International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 3, Issue 9 (April 2014) PP: 1-11

Automatic Parallelization for Parallel Architectures Using Smith Waterman Algorithm-Literature Review

Ananth Prabhu G [1], Dr. Ganesh Aithal [2]
[1]*Assistant Professor, Dept of CS&E, Sahyadri College of Engineering & Management, Mangalore.*
[2]*Professor, HOD CS&E, P.A. College of Engineering, Mangalore.*

Abstract: *This paper presents a literature survey conducted for research oriented developments made till. The significance of this paper would be to provide a deep rooted understanding and knowledge transfer regarding existing approaches for gene sequencing and alignments using Smith Waterman algorithms and their respective strengths and weaknesses. In order to develop or perform any quality research it is always advised to conduct research goal oriented literature survey that could facilitate an in depth understanding of research work and an objective can be formulated on the basis of gaps existing between present requirements and existing approaches. Gene sequencing problems are one of the predominant issues for researchers to come up with optimized system model that could facilitate optimum processing and efficiency without introducing overheads in terms of memory and time. This research is oriented towards developing such kind of system while taking into consideration of dynamic programming approach called Smith Waterman algorithm in its enhanced form decorated with other supporting and optimized techniques. This paper provides an introduction oriented knowledge transfer so as to provide a brief introduction of research domain, research gap and motivations, objective formulated and proposed systems to accomplish ultimate objectives.*

Keywords: *Smith-Waterman (SW), Multiple-Instruction, Multiple-Data (MIMD), Single Instruction, Multiple Data (SIMD), Smith-Waterman algorithm (SW), Diagonal Parallel Sequencing and Alignment Approach (DPSAA).*

I. INTRODUCTION

Genome sequencing is figuring out the order of DNA nucleotides, or bases, in a genome-the order of As, Cs, Gs, and Ts that make up an organism's DNA. The human genome is made up of over 3 billion of these genetic letters. Today, DNA sequencing on a large scale necessary for ambitious projects such as sequencing an entire genome is mostly done by high-tech machines. Much as your eye scans a sequence of letters to read a sentence, these machines "read" a sequence of DNA bases. A DNA sequence that has been translated from life's chemical alphabet into our alphabet of written letters might look like this:

AGTCCGCGAATACAGGCTCGGT

That is, in this particular piece of DNA, an adenine (a) is followed by a guanine (G), which is followed by a thymine (T), which in turn is followed by a cytosine(C), another cytosine(C), and so on. Sequencing the genome is an important step towards understanding it. At the very least, the genome sequence will represent a valuable shortcut, helping scientists find genes much more easily and quickly. A genome sequence does contain some clues about where genes are, even though scientists are just learning to interpret these clues. Scientists also hope that being able to study the entire genome sequence will help them understand how the genome as a whole works-how genes work together to direct the growth, development and maintenance of an entire organism.

Finally, genes account for less than 25 percent of the DNA in the genome, and so knowing the entire genome sequence will help scientists study the parts of the genome outside the genes. This includes the regulatory regions that control how genes are turned on and off, as well as long stretches of "nonsense" or "junk" DNA-so called because we don't yet know what, if anything, it does.

This is matter of fact that as per increase in computational necessities and problems allied with such huge computation, there has been alarming for coming up with certain optimum approach to perform computation task. On the other hand numerous systems in applications demand better computational approach for computing certain task in minimum time requirement and with enhanced performance. Certain applications like aligning genomes in certain sequential format, it is first required to compute the optimum alignment position and likelihood. But in general there are huge data sets to be processed to get the exact alignment of genes or segments. Therefore there is the need to develop an enhanced system that can accomplish such task. This research work has been motivated by this requirement. The ascending section discussed the key components or gene sequencing and alignment with parallel programming or alignment paradigm.

A. Sequence Alignment

In the field of Bioinformatics, in general the sequence alignment is defined as an approach for performing arrangement of sequences of RNA, DNA or other protein so as to identify the regions of similarity that could be certain consequences of operational, functional, architectural or even existing evolutionary relationships existing between the considered genomic sequences. Associated genomic sequences of proteins such as nucleotide or amino acid remains are characteristically presented as various rows inside a matrix. The existing gaps in sequences are in general inserted in between the remains so as to align the identical characters in its successive columns. The process of sequential alignments are employed in applications related to non-biological sequences and few of the dominant fields where it is being employed are natural language or for processing huge financial data. In case two queries or the sequences are in certain alignment when sharing the common antecedent then it could cause mismatches and it can be efficiently interpreted as certain specific point mutations and gaps I terms of indels which are nothing else but the insertion or deletion kind of mutations process and thus it is introduced in one or mutual ancestry in the time as they have been diverged from each other. In the process of sequence alignments, especially in proteins, the factor called degree of similarity existing between various amino acids while taking space of a specific position in the query or gene sequence might be interpreted as a coarse computation of how preserve a scrupulous region or succession design is amongst lineages. The nonexistence of processes such as substitutions or the occurrence of merely very conventional substitutions which is nothing else but the substitution of amino acids possessing its side chains as similar biochemical characteristics in a specific region of the query or sequence, recommend that this region possesses architectural and operational or functional significance. Since, the bases of DNA and RNA nucleotide are in general similar to each other as compared to the other amino acids; the preservation of base pairs might point toward a comparable operational or functional or architectural role.

B. Alignment Methods

This is the matter of fact that the sequences of very small size can be aligned by hands, but then while, majority of the appealing issues need the alignment of higher length query, exceedingly changeable or enormously frequent sequences which can't be performed for alignments solely by means of any human effort. As an alternative, the knowledge of human is implemented for constructing programs or algorithms to generate high-quality and precise sequential alignments, and sporadically in arranging the ultimate results to reproduce designs which are intricate for presenting algorithmically and it becomes much complicate in case of nucleotide sequences. In general there are two broad computational categories of sequence alignment generally, first states for *global alignments* while another refers for *local alignments*. Estimating by approach of global alignment is nothing else but a scheme for global optimization that enforces the genomic alignment to spread across whole length of query sequences. By disparity, the approach of local alignments describes the regions of connection existing within long genomic sequences which are in general divergent. In fact the approach of local alignment is always recommended, but it might be even more intricate to estimate due to the additional confronts of identifying the various regions of similarity. A number of computational approaches have been developed and have been implemented to the problem of sequence alignment. These algorithms encompass slow but properly accurate schemes such as dynamic programming. These also comprised of highly robust, effective and heuristic algorithms and approaches developed for large-scale database search which is not supposed to deliver guarantee matches.

C. Representations

The genomic alignments are in general stated both graphically. In majority of sequences the presentation of sequence alignment are written in certain defined rows so that the designed or aligned residues comes out to be in the fashion of consecutive columns. In text sequences or formats, the aligned columns encompassed with similar characters are presented with a model of conservation symbols. A number of sequence visualization paradigms also takes into consideration of color for displaying information about the characteristics of the personage sequence elements; in DNA and RNA sequences, this equalizes the assigning individual nucleotide in its own inherent color. In protein alignments, the specific color is in general employed for indicating amino acid characteristics for aiding in judging the preservation of the provided amino acid substitution. In case of multiple queries or sequences the last row in individual column is generally states for the accord sequence estimated by performing alignment; the accord sequence is also always indicated in graphical format with a succession logo where the size of individual nucleotide or amino acid in correspondence is associated with its degree of preservation or conservation.

The sequential alignments might be effectively stored in a extensive categories of text-based data formats, numerous of those are mainly developed in combination with a explicit alignment paradigm. Majority of web-based tools permit a confined amount of inputs and outputs data r file formats, few of the predominant formats are FASTA format and GenBank format and the results are not effortlessly editable. A number of

conversion approaches or paradigms are there which delivers graphical and/or command line interfaces that are accessible like READSEQ and EMBOSS. In spite of these there exist numerous programming packages that facilitate such kind of conversion operations like BioPerl and BioRuby.

II. LITERATURE REVIEW

Driga, A. et al [35] advocated for a pair wise series sequencing approach functional for homology search in bioinformatics and large data sequences or datasets. In case of two datasets in case of two DNA strands or sequences of query length m and n, the full-matrix, dynamic programming (DP) sequential alignment approaches like Needleman-Wunsch and Smith-Waterman consume time of O(mn) and takes into consideration of a prohibitive O(n) space factor. The implementation of Hirschberg's algorithm has exhibited a great extent of reduction in memory occupancy to the level of O(min(m,n)), but still this approach needs approximately twice the needs as made by full-matrix approach. The implementation of fast linear space alignment (FastLSA) paradigm efficiently adapts to the quantity of memory accessible by trading memory space for further function. FastLSA algorithm can effectively consider the implementation of either linear or quadratic block or space, which do depends on the quantity of space available. This work illustrated itself as better for optimized caching facility and a fast processing algorithm. The implementation of wave front parallelization causes the fast process in this algorithm.

Oliver, T.F et al [36] presented a new approach to bio-sequence database scanning using re-configurable field-programmable gate array (FPGA)-based hardware platforms to accomplish higher results with minimum possible cost investment. Highly effective process of mapping in Smith-Waterman approach while employing fine-grained and established mechanism and processing for elements (PEs) which are prepared for certain parameters of a sequence or query was developed. In their work, they employed the approach of customization feasibility and possibilities present at run time for reconfiguring the processing elements dynamically. The implementation of FPGA with developed scheme came out with approximate 170 linear gap penalties while in case of affine group it came out to be 125. In this work the author illustrated the approach in which the run-time reconfiguration be employed for additional performance enhancements.

Hsien-Yu Liao et al [37] advocated a scheme for biological gene sequence using Smith-Waterman algorithm which takes into consideration of the dynamic programming approach possessing higher sensitivity. In their work, they facilitated a significant enhancement in enhancing execution speed and it exhibited its superiority over the conventional sequential approach, while preserving similarity in its functional and performance sensitivity.

Arpit, G. et al [38] developed a noble architecture called Diagonal Linear Space Alignment algorithm that was the enhanced form of FastLSA. The unique and robust approach developed was capable of performing with higher datasets sizes. Then also this approach performs process for storing data of the diagonals of the Dynamic Programming matrix in different way as it is done with FastLSA that exhibits storing of the rows and columns. The researchers have analytically and experimentally proved that their algorithm performs better than FastLSA.

Jha, S. et al [39] developed a better approach using certain privacy preserving based system implementation of basic genomic estimation and processing like the estimation of distance and Smith-Waterman correspondence scores existing amid two query sequences. The approach advocated represents the secure approach with crypto graphically scheme and it was more robust and implementable as compared to other existing approaches. The implementation of their developed prototype was evaluated with respect to the sequences from the database called Pfam that represents a protein family. They illustrated that its performance is sufficient for exhibiting real-world sequence-alignment and associated issues in a privacy-preserving approach. Additionally, their developed scheme encompasses more biological computations. The goal they emphasized was to accomplish certain efficient, privacy-preserving system consideration and its implementation for numerous dynamic programming approaches over certain datasets.

Gardner-Stephen et al [40] developed an algorithm for genomic and proteomic sequencing called DASH. The developed scheme consequences into better performance in terms of higher execution speed as compared to reference system NCBI-BLAST 2.2.6. The implementation of dynamic programming causes much enhancement that enhanced system for its highly robust and effective items exploration. Improving the efficiency of DP provides an opportunity to increase sensitivity, or significantly reduce search times and help offset the effects of the enduring high rate growth with varying datasets of different sizes.

Aji, A.M. et al [41] presented n extremely glowing well prepared organize parallelization of the Smith-Waterman algorithm at the Cell Broadband Engine phase, a new cross multicore structural design that constrain the $low-cost\ PlayStation$ 3 (PS3) play game comfort and the $IBM\ BladeCenter\ Q22$, that presently powers the greatest supercomputer into the world, $Roadrunner$ on $Los\ Alamos\ National\ Laboratory$. During an inventive mapping of the most favorable Smith-Waterman algorithm on top of a cluster of PlayStation PS3 nodes, in this research completion delivers twenty one to fifty five folds up speed above an elevated end

multicore structural design as well as up to $449 - fold$ speed-up in excess of the PowerPC processor inside the $PS3$. Subsequently, the researchers estimated the trade-offs flanked through their Smith- Waterman completion on the Cell with obtainable software in adding up to hardware implementations as well as illustrated that the explanation achieves the most outstanding performance-to-price ratio, whilst aligning realistic series dimension and generating or forming the authentic alignments. At last, the researchers demonstrated that the low-cost explanation on a $PS3$ cluster advance the speed of $BLAST$ while attain ideal compassion. To enumerate the association between the 2 algorithms in conditions of speed as well as sensitivity, the researchers formally describe as well as enumerate the sensitivity of homology investigate techniques so that trade-offs among sequence-search explanation can be appraise in a quantitative method.

Popescu, M. et al [42] advocated a standard word sequence alignment approach on the basis of fuzzy kind of Smith-Waterman (SW) dynamic programming approach. The word similarity matrix used in computing the sequence alignment is calculated based on domain ontology (taxonomy). The fuzzy version of the SW algorithm is designed to accommodate words not present in the initial dictionary used to pre-compute the similarity matrix, hence avoiding its recalculation. The researchers apply the developed algorithm for patient retrieval in an EMR. Each patient is described by an ordered sequence of ICD9 diagnoses. The researchers analyze various properties of the proposed algorithm on a patient dataset that contains 107 patients described by ICD9 diagnose sequences.

Rashid et al [43] proposed main most favorable algorithm has been used in succession alignment has been standing by the dynamic programming techniques. Smith-Waterman algorithm is the usually employed dynamic programming based series alignment algorithm. Conversely the algorithm employed quadratic instance as well as space. Heuristic algorithm such as $FASTA$ as well as $BLAST$ these both has been introduced to rapidity up the series alignment algorithm. FASTA is stand on word investigate where the BLAST is stand on utmost segment pairs. During word investigated algorithm, lists of words as of the inquiry and database succession are organism compared to resolve if 2 progressions have a section of satisfactory resemblance to merit additional alignment applying the Smith-Waterman Algorithm. The present algorithm has been applied the removed amino acids alphabet to modify the protein sequences in a sequence of integer as well as employed n-gram to decrease the length of the series. After that the $Smith - Waterman$ algorithm has been utilized to obtain the relationship measure involving 2 sequences.

Changjin Hong et al [44] advocated a highly efficient mechanism for performing updation of local sequence alignments approach with certain affine gap framework. Principally, employing existing results for sequence matching amid two sequences of amino acid, in this work the authors performed a noble approach called forward-backward sequence alignment for generating exploring bands of heuristic types that are in general bounded by a combination of certain suboptimal paths. With provided conditional updated sequence, the authors made prediction ffor new score matrix of the sequence alignment for individual contour for selecting the optimum or potential candidates. At later stage the authors then implemented Smith-Waterman algorithm with certain definite goal. Additionally, the proposed heuristic sequence alignment for an updated sequence illustrated that the proposed system might be further enhanced by implementing certain reusable dynamic programming (rDP). In their work, they validated the usability of $"relative\ node\ tolerance\ boundrdquo\ (RNTB)$ in the pruned exploration space. Additionally, they enhanced the computational efficiency by incorporating quantification and probability analysis of the RNTB tolerance and switching process to rDP on certain perturbation-protected columns. In their searching space derived by a threshold value of 90 percent of the optimal alignment score, the researchers find that 98.3 percent of contours contain correctly updated paths.

Arslan, A.N.et al [45] constructed a weighted finite automaton from a given regular expression, and presents a dynamic programming solution that simulates copies of this automaton in looking for data sequencing with utmost score comprised of the standard expressions. The researchers generalize this approach: 1) The researchers introduce a variation of the problem for multiple sequences, stated by the standard expression inhibited multiple sequence alignment scheme; 2) the authors advocated and developed a robust technique for the situation when the sequential alignments needs are advocated for containing certain query sequence of standard expressions.

Pulka, A et al [46] contract with a really warm difficulty relating to computation biology - the penetrating for a specified orientation prototype in a incredibly extended DNA chain. The software explanation in the argument area is in completed by quantity of resources as well as processing period. To facilitate is why multifaceted programmable campaign are additional with further usually used in the applications regarding microbiology. The dissertation proposed the method that is a customized Smith-Waterman active programming technique. The optimization of the whole algorithm as well as utilized resources is completed with admiration of belongings of $FPGA$ components.

Shi-Yi Shen et al [47] described generalized errors by mutation model which comes from bioinformatics. To deal with the sequences with generalized errors, the researchers also need some type of metric to measure the distance of sequences with various lengths. The query alignment distance is stated by the distance factor existing amid two query/ data sequences possessing certain errors of generalized category. The algorithms to calculate the alignment distance were studied in bioinformatics recently, such as Smith-Waterman dynamic programming algorithm and SPA algorithm. Under the alignment distance, the researchers give the definition of the alignment space. For describing the architecture in precise manner for aforementioned scenario, the author then introduced a mechanism of modular structure theory. A new and stricter proof of triangle inequality for alignment distance is thus given. In this research work the authors analyzed the implementation or usability of the alignment space; the model definition and the associated architecture of generalized error-correcting codes etc. Few of the best performing algorithms and associated codes with small length have been mentioned. The capacity of error-correction in random codes possessing longer query length has been discussed. The researchers discuss fault-tolerance complex in cryptography as another application of the alignment space at the end of the paper

Voss, G. et al [48] proposed a novel method to elevated presentation biological sequence database scanning at graphics dispensation components. Apply modern graphics dispensation units for elevated presentation calculating is making possible through after that- enhanced program capability as well as motivated by their stunning price/presentation ratio along with enormous growth in velocity. To develop a well-organized mapping on top of this kind of structural design, the researchers have reformulated the Smith-Waterman energetic programming algorithm in terms of computer graphics primordial. The outcome of result is in implementation with important runtime savings on 2 standard off-the-shelf computer realistic cards. As for information this is the initial description mapping of biological succession alignment on top of a graphics dispensation unit.

Harris, B. et al [49] explained a narrative $FPGA$ design intended for banded Smith-Waterman, an algorithmic alternative tune to the requirements of $BLASTP$. These intend design is executed in Mercury$BLASTP$, their $FPGA$-accelerated description of the $BLASTP$ algorithm. The researchers demonstrate that Mercury $BLASTP$ executes six-sixteen period quicker than software $BLASTP$ on a contemporary CPU while distribute 99% the same consequences output.

Knowles, G. et al [50] presented a novel system model for exhibiting an efficient protocol for biological as well as proteomic gene sequencing that ultimately accomplishes an enhanced execution time acceleration by twice or three times as compared to Smith-Waterman dynamic programming (DP) in hardware. In their previous papers, the researchers introduce several features of their search algorithm, DASH, which outperforms NCII-Blast (BLAST) by an order of magnitude in software, and it possesses better functional sensitivity. In fact, the proposed system DASH was illustrated with its enhanced sensitivity as compared to Smith-Waterman algorithm. This scheme was developed while employing the concept of biological progression to comprise of regions with higher homology intersperse with regions of comparatively lower homology. The proposed approach DASH, the best suitable solution comprises similar diagonals connected by regions of exact dynamic programming. This is affordable due to the small area of these interconnecting regions.

Weiguo Liu et al [51] proposed an innovative proceed to bio-sequence database scan employing computer graphics hardware to increase elevated presentation at low cost. To obtain a well-organized mapping on top of this kind of structural design, the researchers have re-prepared the Smith-Waterman dynamic programming algorithm in conditions of computer graphics primitives. Their OpenGL completions attain a speedup of roughly 16 on an elevated end graphics certificate over obtainable uncomplicated as well as optimized CPU Smith-Waterman implementations.

Alpern, B. et al [52] explored the use of some standard and novel techniques for improving its performance. The researchers begin by tuning the algorithm using conventional techniques. These make modest performance improvements by providing efficient cache usage and inner-loop code. One novel technique uses the z-buffer operations of the Intel i860 architecture to perform 4 independent computations in parallel. This achieves a five-fold speedup over the optimized code (six-fold over the original). The researchers also describe a related technique that could be used by processors that have 64-bit integer operations, but no z-buffer. Another new technique uses floating-point multiplies and adds in place of the standard algorithm's integer additions and maximum operations. This gains more than a three-fold speedup on the IBM POWER2 processor. This method doesn't give the identical answers as the original program, but experimental evidence shows that the inaccuracies are small and do not affect which strings are chosen as good matches by the algorithm.

Zheng, Fang et al [53] used Compute Unified Device Architecture (CUDA) GPU to accelerate pair wise sequence alignment using the Smith-Waterman (SW) algorithm. Smith-Waterman (SW) is by far the best algorithm for its accuracy in similarity scoring. But the executing time of this algorithm is too long in sequence

alignment. So the researchers describe a multi-threaded parallel design and implementation of the Smith-Waterman (SW) on CUDA to reduce execution time. And according the architecture of CUDA, the researchers have divided the computation of a whole pair wise sequence alignment scoring matrix into multiple sub-matrices, using 32 threads to process on sub matrices, more over the researchers optimized memory distribution scheme, and used reduction to find the maximum element of the alignment scoring matrix. The researchers experiment the algorithm on GeForce 9600 GT, connect to Windows xp 64-bit system. The results show this implementation achieves better performance than the other parallel implementation on the Graphics Processing Unit.

Das, S. et al [54] proposed an algorithm for local alignment between two DNA sequences and compare the performance of the proposed algorithm with Smith-Waterman algorithm. Complexity calculation shows that the proposed algorithm has a much less time complexity and requires very much less amount of memory storage than S-W algorithm.

Junid et al [55] planned a novel advance method has identified to decrease the difficulty of the Smith Waterman algorithm has been used for $FPGA$ implementation. The researchers have proposed the method for the greatest judgment of the 2 DNA sequencing employing Verilog scheduled Xilinx ISE 7.1. The Simulation is executing on the ModelSim XE III 6.0. The combinational stoppage for the presented planned smith waterman algorithm which stands on separate as well as conquers method sequencing is 10.214 ns whilst the innovative is 10.295 ns. The researchers have established that smith waterman algorithm is stands on separate along with conquer method provide improved performance than obtainable method.

III. PROBLEM DEFINITION

In parallel computing scenario the approach of sequential alignment and sequence approach is a robust and elementary model that assists genome scientists to surmise the biological associations or relationships from a huge datasets of associated DNA and allied sequences of protein. Such kind of mammoth works cannot be solved easily by exhibiting conventional string matching operations due to the reason that the genomic queries or sequences which do share similar biological function transmute over time factor when it is out in the open with the evolutionary phenomenon or events, and it might not be effective for matching identically for long time. The other schemes for performing process of genomic sequencing and its comparison like BLAST and Hidden Markov Models (HMM) exhibits its function on the basis of approximation of string matching philosophy that estimates the on the whole resemblance existing between genomic strings and are therefore more robust and forbearing of mismatches. Such kind of fuzzy approaches of string matching issues and applications might be developed or can be formulated in two diverse approaches. Even the similarity existing between two dissimilar or parallel strings can be accomplished unambiguously, by reducing certain factors such as ad hoc cost function or edit distance factor over all probable sequential alignments existing between those parallel strings.

On the other hand, the metrics or the similarity score can also be estimated stochastically, by at first estimating the highest possible path with the help of a hidden Markov model (HMM) which has already been trained from certain sequential input string. These approaches being discussed here are optimization issues and schemes that need certain optimized dynamic programming (DP) approach to accomplish the goal. In fact the dynamic programming approaches are much expensive in terms of computational cost and even it becomes more when it is supposed to perform for comparing large genomic strings. Providentially, the data enslavement in the repetition associations' permit certain confined degree of parallelism and the calculation for certain cell entries of the Dynamic Programming lookup table can further be circulated across certain defined a set of parallel processors.

The Smith-Waterman (SW) algorithm performs exploration for a query or sequence database for identifying the available similarities between certain query sequence and a subject sequences. Then while, such kind of algorithm are least efficient in terms of its performance parameters such as complexity factors in time and space; the highly rated growth of query sequence also encompasses numerous issues related to computation. The Smith-Waterman (SW) algorithm [14][15] presents a kind of dynamic programming approach for identifying the enhanced local sequential alignments of biological gene pairs. Because of its highest optimum sensitivity for accomplishing local alignments, this kind of parallel programming approach is a fundamental process in bioinformatics computation, comprising exploration for biological sequence database and manifold query sequence alignment needs[16][17] and ultimately next-generation query or data sequencing and sequence alignment [18][19].

IV. RESEARCH OBJECTIVE

Considering the requirement of a highly robust and efficient system model and approach for parallel computing applications, here in this research work, the author proposes to develop a highly optimized parallel programming approach that come up with the optimum sequencing scheme in the form of Diagonal alignment of genes. The overall objectives of this research work have been classified into two categories. The first states for General Objectives while another represents Specific Objectives. General objectives are those goal primitives which represents the ultimate goal of the proposed research. These objectives present the overview of the goal, which is being proposed to be developed in this thesis work. While Specific objectives are those objectives which do specifies the exact and precise presentation of approaches and techniques to be used to accomplish the general objectives. These research objectives have been given as follows:

A. GENERAL OBJECTIVES

- To develop a novel scheme for Biological gene sequencing application.
- To develop a highly robust and efficient approach for parallel computing that could deliver higher computational efficiency even with minimum time and space requirements for varying sequential length.
- To develop Diagonal Parallel Sequencing and Alignment Approach (DPSAA) for parallel computing applications.
- To develop an approach for enhancing conventional Smith Waterman (SW) algorithm for gene sequencing and alignment applications.
- To compare the proposed Diagonal sequencing or alignment approach with other existing parallel and serial sequential alignment techniques and justify that diagonal query sequencing can be of immense significance.

B. SPECIFIC OBJECTIVES

- To develop a highly robust and optimized Smith Waterman algorithm based system model for optimal and efficient local sequence alignment applications.
- To develop an Intra-task kernel based parallelization technique for ***Diagonal Parallel Gene Sequence Alignment Approach*** that could bring minimum time complexity with higher computation rate with minimum instruction sets required to execute sequential queries as compared to other parallel as well as serial gene sequencing techniques.
- To implement the optimized Myers-Miller algorithm approach with a goal to minimize the memory occupancy and speed up the processing rate even with enhanced computation.
- To develop an optimization approach for accomplishing enhanced Backtracking facility, optimum sequential distance computation, pairwise estimation and parallelization of progressive sequential alignment facility for Diagonal sequencing alignment applications such as biological gene sequencing.

V. PROPOSED SYSTEM

Considering the requirement of a highly optimized and efficient system for gene alignment and biological sequencing with parallel computing process, the author of this research has proposed Smith Waterman based system developed that could accomplish higher computation rate and gene sequencing. In this research work, the author has proposed a Smith Waterman algorithm based diagonal sequencing approach that could accomplish the higher rate sequencing with diagonal scheme. Here in this research work, the author proposes to implement Myers and Millers algorithm enriched sequencing system that can accomplish optimum performance even with minimum space utilization and intricacy. The author proposes to implement Intra-task parallelization based backtracking and sequencing while optimizing trace-backing and optimum midpoint estimation. In this work, in order to accomplish a fast sequencing and processing facility the author advocate for a sequential distance computation algorithm that estimates the distance between metrics and nodes so as to speed up the overall processing of alignments. For developing enhanced sequential distance estimation algorithm, here in this work the author proposes to develop a novel pairwise alignment scheme that achieves the goal of optimal local alignment by means of a forward pass and a reverse pass on the alignment matrix of two different sequences using the smith watermark algorithm. Here in this work, the author takes into consideration of intra-task parallelogram kernels so as to reduce memory as well as computational counts. On the other hand the consideration of DP lookup table has been proposed only for reducing computational cost and speedup ratio while performing on gene sequence with longer query length. The emphasis on trace backing has been advocated so as to enrich forward and reverse sequencing paradigm and with reduced exploration overheads. The author proposes to develop an enhanced algorithm for optimum trace backing, mean estimation and sequence distance computation. Even here in this work, for precise and efficient alignment purpose the author has proposed for optimized pairwise alignment issue that takes care of every associated parametric enhancement on smith watermark approach. Figs. 1-3 show the sequential, parallel and diagonal approaches.

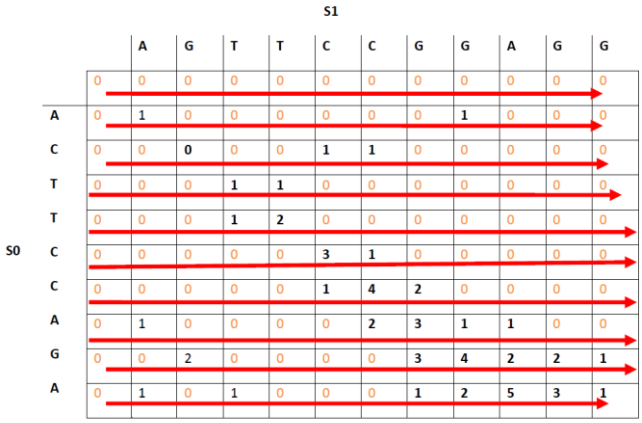

Fig. 1: Sequential approach.

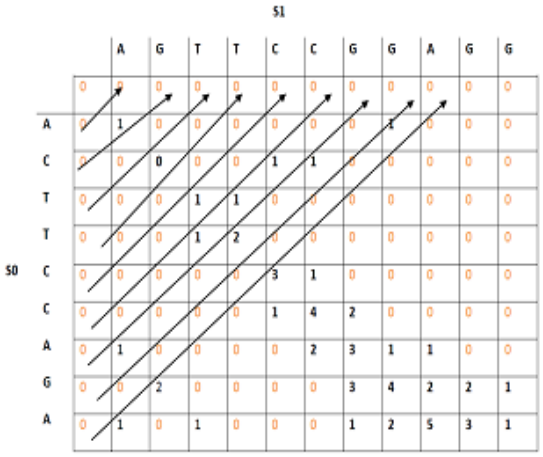

Fig. 2: Parallel approach.

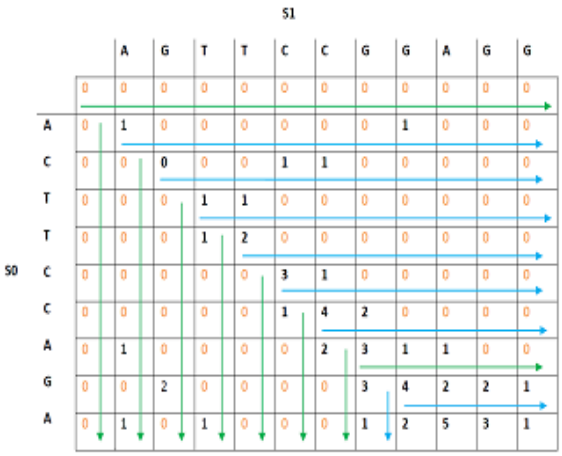

Fig. 3: Diagonal approach.

VI. CONCLUSION

In this work, the predominant factors which has been optimized with its optimum possibility is Smith Waterman algorithm which functions in a unique ***Diagonal Parallel Alignment*** rather being in conventional serial or parallel sequencing approaches. The development of diagonal parallel sequencing makes the system efficient for estimating distance matrices and query matching with neighbour matrices that reduces overall computational count and thus the overall computational cost gets reduced. Similarly, on contrary of conventional serial and parallel sequencing the proposed approach has performed much better in terms of *Query execution time; speed up ratio, number of computation, reduction in computational costs* etc.

REFERENCES

[1] F. Guinand, "Parallelism for computational molecular biology," in ISThmus 2000 Conference on Research and Development for the Information Society, Poznan, Poland, 2000.

[2] L. D'Antonio, "Incorporating bioinformatics in an algorithms course," in Proceedings of the 8th annual conference on Innovation and Technology in Computer Science Education, vol. 35 (3), 2003, pp. 211{214.

[3] H. B. J. Nicholas, D. W. D. II, and A. J. Ropelewski. (Revised 1998) Sequence analysis tutorials: A tutorial on search sequence databases and sequence scoring methods. [Online]. Available: http://www.nrbsc.org/old/education/tutorials/ sequence/db/index.html

[4] X. Huang, Chapter 3: Bio-Sequence Comparison and Alignment, ser. Current Topics in Computational Molecular Biology. Cambridge, MA: The MIT Press, 2002.

[5] S. Needleman and C. Wunch, "A general method applicable to the search for similarities in the amino acid sequences of two proteins," Journal of Molecular Biology, vol. 48, no. 3, pp. 443{453, 1970.

[6] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," Journal of Molecular Biology, vol. 147, no. 1, pp. 195{197, 1981.

[7] O. Gotoh, "An improved algorithm for matching biological sequences," Journal of Molecular Biology, vol. 162, no. 3, pp. 705-708, 1982.

[8] X. Huang and W. Miller, "A time-efficient linear-space local similarity algorithm," Adv.Appl.Math., vol. 12, no. 3, pp. 337{357, 1991.

[9] M. Camerson and H. Williams., "Comparing compressed sequences for faster nucleotide blast searches," IEEE/ACM Transactions on Computational Biology and Bioinformatics, vol. 4, no. 3, pp. 349{364, 2007.

[10] J. D. Frey, "The use of the smith-waterman algorithm in melodic song identification." Master's Thesis, Kent State University, 2008.
[11] J. Potter, J. W. Baker, S. Scott, A. Bansal, C. Leangsuksun, and C. Asthagiri, "Asc: an associative-computing paradigm," Computer, vol. 27, no. 11, pp. 19{25, 1994.
[12] M. J. Quinn, Parallel Computing: Theory and Practice, 2nd ed. New York: McGraw-Hill, 1994.
[13] J. Baker. (2004) Simd and masc: Course notes from cs 6/73301: Parallel and distributed computing - power point slides. [Online]. Available: http://www.cs.kent.edu/wchantam/PDC Fall04/SIMD MASC.ppt
[14] Smith T, Waterman M: Identification of common molecular subsequences. J Mol Biol 1981, 147:195–197.
[15] Gotoh O: An improved algorithm for matching biological sequences. J Mol Biol 1982, 162:707–708.
[16] Thompson JD, Higgins DG, Gibson TJ: CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994, 22:4673–4680.
[17] Liu Y, Schmidt B, Maskell DL: MSA-CUDA: Multiple Sequence Alignment on Graphics Processing Units with CUDA. 20th IEEE International Conference on Application-specific Systems, Architectures and Processors; 2009:121–128.
[18] Li H, Durbin R: Fast and accurate short read alignment with Burrows Wheeler transform. Bioinformatics 2009, 25(14):1755–1760.
[19] Liu Y, Schmidt B, Maskell DL: CUSHAW: a CUDA compatible short read aligner to large genomes based on the Burrows-Wheeler transform. Bioinformatics 2012, 28(14):1830–1837.
[20] Pearson WR, Lipman DJ: Improved tools for biological sequence comparison. Proc. Nat. Acad. Sci. USA 1988, 85(8):2444–2448.
[21] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basical local alignment search tool. J Mol Biol 1990, 215(3):403–410.
[22] Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997, 25(17):3389–3402.
[23] R.S. Harris, "Improved Pairwise Alignment of Genomic DNA," PhD thesis, The Pennsylvania State Univ., 2007.
[24] S. Kurtz, A. Phillippy, A.L. Delcher, M. Smoot, M. Shumway, C. Antonescu, and S.L. Salzberg, "Versatile and Open Software for Comparing Large Genomes," Genome Biology, vol. 5, no. 2, p. R12, 2004.
[25] S. Aluru, N. Futamura, and K. Mehrotra, "Parallel Biological Sequence Comparison Using Prefix Computations," J. Parallel Distributed Computing, vol. 63, no. 3, pp. 264-272, 2003.
[26] S. Rajko and S. Aluru, "Space and Time Optimal Parallel Sequence Alignments," IEEE Trans. Parallel Distributed Systems, vol. 15, no. 12, pp. 1070-1081, Dec. 2004.
[27] R.B. Batista, A. Boukerche, and A.C.M.A. de Melo, "A Parallel Strategy for Biological Sequence Alignment in Restricted Memory Space," J. Parallel Distributed Computing, vol. 68, no. 4, pp. 548-561, 2008.
[28] C. Chen and B. Schmidt, "Computing Large-Scale Alignments on a Multi-Cluster," Proc. IEEE Int'l Conf. Cluster Computing, pp. 3845, 2003.
[29] P. Zhang, G. Tan, and G.R. Gao, "Implementation of the SmithWaterman Algorithm on a Reconfigurable Supercomputing Platform," Proc. First Int'l Workshop High-Performance Reconfigurable Computing Technology and Applications: Held in Conjunction with SC07 (HPRCTA '07), pp. 39-48, 2007.
[30] X. Liu, L. Xu, P. Zhang, N. Sun, and X. Jiang, "A Reconfigurable Accelerator for Smith-Waterman Algorithm," IEEE Trans. Circuits and Systems, vol. 54, no. 12, pp. 1077-1081, Dec. 2007.
[31] A. Boukerche, J.M. Correa, A.C.M.A. de Melo, R.P. Jacobi, and A.F. Rocha, "Reconfigurable Architecture for Biological Sequence Comparison in Reduced Memory Space," Proc. IEEE Int'l Parallel and Distributed Processing Symp. (IPDPS), pp. 1-8, 2007.
[32] C. Chen and B. Schmidt, "An Adaptive Grid Implementation of DNA Sequence Alignment," Future Generation Computer Systems, vol. 21, no. 7, pp. 988-1003, 2005.
[33] F. Sanchez, F. Cabarcas, A. Ramirez, and M. Valero, "Long DNA Sequence Comparison on Multicore Architectures," Proc. 16th Int'l Euro-Par Conf. Parallel Processing (Euro-Par), pp. 247-259, 2010.
[34] A. Sarje and S. Aluru, "Parallel Genomic Alignments on the Cell Broadband Engine," IEEE Trans. Parallel Distributed, vol. 20, no. 11, pp. 1600-1610, Nov. 2009.
[35] Driga, A.; Lu, P.; Schaeffer, Jonathan; Szafron, D.; Charter, K.; Parsons, I., "FastLSA: a fast, linear-space, parallel and sequential algorithm for sequence alignment," Parallel Processing, International Conference on 9-9 Oct. 2003, pp.48-57.
[36] Oliver, T.F.; Schmidt, B.; Maskell, D.L., "Reconfigurable architectures for bio-sequence database scanning on FPGAs," Circuits and Systems II: Express Briefs, IEEE Transactions on Dec. 2005, vol.52, no.12, pp.851-855.
[37] Hsien-Yu Liao; Meng-Lai Yin; Cheng, Y., "A parallel implementation of the Smith-Waterman algorithm for massive sequences searching," Engineering in Medicine and Biology Society,26th Annual International Conference of the IEEE , vol.2, pp.2817,2820, 1-5 Sept. 2004
[38] Arpit, G.; Adiga, R.; Varghese, K., "Space Efficient Diagonal Linear Space Sequence Alignment," BioInformatics and BioEngineering (BIBE), 2010 IEEE International Conference on , vol., no., pp.244,249, May 31 2010-June 3 2010
[39] Jha,S.; Kruger, L.; Shmatikov, V., "Towards Practical Privacy for Genomic Computation," Security and Privacy, 2008. SP 2008. IEEE Symposium on 18-22 May 2008, pp.216,230,
[40] Gardner-Stephen, P.; Knowles, G., "DASH: localizing dynamic programming for order of magnitude faster, accurate sequence alignment," Computational Systems Bioinformatics Conference, IEEE on 16-19 Aug. pp.732-735.
[41] Aji, A.M.; Wu-chun Feng, "Optimizing performance, cost, and sensitivity in pairwise sequence search on a cluster of PlayStations," BioInformatics and BioEngineering 8th IEEE International Conference on 8-10 Oct. 2008, pp.1-6.
[42] Popescu,M., "An ontological fuzzy Smith-Waterman with applications to patient retrieval in Electronic Medical Records," Fuzzy Systems (FUZZ), IEEE International Conference on 18-23 July 2010, pp.1-6.
[43] Rashid, N.A.A.; Abdullah, R.; Talib, A.Z.H.; Ali, Z., "Fast Dynamic Programming Based Sequence Alignment Algorithm," Distributed Frameworks for Multimedia Applications, 2006. The 2nd International Conference on , vol., no., pp.1,7, May 2006
[44] Changjin Hong; Tewfik, A.H., "Heuristic Reusable Dynamic Programming: Efficient Updates of Local Sequence Alignment," Computational Biology and Bioinformatics, IEEE/ACM Transactions on Oct.-Dec. 2009, vol.6, no.4, pp.570-582.
[45] Arslan A.N; "Multiple Sequence Alignment Containing a Sequence of Regular Expressions"; Computational Intelligence in Bioinformatics and Computational Biology, 2005. Proceedings of the IEEE Symposium on. 14-15 Nov. 2005, pp.1-7.
[46] Pulka, A.; Milik, A., "A new hardware algorithm for searching genome patterns," Signals and Electronic Systems, 2008.; International Conference on 14-17 Sept. 2008, pp.181-184.
[47] Shi-Yi Shen; Kui Wang; Gang Hu; Shu-Tao Xia; "On the Alignment Space and its applications," Information Theory Workshop on 22-26 Oct. 2006, vol., no., pp.165-169.
[48] Voss, G.; Muller-Wittig, W.; Schmidt, B., "Using Graphics Hardware to Accelerate Biological Sequence Database Scanning," TENCON 2005 2005 IEEE Region 10 , vol., no., pp.1,6, 21-24 Nov. 2005

[49] Harris, B.; Jacob, A.C.; Lancaster, J.M.; Buhler, J.; Chamberlain, R.D., "A Banded Smith-Waterman FPGA Accelerator for Mercury BLASTP," Field Programmable Logic and Applications, International Conference on 27-29 Aug. 2007, pp.765-769.

[50] Knowles, G.; Gardner-Stephen, P; "A new hardware architecture for genomic and proteomic sequence alignment," Computational Systems Bioinformatics Conference, 2004. CSB 2004. Proceedings. 2004 IEEE , vol., no., pp.730,731, 16-19 Aug. 2004

[51] Weiguo Liu; Schmidt, B.; Voss, G.; Schroder, A.; Muller-Wittig, W., "Bio-sequence database scanning on a GPU," Parallel and Distributed Processing Symposium on 25-29 April 2006, pp.8

[52] Alpern B.; Carter, L.; Kang Su Gatlin, "Micro-parallelism and High-Performance Protein Matching"; Supercomputing Proceedings of the IEEE/ACM SC95; pp.24-24.

[53] Zheng, Fang; Xu, Xianbin; Yang, Yuanhua; He, Shuibing; Zhang, Yuping, "Accelerating Biological Sequence Alignment Algorithm on GPU with CUDA," Computational and Information Sciences (ICCIS), International Conference on 21-23 Oct. 2011, pp.18-21.

[54] Das, S.; Dey, D., "A new algorithm for local alignment in DNA sequencing"; India Annual Conference; Proceedings of the IEEE INDICON on 20-22 Dec. 2004; pp.410-413.

[55] Junid, S.A.M.; Majid, Z.A.; Halim, A.K., "Development of DNA sequencing accelerator based on Smith Waterman algorithm with heuristic divide and conquer technique for FPGA implementation," Computer and Communication Engineering; International Conference on on 13-15 May 2008; pp.994-996.

[56] Boukerche, A.; Correa, J.M.; de Melo, A.C.M.A.; Jacobi, R.P.; Rocha, A.F., "Reconfigurable Architecture for Biological Sequence Comparison in Reduced Memory Space," Parallel and Distributed Processing Symposium, IPDPS 2007; IEEE International on 26-30 March 2007; pp.1-8.

[57] Delgado, G.; Aporntewan, C.; "Data dependency reduction in Dynamic Programming matrix," Computer Science and Software Engineering (JCSSE)Eighth International Joint Conference on 11-13 May 2011; pp.234-236.

[58] Riedel, D.E.; Venkatesh, S.; Wanquan Liu, "A Smith-Waterman Local Alignment Approach for Spatial Activity Recognition"; Video and Signal Based Surveillance, AVSS '06; IEEE International Conference on Nov. 2006; pp.54-54.

[59] Fa Zhang; Xiang-Zhen Qiao; Zhi-Yong Liu; "A parallel Smith-Waterman algorithm based on divide and conquer" Algorithms and Architectures for Parallel Processing Proceedings. Fifth International Conference on 23-25 Oct. 2002; pp.162-169.

[60] Sebastiao, N.; Dias, T.; Roma, N.; Flores, P., "Integrated accelerator architecture for DNA sequences alignment with enhanced traceback phase," High Performance Computing and Simulation (HPCS), 2010 International Conference on June 28 2010-July 2 2010; pp.16-23.

[61] Allred, J.; Coyne, J.; Lynch, W.; Natoli, V.; Grecco, J.; Morrissette, J., "Smith-Waterman implementation on a FSB-FPGA module using the Intel Accelerator Abstraction Layer," Parallel & Distributed Processing IPDPS in IEEE International Symposium on 23-29 May 2009; pp.1-4.

[62] Hasan,L.; Al-Ars, Z., "An efficient and high performance linear recursive variable expansion implementation of the smith-waterman algorithm," Engineering in Medicine and Biology Society, Annual International Conference of the IEEE3-6 Sept. 2009; pp.3845-3848.

[63] Razmyslovich, D.; Marcus, G.; Gipp, M.; Zapatka, M.; Szillus, A., "Implementation of Smith-Waterman Algorithm in OpenCL for GPUs," Parallel and Distributed Methods in Verification, 2010 Ninth International Workshop on, and High Performance Computational Systems Biology, Second International Workshop on Sept. 30 2010-Oct. 1 2010; pp.48-56.

[64] Cheng Ling; Benkrid, K.; Hamada, T; "A parameterizable and scalable Smith-Waterman algorithm implementation on CUDA-compatible GPU"; Application Specific Processors IEEE 7th Symposium on 27-28 July 2009; pp.94-100.

[65] Steinfadt, S.I., "SWAMP+: Enhanced Smith-Waterman Search for Parallel Models," Parallel Processing Workshops (ICPPW), 2012 41st International Conference on 10-13 Sept. 2012; pp.62-70.

[66] Nordin, M.; Rahman, A., "Utilizing MPJ Express Software in Parallel DNA Sequence Alignment," Future Computer and Communication, ICFCC 2009. International Conference on 3-5 April 2009; pp.567-571.

[67] Qian Zhang; Hong An; Gu Liu; Wenting Han; Ping Yao; Mu Xu; Xiaoqiang Li, "The optimization of parallel Smith-Waterman sequence alignment using on-chip memory of GPGPU," Bio-Inspired Computing: Theories and Applications (BIC-TA), 2010 IEEE Fifth International Conference on 23-26 Sept. 2010; pp.844-850.

[68] W. Liu; B. Schmidt; G. Voss; A. Schroder; W. Muller-Wittig; "Bio-Sequence Database Scanning on a GPU" Proc. 20th Int'l Conf. Parallel and Distributed Processing (IPDPS), 2006.

[69] Ali Khajeh-Saeed; Stephen Poole; J. Blair Perot; "Acceleration of the Smith–Waterman algorithm using single and multiple graphics processors"; Journal of Computational Physics, 229 (2010) 4247–4258.

[70] Sarje and S. Aluru, "Parallel Genomic Alignments on the Cell Broadband Engine," IEEE Trans. Parallel Distributed, vol. 20, no. 11, pp. 1600-1610, Nov. 2009.

[71] Y. Liu, W. Huang, J. Johnson, and S. Vaidya, "GPU Accelerated Smith-Waterman," Proc. Sixth Int'l Conf. Computational Science (ICCS), vol. 3994, pp. 188-195, 2006.

[72] Sheng-Ta Lee; Chun-Yuan Lin; Che Lun Hung; "GPU-Based Cloud Service for Smith-Waterman Algorithm Using Frequency Distance Filtration Scheme"; BioMed Research International Volume 2013 (2013), Article ID 721738, 8 pages.

[73] Sean O. Settle; "High-performance Dynamic Programming on FPGAs with OpenCL"; IEEE 2013.

[74] David Uliana; Krzysztof Kepa; Peter Athanas; "FPGA-based HPC application design for non-experts"; 2013 IEEE 24th International Conference on Application-Specific Systems, Architectures and Processors on June 05-June 07.

[75] Cehn, C.; Schmidt, B., "Computing large-scale alignments on a multi-cluster," Cluster Computing; Proceedings. 2003 IEEE International Conference on 1-4 Dec. 2003, pp.38-45.

[76] Batista, R.B.; Magalhaes Alves de Melo, Alba Cristina, "Z-align: An Exact and Parallel Strategy for Local Biological Sequence Alignment in User-Restricted Memory Space," Cluster Computing, 2006 IEEE International Conference on 25-28 Sept. 2006, pp.1-10.

[77] Rajko, S.; Aluru, S., "Space and time optimal parallel sequence alignments," Parallel and Distributed Systems, IEEE Transactions on Dec. 2004, vol.15, no.12, pp.1070-1081.

[78] Michael S. Farrar; "Optimizing Smith-Waterman for the Cell Broadband Engine.

[79] Gabor Ivan; Daniel Bank; Vince Grolmusz; "Fast and Exact Sequence Alignment with the Smith-Waterman Algorithm: The swissAlighn Webserver"; 7 Sep 2013.

[80] Yoshiki Yamaguchi; Hung Kuen Tsoi; Wayne Luk; "FPGA-based smith-waterman algorithm: analysis and novel design"; Proceeding ARC'11 Proceedings of the 7th international conference on Reconfigurable computing: architectures, tools and applications; Pages 181-192, 2011.

[81] A Surendar; M Arun; C Bagavathi; "EVOLUTION OF RECONFIGURABLE BASED ALGORITHMS FOR BIOINFORMATICS APPLICATIONS: AN INVESTIGATION"; International journal of life sciences, Biotechnology and Pharma Research; Vol. 2, No. 4, October 2013.
[82] Doug Hains; Zach Cashero; Mark Ottenberg;Wim Bohm; Sanjay Rajopadhye; "Improving CUDASW++, a Parallelization of Smith-Waterman for CUDA Enabled Devices.
[83] Michael Christopher Schatz; "HIGH PERFORMANCE COMPUTING FOR DNA SEQUENCE ALIGNMENT AND ASSEMBLY"; Thesis University of Maryland in 2010.
[84] Talal Bonny; M. Affan Zidan; Khaled N. Salama; "An Adaptive Hybrid Multiprocessor Technique for Bioinformatics Sequence Alignment".
[85] Ligowski, L.; Rudnicki, W.; "An efficient implementation of Smith Waterman algorithm on GPU using CUDA, for massively parallel scanning of sequence databases," Parallel & Distributed Processing, 2009; IEEE International Symposium on 23-29 May 2009, pp.1-8.
[86] Mohd Nazrin Md Isa, "High Performance Reconfigurable Architectures for Biological Sequence Alignment"; Thesis; University of Edinburgh, March 2013.
[87] Brian Hang; Wai Yang; "A Parallel Implementation of Smith-Waterman Sequence Comparison Algorithm"; December 6, 2002.
[89] Jay Shendure; Hanlee Ji; "Next-generation DNA sequencing"; nature biotechnology volume 26 number 10; October 2008.
[90] Xiandong Meng; Chaudhary, V., "Exploiting Multi-level Parallelism for Homology Search using General Purpose Processors," Parallel and Distributed Systems, 2005. Proceedings. 11th International Conference on 22-22 July 2005; vol.2, pp.331,335,
[91] Zhihui Du; Zhaoming Yin; Bader, D.A., "A tile-based parallel Viterbi algorithm for biological sequence alignment on GPU with CUDA," Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on 19-23 April 2010; pp.1-8.
[92] Shucai Xiao; Aji, A.M.; Wu-chun Feng, "On the Robust Mapping of Dynamic Programming onto a Graphics Processing Unit," Parallel and Distributed Systems (ICPADS), 2009 15th International Conference on 8-11 Dec. 2009; pp.26-33.
[93] Nawaz, Z.; Al-Ars, Z.; Bertels, K.; Shabbir, M., "Acceleration of Smith-Waterman using Recursive Variable Expansion," Digital System Design Architectures, Methods and Tools, 11th EUROMICRO Conference on 3-5 Sept. 2008, pp.915-922.
[94] Sadi, M.S.; Sami, A.Z.M.; Ahmed, I.U.; Ruhunnabi, A.; Das, N., "Bioinformatics: Implementation of a proposed upgraded Smith-Waterman algorithm for local alignment," Computational Intelligence in Bioinformatics and Computational Biology, IEEE Symposium on March 30 2009-April 2 2009; pp.87-91.
[95] Khairudin, N.; Mahmod, N.; Halim, A.K.A.; Junid, S. A M A; Idros, M. F M; Hassan, S. L M; Majid, Z.A., "Design and Analysis of High Performance Matrix Filling for DNA Sequence Alignment Accelerator Using ASIC Design Flow," Computer Modeling and Simulation (EMS), Fourth UKSim European Symposium on 17-19 Nov. 2010; pp.108-114.
[96] Gardner-Stephen, P.; Knowles, G., "DASH: localising dynamic programming for order of magnitude faster, accurate sequence alignment," Computational Systems Bioinformatics Conference Proceeding IEEE on 16-19 Aug. 2004; pp.732-735.
[97] Flynn and Laurie J.: Intel Halts Development of 2 New Microprocessors. The New York Times, 2004.
[98] Moore's Law to roll on for another decade, http://news.cnet.com/2100-1001984051.html. Retrieved 2012-2-10.
[99] Amdahl G.M.: The validity of the single processor approach to achieving large scale computing capabilities. In Proceedings of AFIPS Spring Joint Computer Conference, Atlantic City, N.J., AFIPS Press, pp.483–85.
[100] Gotoh O.: An improved algorithm for matching biological sequences. J Molecular Biology 1982, 162(3):705-708
[101] D.S. Hirschberg, "A Linear Space Algorithm for Computing Maximal Common Subsequences," Comm. ACM, vol. 18, no. 6, pp. 341-343, 1975.
[102] O. Gotoh, "An Improved Algorithm for Matching Biological Sequences," J. Molecular Biology, vol. 162, no. 3, pp. 705-708, Dec. 1982.
[103] Saitou M. and Nei N.: The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol, 1987, 4:406-425.

BIOGRAPHY

Ananth Prabhu G received the BE degree in computer science & engineering in 2007 from VTU, Belgaum and the MTech degree in computer science & engineering in 2010, from the Manipal University, Manipal. Currently, he is working towards the PhD degree in the department of computer science & engineering from VTU, Belgaum. He works as an assistant professor in the department of computer science & engineering in Sahyadri College of Engineering & Management, Mangalore. His research interests include high performance computing, parallel programming, GPGPUs, and load balancing.

Dr. Ganesh Aithal received the Ph.D. degree in computer science from the National Institute of Technology, Karnataka (NITK), India. Currently, he is the professor in the computer science department and vice principal in P.A. College of Engineering, Mangalore. He has 23 years of teaching experience and has published many papers in national and international level. His research interests include cryptography and parallel programming.