IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.12534

# NANOPORE TECHOLOGY AND ITS APPLICATIONS IN GENRATION SEQUENCING

**Meghana C.V[1], Chandini A.G[2]**

UG Student, Dept of ECE SJCIT, Chickballapura, India[1]

Assistant Professor, Dept of ECE, SJCIT, Chickballapura, India[2]

**Abstract:** Over the last few years, Nanopore sequencing has gained so much importance in science and in biomedical research.Oxford Nanopore Technology is one of the best application of nanopores.Oxford Nanopore Technology enables the identification of a broad range of analytes including DNA,RNA,micro RNA and proteins. In Oxford nanopore system the nanopore is inserted into an electrically resistant membrane created from synthetic polymers. A potential is applied across the membrane resulting in current which flows only through the aperture of nanopore. Single molecules that enter the nanopore causes disruption in the current, by measuring the disruption caused a molecule can be identified.

**Keywords:** nanopore, membrane, DNA, RNA,analytes.

## I. INTRODUCTION

For DNA sequencing,Oxford nanopore uses a strand sequencing method in which intact DNA strands are processed by nanopores and analysed in real time.

The nanopore sequences whatever fragments are presented toit regardless of their length rather than generating reads of a specific length.This is different from traditional cyclical sequencing chemistries that deliver a set of data at the end ofa fixed run time with fragments of a set length.

The DNA strandsto be sequenced are mixed with copies of aprecessive enzymes as the DNA enzyme complexes approached the nanopore. The single-stranded DNA is pulledthrough the aperture of the nanopore,the enzyme binds toa single-stranded leader at the end of the double-stranded DNAtemplate and unzips the double-strand feeding it through thenanopore. As the DNA moves through the pore the combination of nucleotides called a K-mer within the narrowest part of the pore-barrel creates a characteristic disruption in electrical current.This information can be used to determine the order of the basis on that DNA strand.The below figure shows the nanopore sequencing.

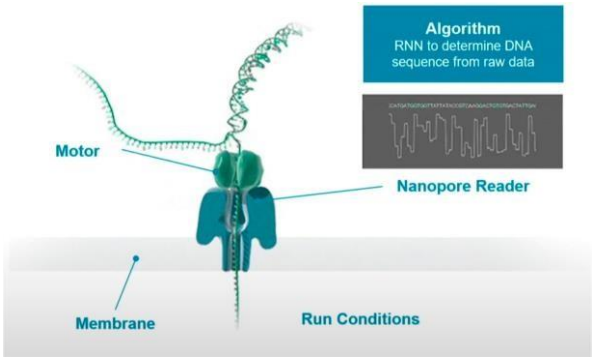



Figure 1.1:The process of nanopore sequencing.

The speed of the enzyme can be controlled.The faster it runsthe more data is yielded per second.

The strand sequencing method sequences an intact strand of DNA as it passes through the nanopore.Nanopores have processed read lengths of hundrerds of kilobases and when ananopore has processed a complete read it will start a new one.

There is no deterioration of accuracy as the DNA strandis sequenced.The system can read both strands of the double-

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.12534

stranded DNA in one continuous read. This given advantagesin data analysis and improves the accuracy of the sequence produced.

## II. RELATED WORK

Jianan Hui et al. [1] introduced a nanopore sequencing technology to investigate the nucleic acids and bio-macromolecules.The principle technology of signal processing can be used in the diagnosis of cancer and detection of microbes using nanopore sequencing technology.

David Rodriguez-Larrea et al.[2] published single-amino acid discrimination in proteins using homogeneous nanopore sensors where the proteins can be identified by training simple nural networks.A thermal protocol is used on the nanopore strand which examines proteins of the same strand which are aimed at reducing data variability that increases the generosity of the trained network which results in 100% correct assignement.The nanopores are engineered with homogeneity which is high and coupled with state of art of the iconic current.

Yunhao Wang et al.[3] made rapid advancement in nanoporetechnologies for sequencing where the DNA and RNA sequences are very lengthy which led to the substantial growth in the accuracy,read length and throughput.The bio-informics methods required extensive development to fully detect the long reads while investigating genomes.The nanopore can also be applied in genomes,transcript detectionand modification of base and detection.

Xixin Fu et al.[4] introduced the analysis and classification of nanopore data based multi-modality.DNA sequencing uses the single molecule analysis.The biomoleular recognition also follows the same molecular analysis.The analysis of DNA ,peptides,proteins and organic small molecules are realized using this analysis technology.The bio-marker analysis and DNA sequencing and oher various aspects use the same technology at present.

Christian Rohrandt et al.[5] published nanopore simulation where a raw data simulator is used for nanopore sequencing.

The second generation sequencing permits the determinationof longer DNA single molecules longer as 10000 molecules enabling nanopore DNA sequencing.The feature being unique has been

poised for further development of applications in novel biomedical which were previously unfeasible.

Basecaller: A wide range of basecaller algorithms are usedin the basecalling process of nanopore sequencing.Oxford nanopore provides open source research release

3. basecallers which implements new algorithms to improve accuracy of the reads.

## III. MOTIVATION

The DNA sequencing method is used to determine the nucleotides.There are four bases of nucleotides namely adenine,guanine,cytosine and thymine. A strand of DNA is made up of all the four bases.They provide the genotype telling a cell what to do and the phenotype.Each individual has a unique and specific base sequence.The contribution of proposed work is described as:

1. To detect the arrangement of nucleotides in thestrand of DNA and sequencing of fragments of all lengths.
2. The Oxford Nanopore Technology (ONT) to generate the microbial,animal,human and genomes of plant and complete the nanopore sequencing reads.
3. Real-time analysis of long DNA and RNA fragments which can be done directly by naopore sequencing technology.

## IV. PROBLEM STATEMENT

The traditional methods were capable of reading and sequencing the fragents which are of only short lengths of DNA which again has to be reassembled.The difficuly to read long sequences of DNA has appeared as the methods cannot afford to read long data sequences. The short sequencing couldn't deliver enough overlap between the DNA fragments.The old nanopore sequencing is limited only to the length of RNA/DNA fragments presented to the pore.

## V. PROPOSED SYSTEM

Nanopore technology is a next-generation sequencing(NGS)third generation approach is the most promising technology in genomic forward in future.The sequencing provides cost effective genotyping and testing.The nanopore sequencing provides real-time processing of samples of longer reads.

The process of nanopore sequencing consists of three parts namely:

1. Library preparation: In the nanopore sequencing library preparation is very much important in the process of sequencing.The fragments of DNA should be repaired if they are sheared or not.The DNA protein complex along with helicase or polymerase acts as a repaired connector during repairing.
2. Sequencing process:The copies of processive enzyme is mixed with the DNA fragment to be sequenced.The

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.12534

enzyme binds with single stranded double end of the fragment.The current produced while the process can be used to detect the sequence.

Capturing the signal:

During the process of RNA and DNA sequencing,the change in current knows as disruption caused by the movement of a single molecule passing through nanopore which is recorded by MinKNOW$^{TM}$ a software which executes all the programs of oxford sequencing devices.“The squiggle” a continuous change in current caused due tothe pore when bases pass through it.MinKNOW sequences the reads in realtime where each read refers to a strand of DNA/RNA.

Flow Cells:

The nanopore sequencing devices of oxford nanopore reckon on flow cells to undertake the process of sequencing.These cells are compatible,pocket-sized and portable with MinION.The reports shows that five MinION can generate upto 250 Gb of data.The adapter for the flow cells which enables real-time sequencing,direct and also provides single use Flonge cells.

Nanocall Basecalling program:

Nanocall is a program where the algorithms are written using c language while the metrichor was still in service.Nanocalls can be operated offline on the other hand metrichor must be connected to internet.To detect bases more accurately nanocall uses Expectation Maximum(EM) algorithm for several rounds.

SACall basecalling program:

SAcall is a CTC decoder and an end-to-end basecaller program which contains self-transformation layer and a convolution layer.A beam search algorithm is used by the decoder to sequence and read the bases.The consensus accuracy,assembly quality are far more better than the other two programs.

Applications of Nanopore technology:

Clinical Research-to identify and phase genetic variants and fusion transcripts.Provide insights about human health ,immunology and neuroscience.

Detection of microbes-It helps researchers in the field of sequencing to classify and monitor the microbes.

Assemble genomes-sequencing can taken for long reads which overcomes the difficulty of short read lengths.

Environmental genomics-The portability and affordability makes the nanopore sequencing to be carried out either in fields or in lab.

## V. RESULTS AND DISCUSSIONS

Nanopore sequencing technology provides a new way for identification of genomes and single-molecule proteins such as DNA/RNA.The technology is having tremendous impact on gene sequencing.It plays a vital role in diagnosing human cancer and immunal diseases.As it is called as the fourth- generation sequencing technology also referred to next generation approach will take over the other approaches in future.

The below figures 5(a) and 5(b) shows the sequencing readsof a DNA/RNA strands.

## VI . CONCLUSION

Nanopore sequencing has enabled bio-medical researches and many bio-medical studies as it provides reads for ultralong sequences.Nanopore sequencing provides results in real-time and direct sequences are allowed.Overcomingall the challenges with the traditional methods of sequencingnanopore technology has made a breakthrough into the future.It is also called as Next Generation Sequencing as itis the fourth generation of the development.The main concern in all the methods is the error rate which is about 6- 15% in R9.0 nanopore which is much greater than the illumine sequencing which is about 0.1-1%.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

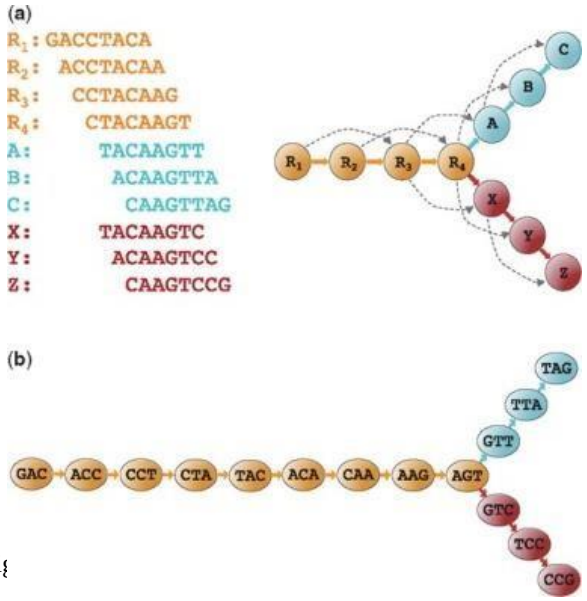
International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.12534

The accuracy can be further improved in future by optimizing the transcription reads.Improved accuracy would help in single molecule omics studies.Haplo-type resolved genome assembly can be achieved by ONT method using MinION.The ONT system will get benefit in end-to-endsystem sequencing.

More robust and bio-informatics user friendly,cloud storage and computing,real time-analysis are the key points in the nanopore sequencing.




Fig

In the field of life science,nanopore sequencing plays a major role in genomics,epigenetics research,detecting microbes relationship to human phenotypes and other fields.Oxford nanopore technologies(ONT) developed a direct sequencing technology which can sequence the native RNA molecules,providing full-length RNA reads in its native context.The library preparation for direct sequencing was made available in 2017 only a modest number of users were using it.To overcome the difficulties a new tool was introduced known as Nextflow Workflows.

Due to the adoption of deep learning algorithms ,thenanopore read basecalling has grown significantly over thelast few years.The most commonly used strategies forbasecalling are convolution neural networks(CNN) andrecurrent neural networks(RNN).By using both strategiesthe accuracy of basecall reads has improved.The Graphicprocessing units (GPUs) provide greater acceleration incertain computational tasks which allows 50-500 reads persecond.By using GPU process while nanopore sequencingwhen in time will be cost-efficient as well as time-efficient. The companies that use nanopore sequencing are as followsPacBio,ONT,Illumina,Agilent,ThermoFisherScientific,Qiag en and so on.

## REFERENCES

[1] Rodriguez-Larrea, D. Single-aminoacid discrimination inproteins with homogeneous nanopore sensors and neural networks. Biosens. Bioelectron. 2021, 180, 7.
[2] Fang, C.; Zhong, H.; Lin, Y.; Chen, B.; Han, M.; Ren, H.; Lu, H.; Luber, J.M.; Xia, M.; Li, W.; et al. Assessment of the cPAS- based BGISEQ-500 platform for metagenomic sequencing. Gigascience 2018, 7, 1–8.
[3] Miga, K.H.; Koren, S.; Rhie, A.; Vollger, M.R.; Gershman, A.; Bzikadze, A.; Brooks, S.; Howe, E.; Porubsky, D.; Logsdon, G.A.; et al. Telomere-to-telomere assembly of a complete human X chromosome. Nature 2020, 585, 79–84.
[4] Sun, X.; Song, L.; Yang, W.; Zhang, L.; Liu, M.; Li, X.; Tian, G.; Wang, W. Nanopore sequencing and its clinical applications. Methods Mol. Biol. 2020, 2204, 13–32.
[5] Ameur, A.; Kloosterman, W.P.; Hestand, M.S. Single- Molecule Sequencing: Towards Clinical Applications. Trends Biotechnol. 2019, 37, 72–85.
[6] Logsdon, G.A.; Vollger, M.R.; Eichler, E.E. Long-read human genome sequencing and its applications. Nat. Rev. Genet. 2020,21, 597–614.
[7] Fried, J.P.; Swett, J.L.; Nadappuram, B.P.; Mol, J.A.; Edel, J.B.; Ivanov, A.P.; Yates, J.R. In situ solid-state nanopore fabrication. Chem. Soc. Rev. 2021, 50, 4974–4992.
[8] Ma, Q.; Si, Z.X.; Li, Y.; Wang, D.G.; Wu, X.L.; Gao, P.C.; Xia, F. Functional solid-state nanochannels for biochemicalsensing. Trac. Trends Anal. Chem. 2019, 115, 174–186.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.12534

[9] Brown, E.; Freimanis, G.; Shaw, A.E.; Horton, D.L.; Gubbins, S.; King, D. Characterising foot-and-mouth disease virus in clinicalsamples using nanopore sequencing. Front. Vet. Sci. 2021, 8, 10.
[10] Hu, Z.L.; Huo, M.Z.; Ying, Y.L.; Long, Y.T. Biological nanopore approach for single-molecule protein sequencing. Angew. Chem. Edit. 2020, 133.
[11] Oxford Nanopore Technologies. Introduction to Real Time Analysis.2020. Available online: https://www.youtube.com/watch?v=8oNEjt5Ov_Q.
[12] Wick, R.R.; Judd, L.M.; Holt, K.E. Performance of neural networks tools for Oxford Nanopore sequencing. Genome Biol. 2019.
[13] [1] Rang, F.J.; Kloosterman, W.P.; de Ridder, J. From squiggle to basepair: Computational approaches for improving nanopore sequencing read accuracy. Genome Biol. 2018, 19, 11.
[14] [2] Karst, S.M.; Ziels, R.M.; Kirkegaard, R.H.; Sørensen, E.A.; McDonald, D.; Zhu, Q.,High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. Nat. Methods 2021, 18, 165–169.
[15] [3] van Dijk, E.L.; Jaszczyszyn, Y.; Naquin, D.; Thermes, C. The third revolution in sequencing technology. Trends Genet. 2018, 34, 666–681.
[16] [4] Meng, J.; Victor, B.; He, Z.; Liu, H.; Jiang, T. Detecting somatic small variants in paired tumor and normal sequencing data with convolutional neural network. Bioinform. 2020.
[17] [5] Abiodun, O.I.; Jantan, A.; Omolara, A.E.; Dada, K.V.; Mohamed, N.A.; Arshad, H. State-of-the-art in artificial neural network applications: A survey. Heliyon 2018, 4, e00938.
[18] [6] Huang, N.; Nie, F.; Ni, P.; Luo, F.; Wang, J. SACall: A neural network for Oxford Nanopore sequencing data based on self- attention mechanism. IEEE/ACM Trans. Comput. Biol. Bioinform. 2020.
[19] [7] Teng, H.; Hall, M.B.; Duarte, T.; Wang, S.; Coin, L.J.M. Chiron: Translating nanopore raw signal directly into nucleotide sequence using deep learning. Gigascience 2018, 7.
[20] [8] Chan, W.S.; Chan, T.L.; Au, C.H.; Leung, C.P.; To, M.Y.; Ng, M.K.; Ma, E.S.K.; Tang, B.S.F. An economical Nanopore sequencing assay for human papillomavirus (HPV) genotyping. Diagn. Pathol. 2020, 15, 18.
[21] [9] Ren, Z.L.; Zhang, J.R.; Zhang, X.M.; Liu, X.; Lin, Y.F.; Bai, H.; Wang, M.C.; Li, P.; et al. Forensic nanopore sequencing of STRs and SNPs using Verogen's ForenSeq DNA Signature Prep Kit and MinION. Int. J. Legal Med. 2021, 9.
[22] [10] Wu, X.W.; Luo, H., Wiedmann, M.; Baker, R.C.; Stevenson, A.; Zhang, G.T.; et al. Evaluation of salmonella serotype prediction with multiplex nanopore sequencing. Front. Microbiol. 2021, 12, 13.
[23] [11] Williams, M.S.; Basma, N.J.; Amaral, F.M.R.; Williams, G.; Weightman, J.P, T.C.P. Targeted nanopore sequencing for the identification of ABCB1 translocations in cancer. BMC Cancer 2020, 20, 9.
[24] [24 Tan, S.Y., Dvorak, M.P. Characterization of Emerging swine viral diseases through oxford nanopore sequencing using senecavirus A as a model. Viruses Basel. 2020, 12, 1136.
[25] [25] Brandt,C., Bongcam-Rudloff, E., Muller, B. Abundance tracking by long-read nanopore sequencing of complex microbial communities in samples. Appl. Sci. Basel 2020, 10, 7518.