An Adaptive Implementation of Multi-core K-Nearest Neighbour Ensemble Classifier Using Dynamic Partial Reconfiguration

Hana M. Hussain[1], Dr. Khaled Benkrid[1], Chan Hong[1], Dr. Huseyin Seker[2]

The Edinburgh University, UK
{ h.hussain,k.benkrid. C.Hong}@ed.ac.uk
[2]De Montfort University,UK
hseker@dmu.ac.uk

Introduction

K-nearest Neighbour (K-NN) classifier is a supervised machine learning method used to assign a class label to a query of un-known class. When datasets are large and highly dimensional, General Purpose Processors (GPPs) execute the K-NN classifier slowly. Microarray, a high throughput biotechnology applies K-NN to predict the class label of un-known gene expression sample e.g. identify if the sample is of cancerous or non-cancerous tissue, diagnose or predict disease prognosis etc. GPP limitations in terms of execution time and power consumption calls for applying high performance computing (HPC) methods. As such, a high performance adaptive FPGA-based implementation of K-NN classification using Dynamic Partial Reconfiguration (DPR) is proposed in this paper.

Background on K-NN Classification

Given X_{NM}= {X_{N1}, X_{N2}, XN3, XN4,....X_{NM}}is a training set of M dimensions, and N patterns, each pattern X has a known class label L, there are C possible classes e.g., C=2 identifies cancerous vs. non-cancerous tissues, or diseased vs. Non-diseased tissue. Y is a query of M dimensions. Manhattan distance is used to compute the distance between Y and all patterns in X to determine the closest number of neighbourhood (K) patterns (KNNs) to the query and assign the query to the most frequently encountered L. The Manhattan Distance is:

$$D(X,Y) = \sum_{i=1}^{M} |Y_i - X_{Ni}|$$

Microarray classification problems are highly dimensional imposing high computational demands, thus are candidate for FPGA acceleration.

Systolic Array Architecture

The hardware implementation was captured in Verilog HDL, fully parameterized and pipelined. Parameters are: number of neighbourhoods (K), number of dimensions (M), number of training vectors (N), each dimensions has a wordlength of B, number of class labels (C). The design is based on two types of systolic arrays to perform the distance computation and finding the KNNs, lastly the class-label is determined.
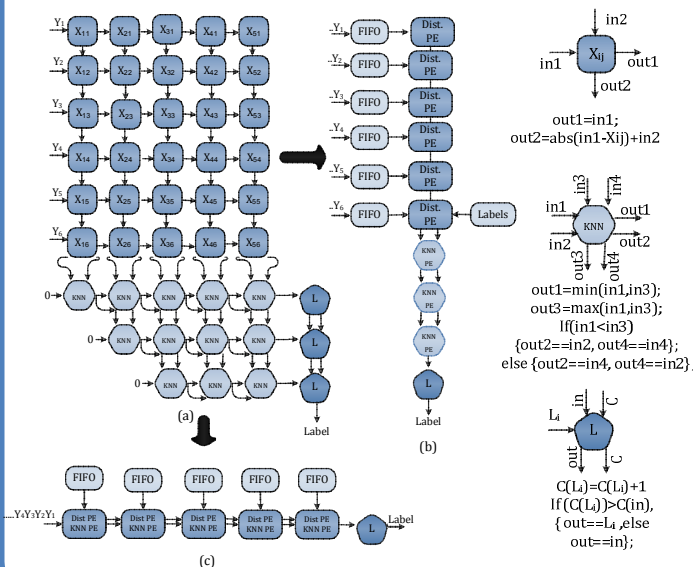

FPGA vs. GPP implementations Results

❖ FPGA implementation achieved~ 60x speed-up over (GPP).
❖ FPGA board used is ML403 having Xilinx XC4VFX12. The single K-NN core achieved 150 MHz clock speed and occupied 25% CLB slices based on (B-16, M=8, N=1024, K=13).
❖ The GPP used is a 2.6 GHz Intel Dual Core E5300 running 3GB RAM.
❖ Effect of increasing M was studied as shownbelow:
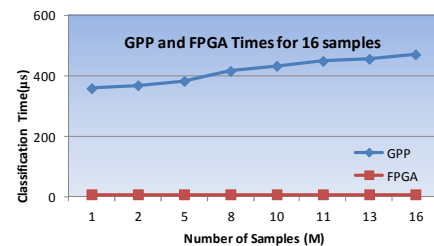
DPR Architecture of three-core Ensemble K-NN classifier
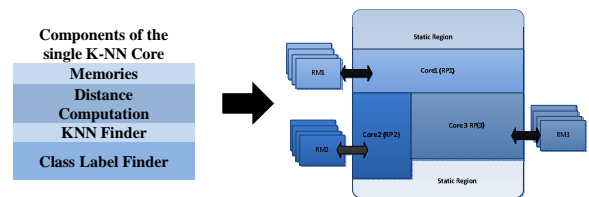
A configuration library based on cores configured with different K's was created whereby the result of the three configurations are combined to obtain the best classification.
The DPR Implementation was ~5x quicker in reconfiguration time than full FPGA reconfiguration.
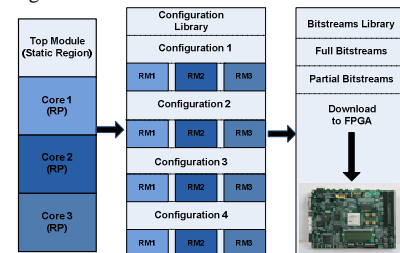
Summary & Conclusion

❖ An FPGA implementation of the K-NN classification offers significantly higher performance in terms of execution time over equivalent GPP implementation (between 60x-92x speed-up).
❖ The DPR Multi-core Ensemble implementation offers the flexibility to alter the parameters of any core without interrupting the operation of others at ~5x speed-up in reconfiguration time over full-chip reconfiguration.
❖The Ensemble K-NN classifier allows combining the results of multiple K-NN cores each based on different K to improve the classification accuracy as K is known to affect the classification accuracy

Acknowledgment