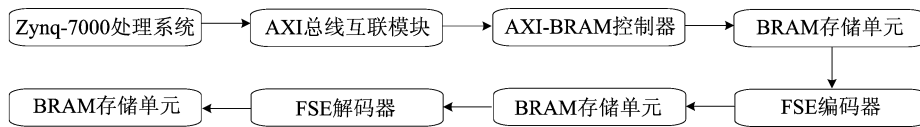


图 1 hFSE 的非对称数字系统架构

1.2 FSE 算法性能测试

FSE 算法相对于传统哈夫曼的优势是压缩率有所提升^[11],ANS 算法的特例即非对称二进制算法 (asymmetric binary system, ABS) 已在数学上被证明收敛于源熵^[12],即收敛于理想压缩率,因此本文着重对 FSE 算法的压缩率进行测试。

输入符号长度对压缩率的影响如图 2 所示,对称分布情况下,改变输入符号长度,可以得到 FSE 算法的实际压缩率情况。

由图 2 可知,当输入符号的量级越大时,FSE 算法的压缩率越接近于理想压缩率,即压缩性能越好,而实时压缩领域的符号数量量级都在 10^5 以上,这使 FSE 算法在工程中能达到几乎香农极限的压缩率。

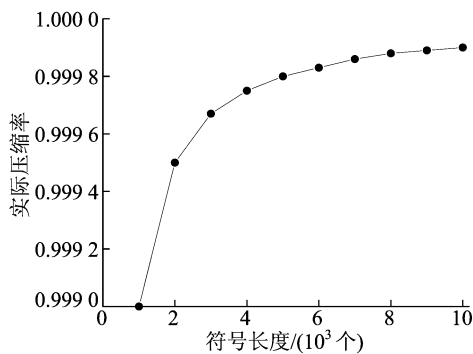


图 2 输入符号长度对压缩率的影响

考虑输入符号类型对压缩率的影响。输入符号近似均匀分布 (uniform)、2 次方分布 (power-2)、10 次方分布 (power-10) 和 50 次方分布

(power-50) 的情况下,算法压缩率的变化见表 1 所列。

表 1 非对称数字系统算法仿真结果

分布类型	符号长度	实际压缩率	理想压缩率
uniform	10	0.666 7	1.000 0
power-2	10	0.706 2	1.083 8
power-10	10	0.833 3	1.358 1
power-50	10	1.089 3	2.089 7
uniform	100	0.952 4	1.000 0
power-2	100	1.031 4	1.083 8
power-10	100	1.283 0	1.358 1
power-50	100	1.924 6	2.089 7
uniform	500	0.992 1	1.000 0
power-2	500	1.072 8	1.083 8
power-10	500	1.343 7	1.358 1
power-50	500	2.055 6	2.089 7
uniform	1 000	0.995 0	1.000 0
power-2	1 000	1.078 5	1.083 8
power-10	1 000	1.350 7	1.358 1
power-50	1 000	2.074 0	2.089 7
uniform	10 000	0.999 5	1.000 0
power-2	10 000	1.083 2	1.083 8
power-10	10 000	1.357 5	1.358 1
power-50	10 000	2.088 5	2.089 7

理想压缩率通过香农极限公式计算,实际压缩率使用压缩后的比特数除以压缩前的比特数计算。

理想压缩率受到符号分布的影响,而符号长度不会影响理想压缩率。实际压缩率受到 2 种因素的共同影响。

流,输出为解码符号的二进制表示。

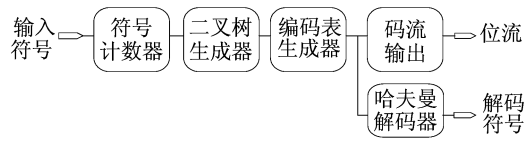


图 4 哈夫曼编解码器系统框图

2.3 测试平台

硬件测试平台为:SoC 为米联客 mz7xb,核心板为 xc7z020clg400-2。

软件测试平台为:vivado2019.2,python3.10。

仿真环境为:centOS7,vcs2018.09,verdi2018.09。

2.4 性能对比分析

考虑到 hFSE 是新的熵编码架构,本文从压缩率、速度、资源消耗、功耗 4 个方面对传统硬件哈夫曼架构与本文提出的架构进行比较。编码器输入均为 400 bit 相同的近似均匀分布符号,输出为二进制码流;解码器的输入均为各自的编码二进制码流,输出均为原来的 400 bit 符号。

在压缩率方面,Huffman 编码器的压缩率为 1.069 5,hFSE 编码器的压缩率为 1.176 5,优于传统哈夫曼编码器。在速度方面,hFSE 的硬件电路最大频率为 200.00 MHz,而 Huffman 的硬件电路最大频率为 128.21 MHz,具体而言,Huffman 编码器的吞吐率为 9 441.95 MiB/s,解码器吞吐率为 4 384.36 MiB/s,而 hFSE 编码器的吞吐率为 96 689.36 MiB/s,解码器吞吐率为 82 185.96 MiB/s,因此,从吞吐量方面考虑,hFSE 编码器的吞吐率提高 10.2 倍,hFSE 解码器性能提升 18 倍。

2 种硬件架构各模块在片查找表(slice look-up table,slice LUT)、片寄存器(slice register)、多路选择器(multiplexer,MUX)消耗资源情况见表 2 所列,Huffman 编码器所消耗的 slice LUT 数量是 hFSE 编码器的 51.11 倍,而 Huffman 解码器所消耗的 slice LUT 数量是 hFSE 解码器的 12.89 倍。

表 2 hFSE 和 Huffman 消耗数量对比

模块	片查找表	片寄存器	多路选择器
Huffman 编码器	10 784	924	85
Huffman 解码器	1 366	468	0
hFSE 编码器	211	61	16
hFSE 解码器	106	61	16

在功耗方面的对比结果如图 5 所示,当电路频率为 128 MHz 时,Huffman 架构消耗的总功耗是 hFSE 的 2.76 倍,图 5 柱状图上侧代表静态功耗,下侧代表动态功耗。

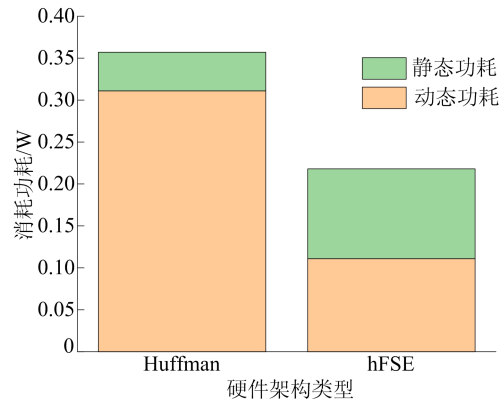


图 5 hFSE 和 Huffman 消耗功耗对比

3 结 论

本文测试 FSE 算法在不同输入情况下的压缩率变化情况,结果表明,FSE 算法的输入数据非对称性越高,输入数据量越大,压缩率越高。

本文实现了基于 SoC 的 FSE 算法的硬件架构设计,并将设计的 hFSE 编码器、hFSE 解码器和传统的基于状态机设计的 Huffman 编码器、Huffman 解码器进行全方位对比。结果表明,FSE 算法的硬件架构在吞吐量和资源消耗方面相对传统 Huffman 算法的硬件架构具有巨大优势。在未来的视频压缩器和图像压缩器中,hFSE 可以完全取代原先的 Huffman 熵编码器和解码器。

[参 考 文 献]

[1] DUDA J,TAHBOUB K,GADGIL N J,et al.The use of asymmetric numeral systems as an accurate replacement for huffman coding [C]//2015 Picture Coding Symposium (PCS). [S.l.:s.n.],2015:65-69.

[2] BAEZA-YATES R A.Expected behaviour of B+-trees under random insertions [J]. Acta Informatica,1989,26:439-471.

[3] KOSARAGU S R,PRZYTYCKA T M,BORGSTROM R.On an optimal split tree problem[C]//The 6th International Workshop on Algorithms and Data Structures. [S.l.:s.n.],2002:157-168.

[4] ASPAR Z,YUSOF Z M,SULEIMAN I.Parallel huffman decoder with an optimized look up table option on FPGA[C]//2000 TENCON Proceedings Intelligent Systems and Technologies for the New Millennium. [S.l.:s.n.],2000:73-76.

(下转第 677 页)

[参 考 文 献]

- [1] 汪馨. *Psp* 系统在丙二酸盐克罗诺杆菌干燥胁迫中的作用探讨[D]. 合肥:合肥工业大学,2021.
- [2] 战捷. 调控因子 *RpoS* 和 *SspA* 在阪崎克罗诺杆菌抗环境胁迫中的作用[D]. 无锡:江南大学,2021.
- [3] 凌娜. 中国即食蔬菜食品中克罗诺杆菌的遗传多样性与酸胁迫响应机制研究[D]. 广州:华南理工大学,2019.
- [4] FG A, MPA B, MIGUEL P, et al. The role of the general stress response regulator *RpoS* in *Cronobacter sakazakii* biofilm formation-science direct[J]. Food Research International, 2020, 136(1):0963-9969.
- [5] 梁运改, 桂荫, 王顺, 等. 鼠伤寒沙门菌 *rpoS* 基因缺失菌株的构建及 *rpoS* 因子在环境胁迫下的作用[J]. 食品科学, 2018, 39(6):141-147.
- [6] ADNAN M, SOUSA A M, MACHADO I, et al. Role of *boLA* and *rpoS* genes in biofilm formation and adherence pattern by *Escherichia coli* K-12 MG1655 on polypropylene, stainlesssteel, and silicone surfaces[J]. Acta Microbiologica Et Immunologica Hungarica, 2017, 64(2):1-11.
- [7] BAZIRE A, DIAB F, JEBBAR M, et al. Influence of high salinity on biofilm formation and benzoate assimilation by *Pseudomonas aeruginosa*[J]. Journal of Industrial Microbiology & Biotechnology, 2007, 34:5-8.
- [8] ALVAREZ-ORDONEZ A, BEGLEY M, HILL C, et al. Polymorphisms in *rpoS* and stress tolerance heterogeneity in natural isolates of *Cronobacter sakazakii*[J]. Applied & Environmental Microbiology, 2012, 78(11):3975-3984.
- [9] AMEELAH M, DEWANTI-HARIYADI R, NURJANAH S, et al. Expression of *rpoS*, *ompA* and *hfq* genes of *Cronobacter sakazakii* strain Yrt2a during stress and viable but nonculturable state[J]. Food Science and Biotechnology, 2018, 27:915-920.
- [10] XU Z, LIU Z, SOTEYOME T, et al. Impact of *pmrA* on *Cronobacter sakazakii* planktonic and biofilm cells: a comprehensive transcriptomic study[J]. Food Microbiology, 2021, 98:103785-103789.
- [11] GAO Z J, GE C T, BAKER R C, et al. Evaluation of potential for butyl and heptyl para-hydroxybenzoate enhancement of thermal inactivation of *Cronobacter sakazakii* during rehydration of powdered infant formula and nonfat dry milk[J]. Journal of Food Protection, 2022, 85(8):1133-1141.
- (责任编辑 张 镅)
-
- (上接第 659 页)
- [5] BEAK S, VAN HIEU B, PARK G, et al. A new binary tree algorithm implementation with huffman decoder on FPGA [C]//2010 Digest of Technical Papers International Conference on Consumer Electronics (ICCE). [S. l. : s. n.], 2010: 437-438.
- [6] RIGLER S, BISHOP W, KENNINGS A. FPGA-based lossless data compression using huffman and LZ77 algorithms [C]//2007 Canadian Conference on Electrical and Computer Engineering. [S. l. : s. n.], 2007:1235-1238.
- [7] BONNY T, HENKEL J. Efficient code density through look-up table compression[C]//Automation & Test in Europe Conference & Exhibition 2007 Design. [S. l. : s. n.], 2007:1-6.
- [8] DUDA J. Asymmetric numeral systems[EB/OL]. [2023-02-05]. <https://www.semanticscholar.org/paper/Asymmetric-numeral-systems-Duda/3b003beb9d8186c51b4912933b1ff75eef0a3c6c>.
- [9] DUDA J, NIEMIEC M. Lightweight compression with encryption based on asymmetric numeral systems [EB/OL]. [2023-02-05]. <https://dl.acm.org/doi/10.34768/amcs-2023-0004>.
- [10] NAJMABADI S M, TUNGAL H S, TRAN T H, et al. Hardware-based architecture for asymmetric numeral systems entropy decoder[C]//2017 Conference on Design and Architectures for Signal and Image Processing (DA-SIP). [S. l. : s. n.], 2017:1-6.
- [11] DUDA J. Asymmetric numeral systems: entropy coding combining speed of huffman coding with compression rate of arithmetic coding [EB/OL]. [2023-02-05]. <https://www.semanticscholar.org/paper/Asymmetric-numeral-systems%3A-entropy-coding-speed-of-Duda/bbcd5711d2f02e5fe07316669264334d193e88>.
- [12] YOKOO H. On the stationary distribution of Asymmetric Binary Systems[C]//2016 IEEE International Symposium on Information Theory (ISIT). [S. l.]: IEEE, 2016: 11-15.
- (责任编辑 张 镅)