



Real-time, Portable and Lightweight Nanopore DNA Sequence Analysis using System-on-Chip

Hasindu Gamaarachchi^{1,2}, Martin A. Smith^{2,3}, Sri Parameswaran¹

¹ School of Computer Science and Engineering, UNSW, Sydney

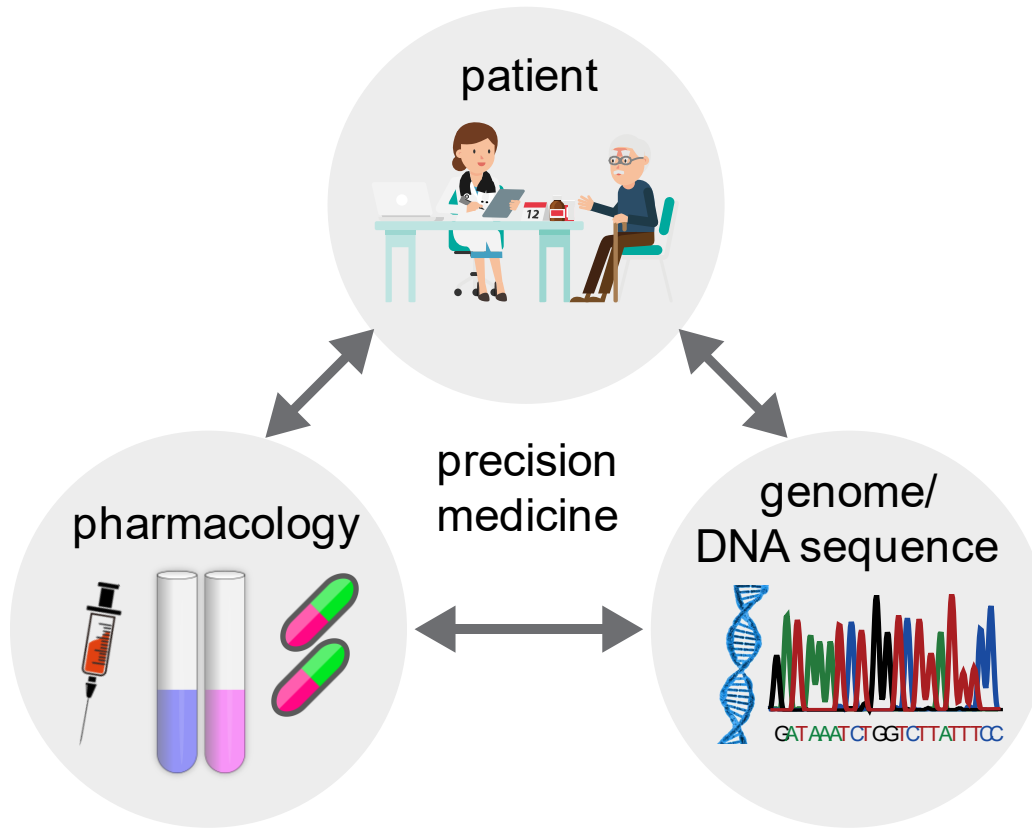
² Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney

³ St-Vincent's Clinical School, UNSW, Sydney

hasindu@unsw.edu.au

Introduction and Motivation

Precision (personalised) Medicine



**DNA sequence analysis is the key
to precision medicine**

Evolution of DNA Sequencing & Analysis

**DNA
sequencing**

year 2000



year 2019



DNA sequencers
have become
ultra portable

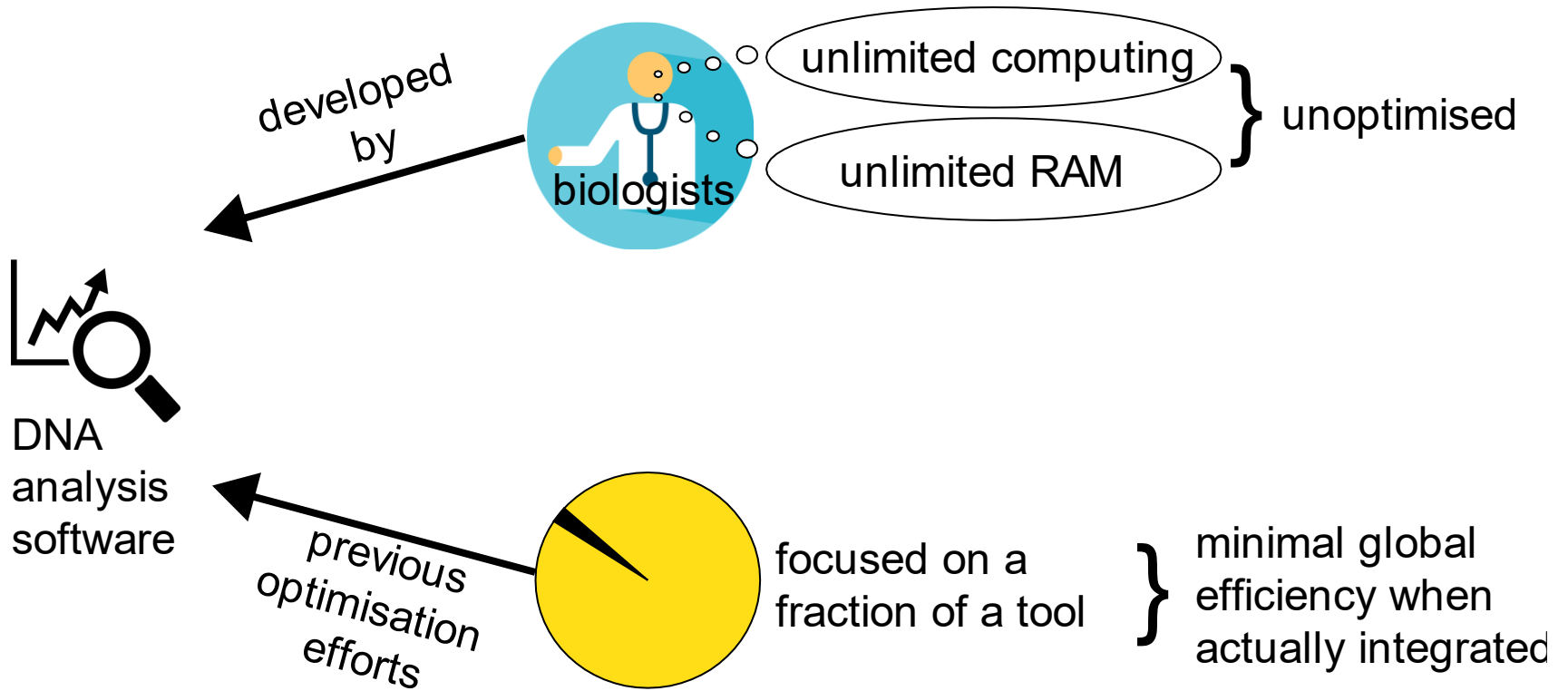


**DNA
analysis**



Data analysis is
still on done HPC

Why?

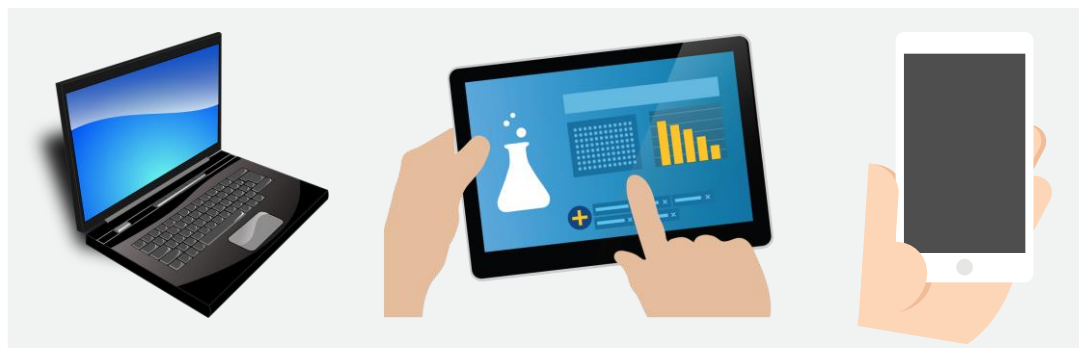


We, for the First Time:

- optimise a complete Nanopore DNA analysis workflow
- to execute on portable and lightweight embedded systems



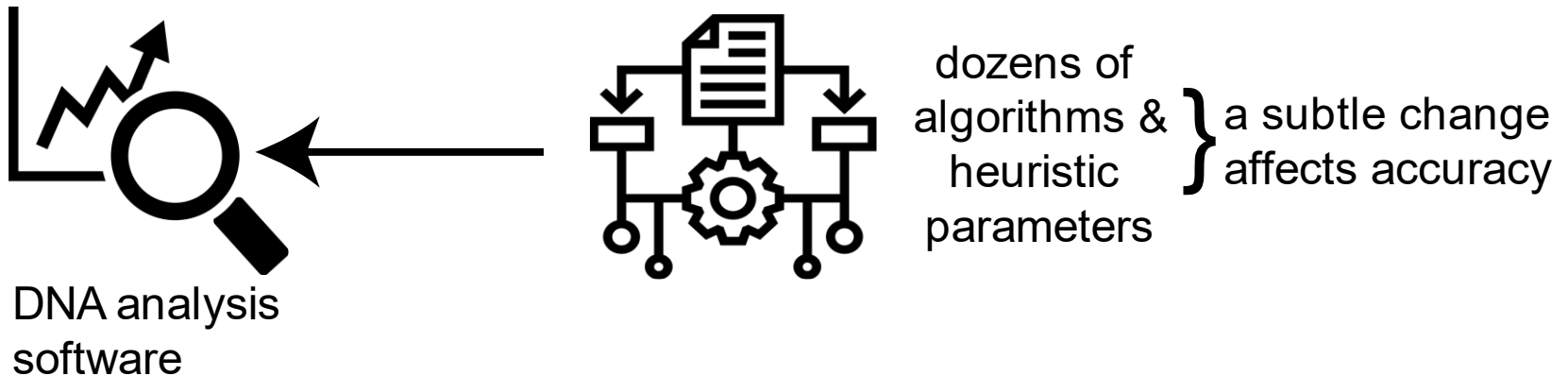
terabytes of data



Challenges

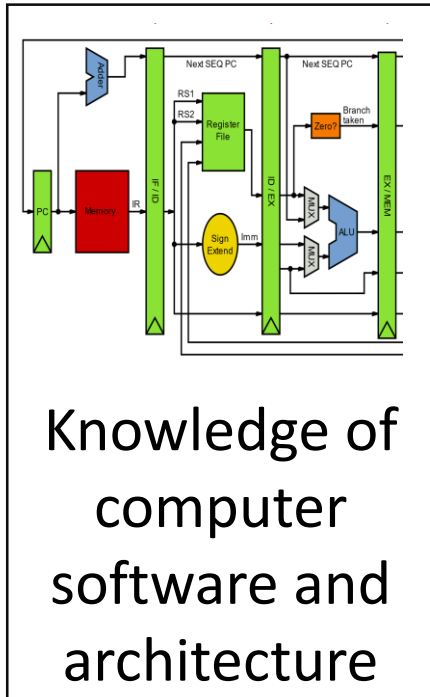
DNA analysis tools are extremely complex.

- dozens of various algorithms
- numerous heuristically determined parameters

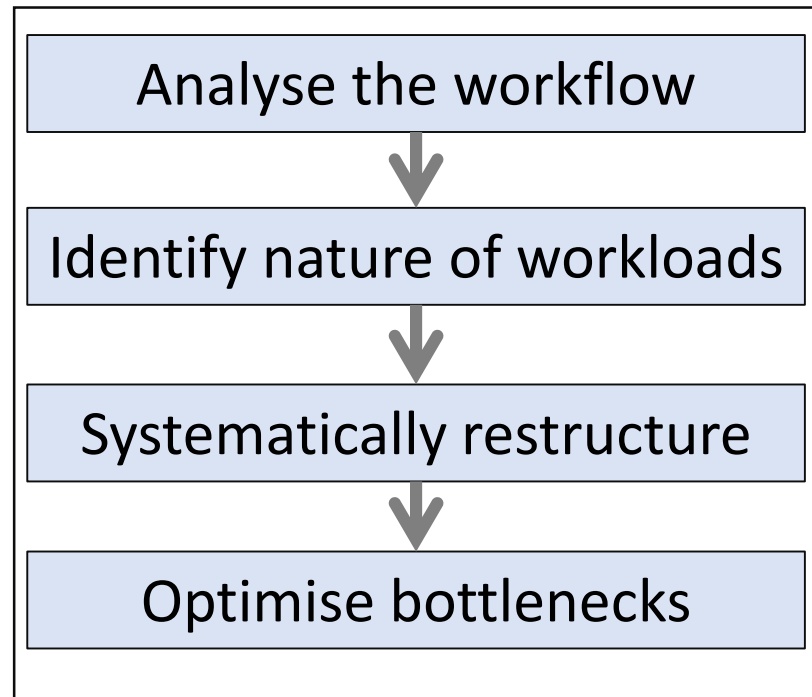


Methodology and Immediate Results

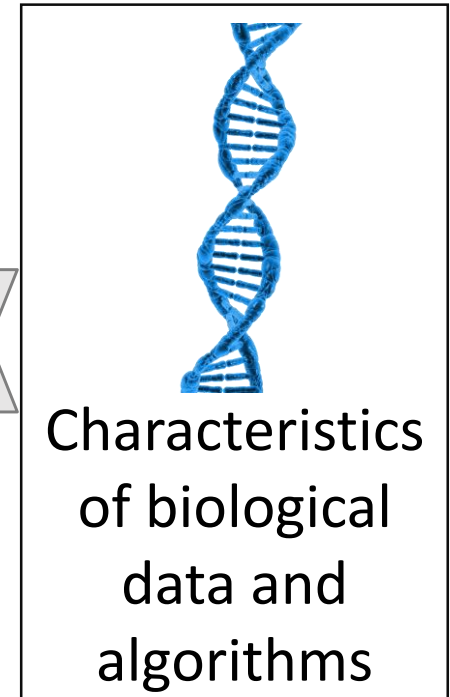
Simplified Overview



Efficiently leverage resources

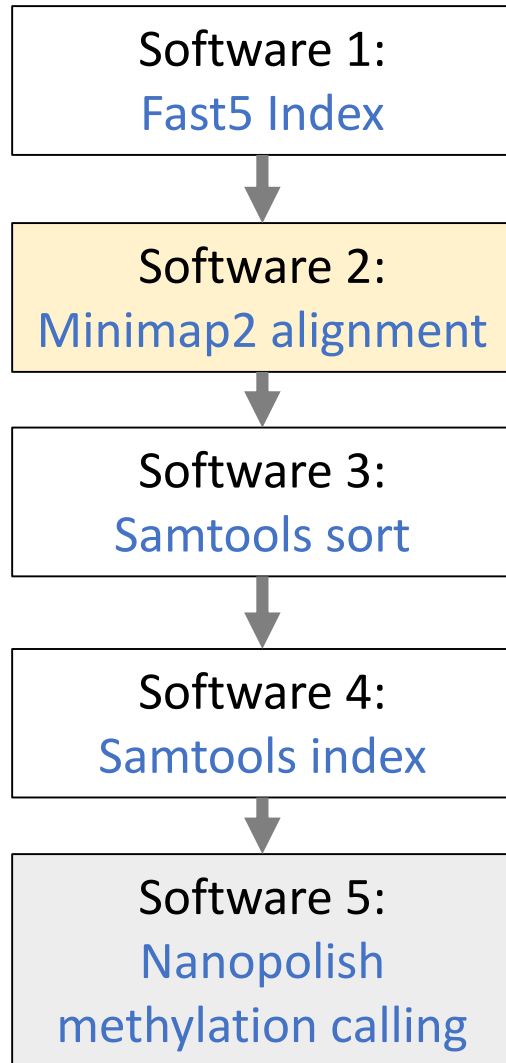


Execute efficiently on embedded systems

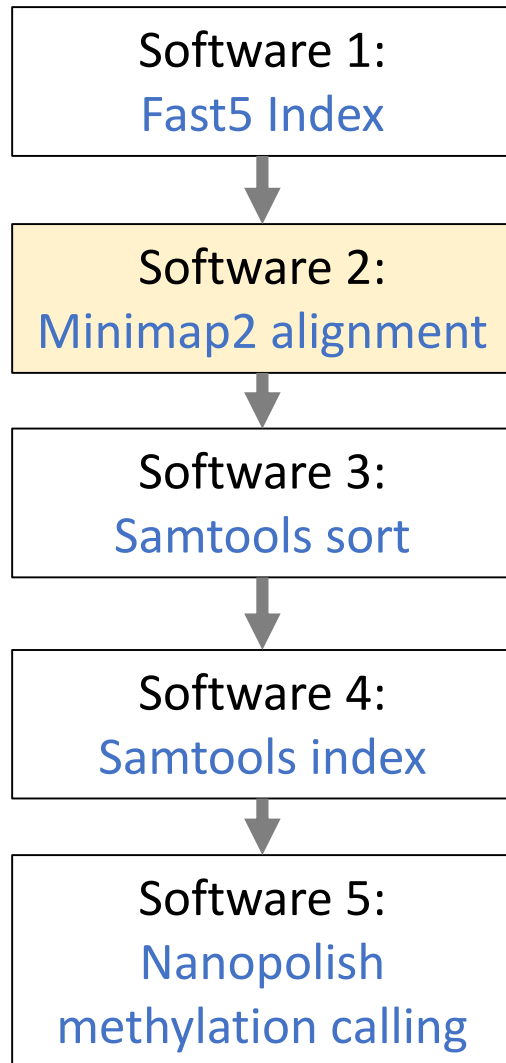


No impact on accuracy

Applied to a Popular Nanopore DNA Sequence Analysis Workflow

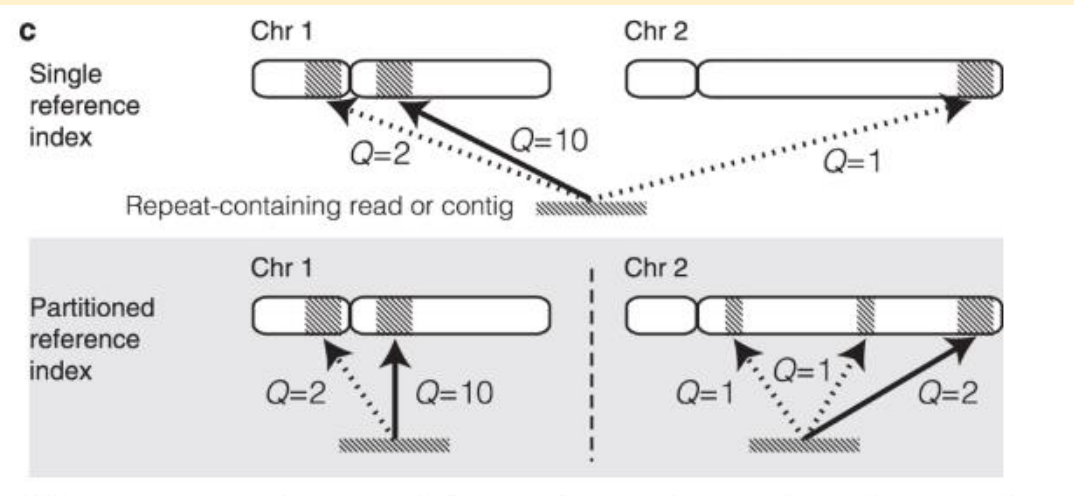


Applied to a Popular Nanopore DNA Sequence Analysis Workflow



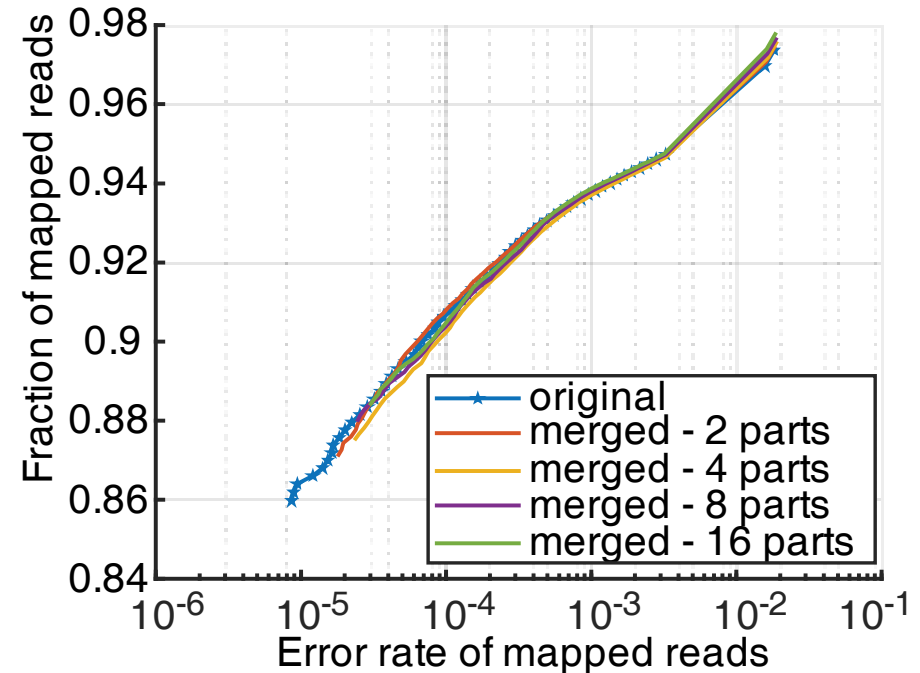
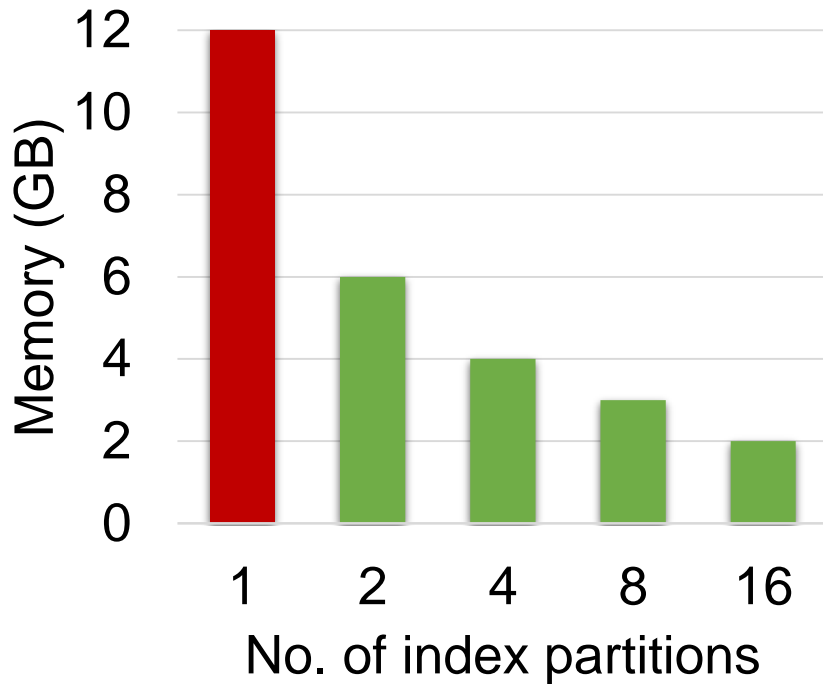
Memory Intensive

- divide and conquer strategy to reduce peak RAM



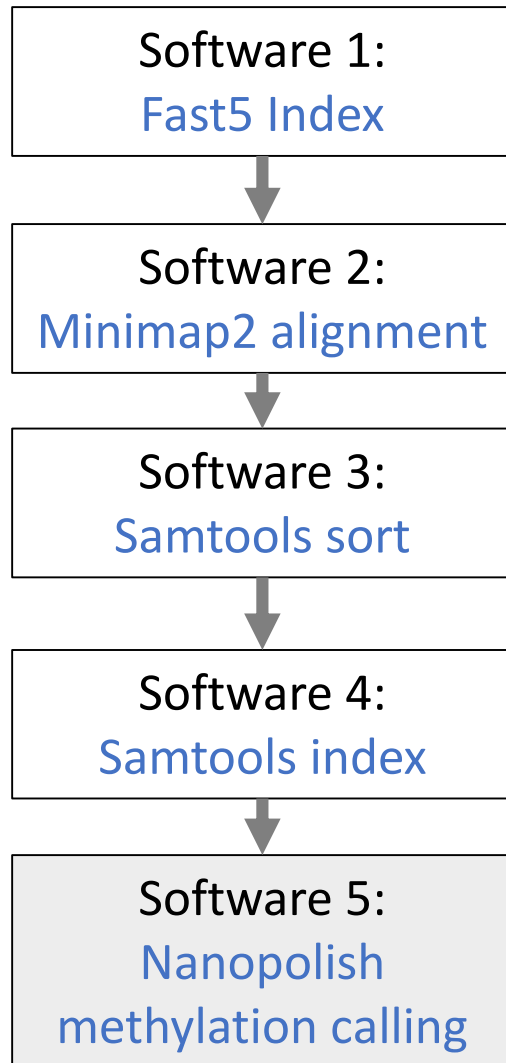
- proposed merging strategy for retaining accuracy

Optimised Minimap2



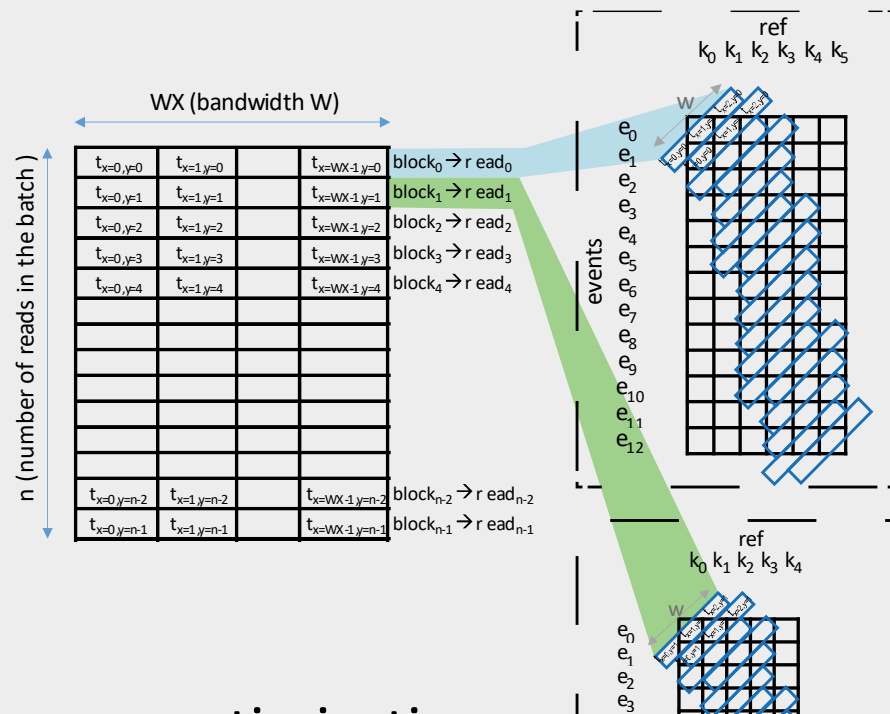
Hasindu Gamaarachchi, Sri Parameswaran & Martin A. Smith,
Featherweight long read alignment using partitioned reference indexes,
<https://www.nature.com/articles/s41598-019-40739-8>

Applied to a Popular Nanopore DNA Sequence Analysis Workflow



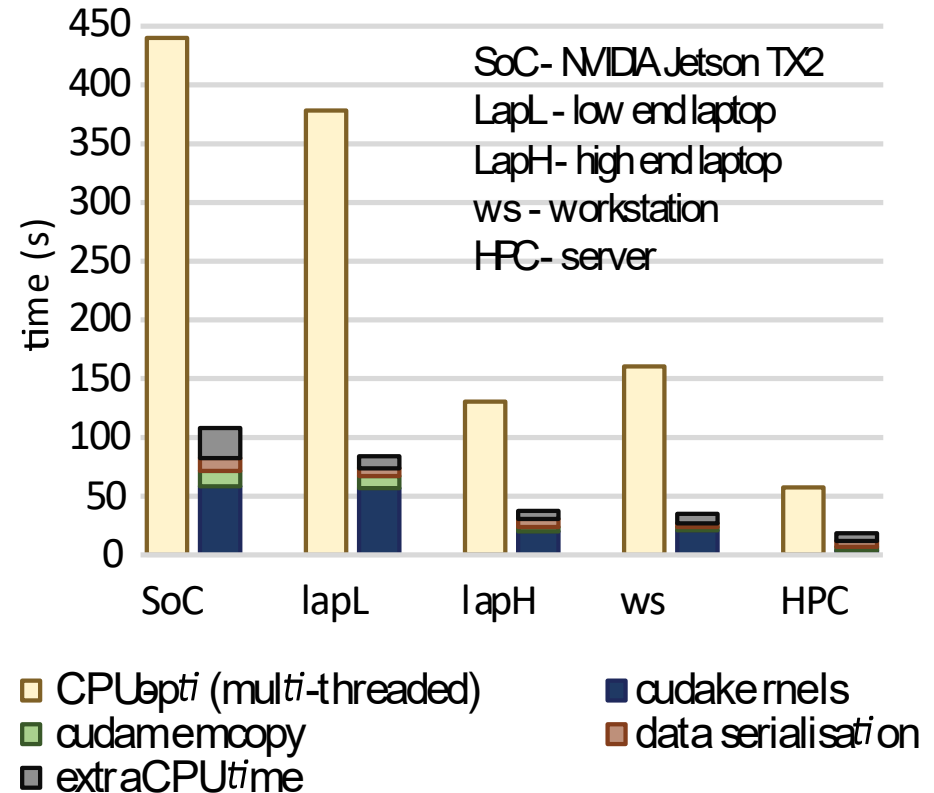
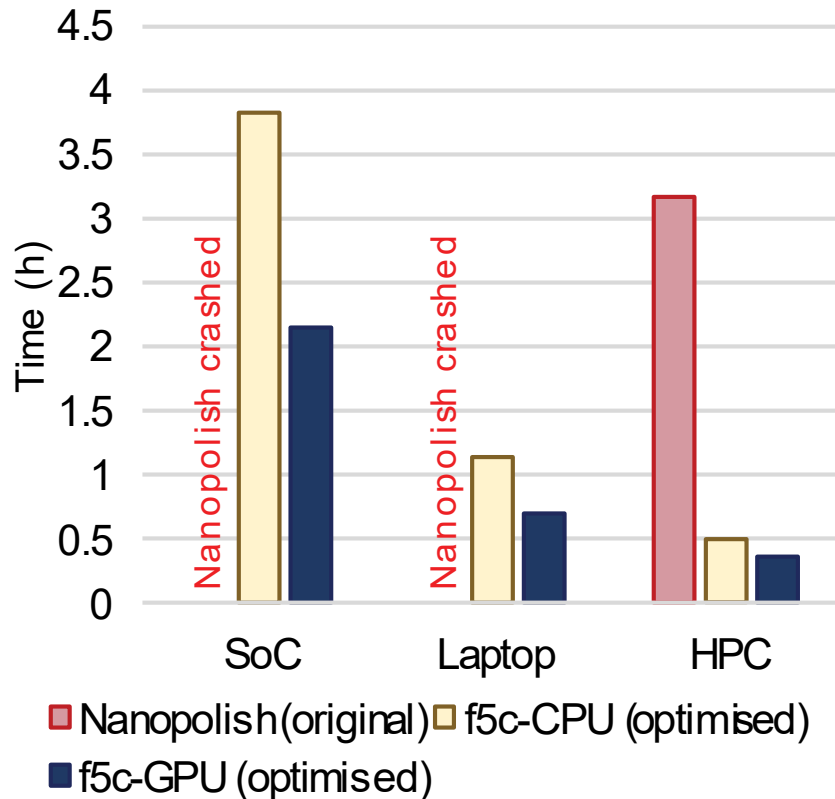
Compute intensive component

- Parallelisation for the GPU



- Memory optimisations
- Work scheduling between CPU-GPU

Optimised Nanopolish



Hasindu Gamaarachchi Et al.,

GPU Accelerated Adaptive Banded Event Alignment for Rapid Comparative Nanopore Signal Analysis,

<https://www.biorxiv.org/content/10.1101/756122v1>

Open Source

Available at :

<https://github.com/hasindu2008/f5c/>

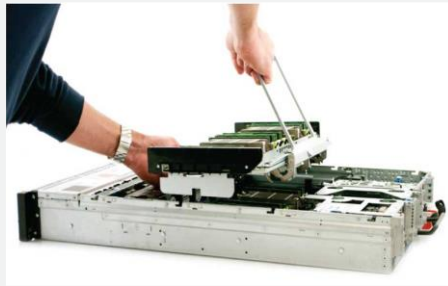
Overall Results and Outcome

DNA Analysis on a SoC is Now Possible!!

HPC

>20,000\$

>20 kg



Before our optimisations



After our optimisations



~9X faster than before

SoC

<200\$

<200 g



due to high peak memory



now possible!!

HPC: eg, server blade: 20 core Xeon, 256 GB RAM

SoC: eg, NVIDIA Jetson Nano: Quad core ARM, 4GB RAM+GPU memory, 128 CUDA cores

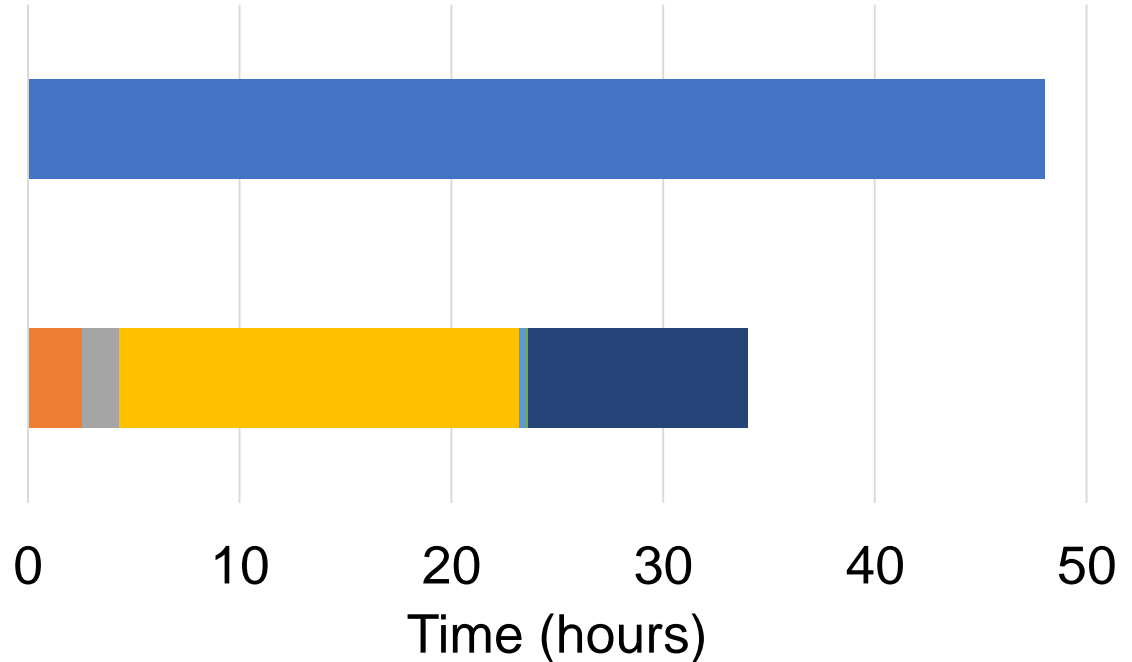
Real-time Processing Capability



Sequencing on MinION (data generation)



Analysis on Jetson Nano SoC



- MinION sequencing run
- Fast5 index
- Samtools sort
- Nanopolish methylation calling

- File transfer
- Minimap2 alignment
- Samtoool index

An Era of Fully Portable Genomics



Portable Nanopore
MinION sequencer



Real-time, portable
and lightweight SoC

Complete DNA analysis in-the-field

1. Remote locations without Internet
2. Point-of-care (hospital or clinic)

Questions?
