Optimized and Cost Considering Huffman Code For Biological Data Transmission

Abstract

Keywords:

1. Introduction

2. Background

2.1. Issues on Biological Data Transmission

The size of biological data including DNA sequences increase with an ever expanding rate and will be bigger and bigger in the future. These Biological data are stored in biology database, the exponential growth of these database become a big problem to all biological data processing methods. Different operation will be applied to these data such as, searching [],e-mail attachment [], alignment [], and transmission on distributed computing []. Interestingly, biological data compression can play a key role in all biological data processing.

A recent deluge of interest in the development of new tools for biological data processing, these all algorithms needs an efficient methods for data compression. The main objective of data compression methods is minimizing the number of bits in the data representation. In [Marty C. Brandon] authors propose a new general data structure and data encoding approach for the efficient storage of genomic data. This method encode only the differences between a genome sequence and a reference sequence, the method use different encoding scheme from fixed codes such as Golomb and Elias codes, to variables codes, such as Huffman codes. Other methods based on same idea to encode only the difference between reference sequence and the target one, Authors in [Scott Christley1] uses Huffman code for encoding difference between sequence to sent it as an email attachment, but these methods suffer that they must sent the reference sequence for at least one time for each

species.
Wang and Zhang (2011) proposed a new scheme for referential compression of genomes based on the chromosome level. The Algorithm aim to search for longest common subsequence between matching parts and the differences encoded using Huffman coding.
All previous studies focus only on the differences and the relation between continuation of the sequence, and without improvement of the encoding scheme.

2.2. Huffman Code

2.3. Power Efficient on Data Transmission

In the recent years, application of battery-powered portable devices, e.g. laptop computers and mobile phones has increased rapidly. Power dissipation has become a primary concern for digital community because it affects the performance,reliability, and the cost of computation in both portable and non-portable devices. CMOS technologies were developed in order to reduce the power consumption in devices. Digital CMOS circuits have three major sources of power dissipation and are summarised in the following equation [1]:

$$\begin{aligned} P_{avg} &= P_{dynamic} + P_{short-circuit} + P_{leakage} \\ &= \alpha_{0\to1} \cdot C_L \cdot V_{dd}^2 \cdot f_{clk} + I_{sc} \cdot V_{dd} + I_{leakage} \cdot V_{dd} \end{aligned} \tag{1}$$

2.4. Unequal Bit Considering On data Transmission

3. Approach

3.1. Proposed Scheme

3.2. Power Efficient Huffman code

3.3. Optimisation of the Codes

4. Results And Discussion

5. Conclusion

[1] N. H. E. Weste, K. Eshraghian, Principles of CMOS VLSI design: a systems perspective, Addison-Wesley, 1988.