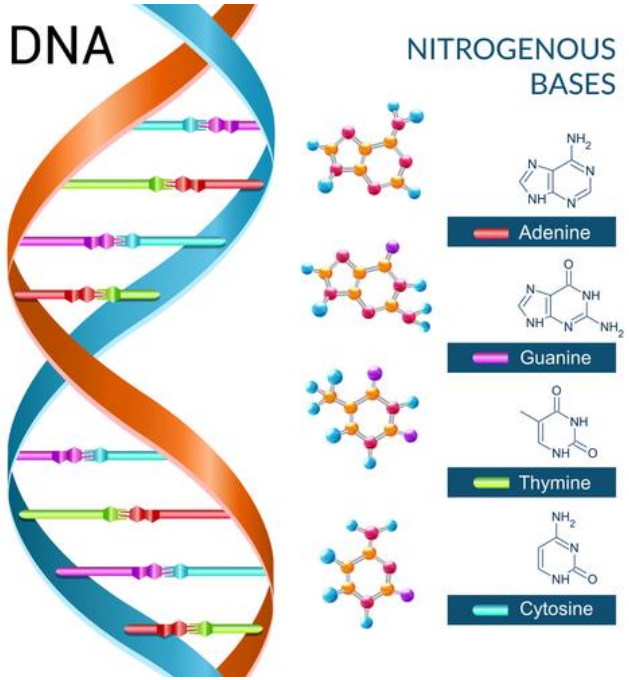
Robust RNA Tail Detection Techniques for High-Density DNA Data Storage Using Solid-State Nanopores

Rajat Chakraborty, Mingye Xiong & Jean-Pierre Leburton
Department of Electrical and Computer Engineering
University of Illinois Urbana-Champaign

2024 Frontera User Meeting

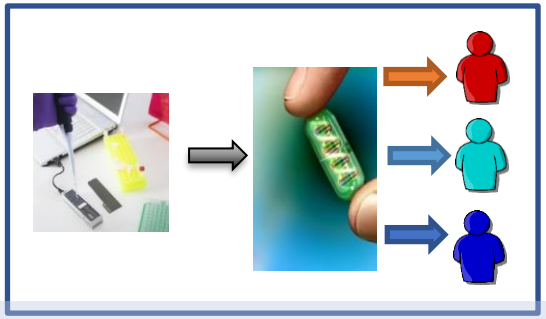
August 5, 2024

DNA Sequencing : Applications



Genome Sequence

…AGTCGTACTTTCCGATCG
ATCGAGTCGTACTTTCCGA
TCGATCGAGTCGTACTTTC
CGATCG…



Personalized Medicine

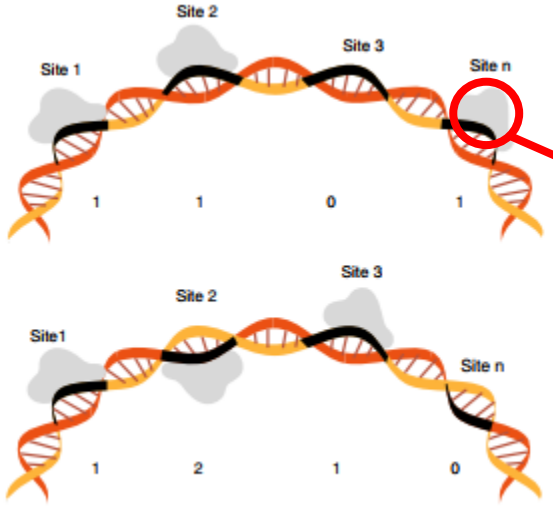
Pharmaceutical Research

Data Storage

DNA Data Storage

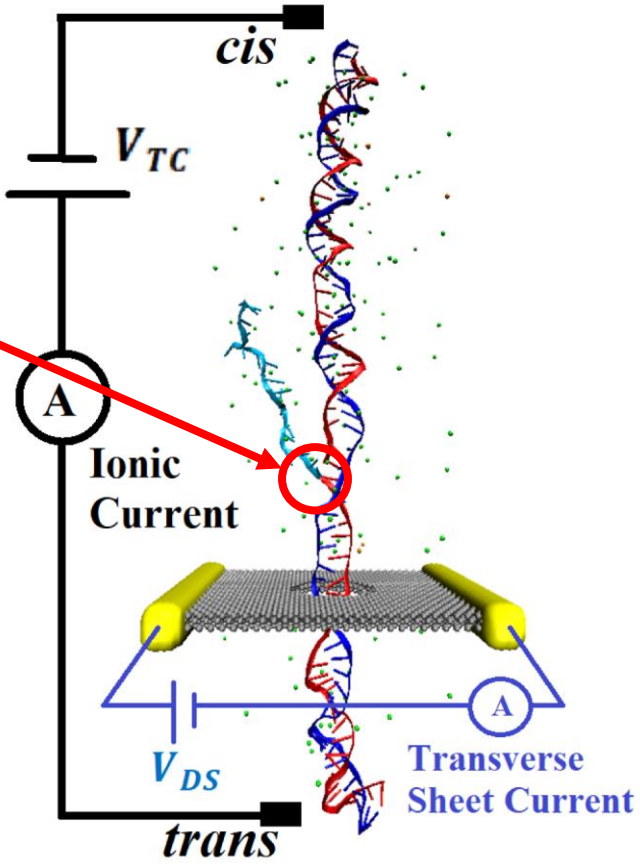
- Nature's way of storing data
- Encode using synthetic DNA, then decode by sequencing
- Using synthetic DNA is time consuming and labor intensive

Positional Encoding – DNA Nicks



S.K. Tabatabaei et. al., Nat Commun 11, 1742 (2020)

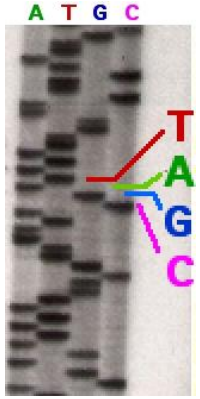
RNA Tails

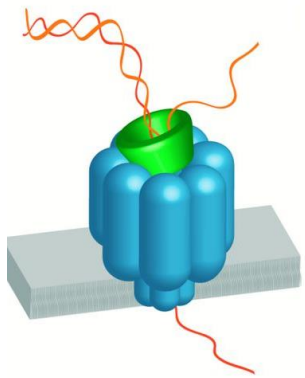


Nanopore Sequencing

Sanger Sequencing

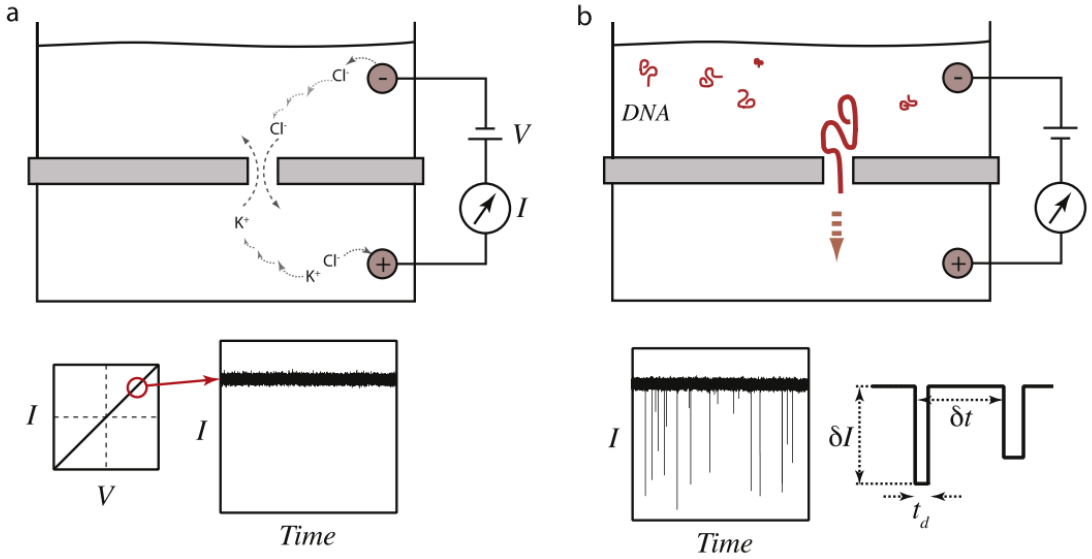
- High cost
- Intensive lab process
- Time consuming





Biological Nanopore

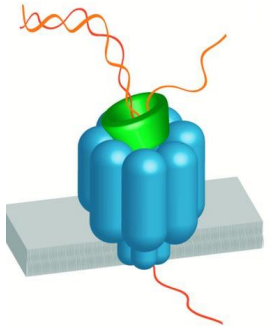
- Cost reduced
- Speed increased

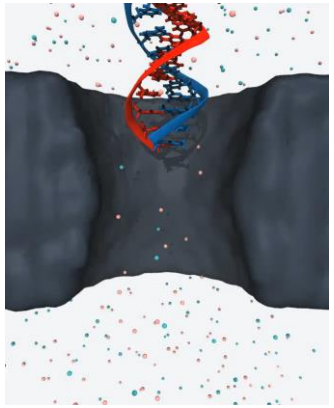


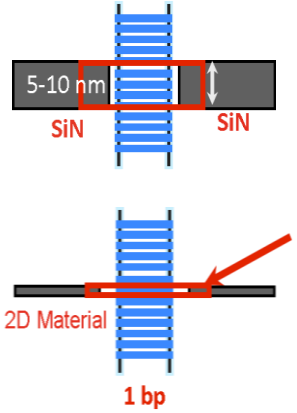
- Steady-state current unless a biomolecule passes through
- Information from magnitude and duration of current

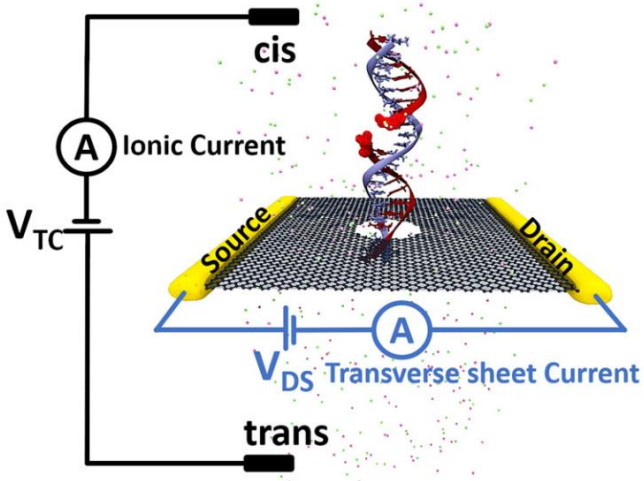
Wanunu, M., *Physics of life reviews,* 9, 125 (2012)

Solid-State Nanopores









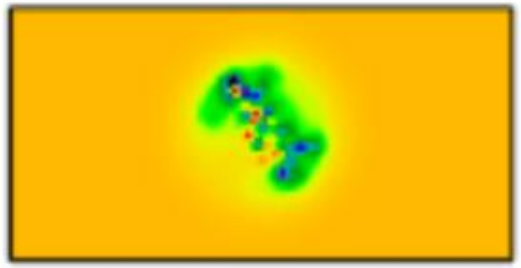
Biological Nanopore → **Solid State Nanopore** → **2D Material Nanopore**

- Cost reduced
- Constraints on exp. Conditions and size
- Limited stability

- Increased stability
- Less sensitive to conditions
- Tunable size
- Limited spatial resolution

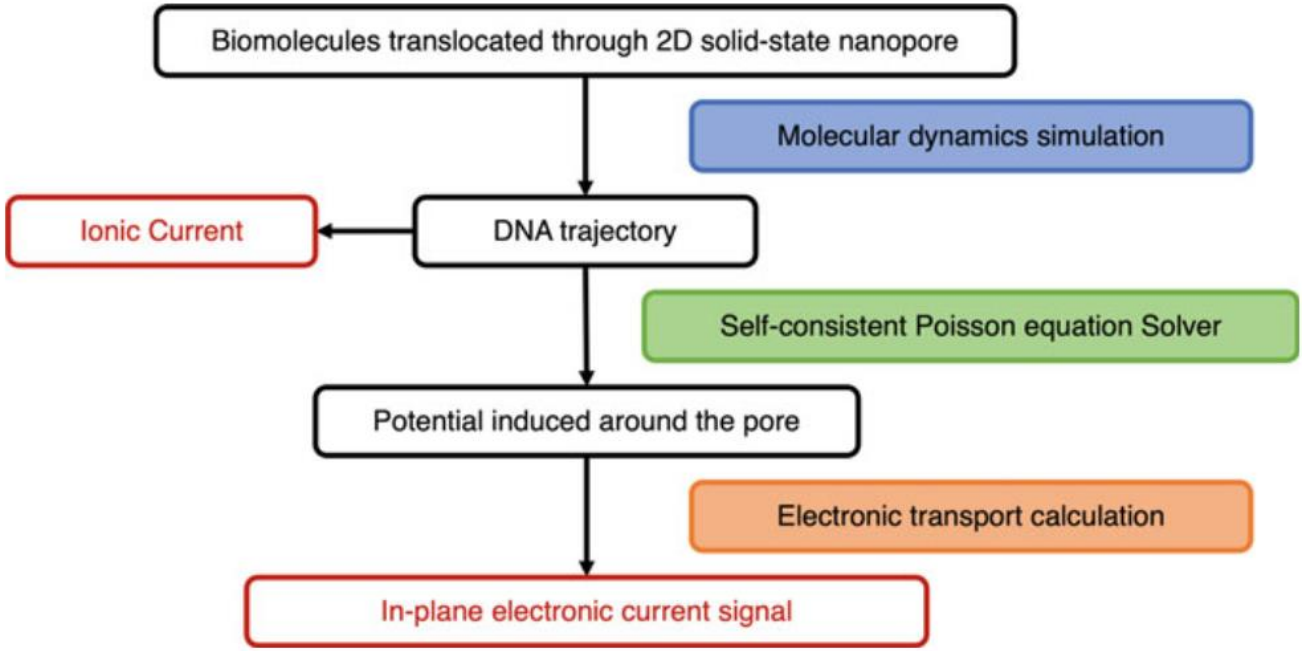
- Better spatial resolutions
- In-plane current

$$\nabla \cdot [\epsilon(\vec{r})\nabla\phi(\vec{r})] = -e[K^+(\vec{r},\phi) - Cl^-(\vec{r},\phi)] - \rho_{DNA}(\vec{r})$$



- Better resolution
- Less noise

Computational Methods



- **Molecular dynamics** - numerical method for studying many-body systems
- Powerful computational tool to investigate interactions at an atomic scale, extensively used to study biophysical systems
- Interaction energy and force calculated to obtain a trajectory in a system
- **Poisson-Boltzmann solver** – to obtain induced potential
- **Electronic transport calculation** – in-plane signal

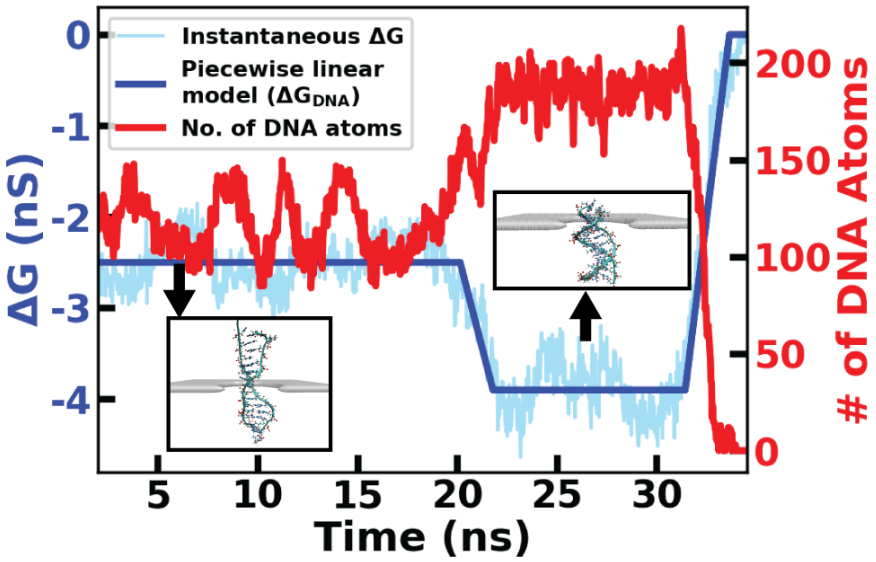
Detection of RNA Tails

Three Different Tail Lengths

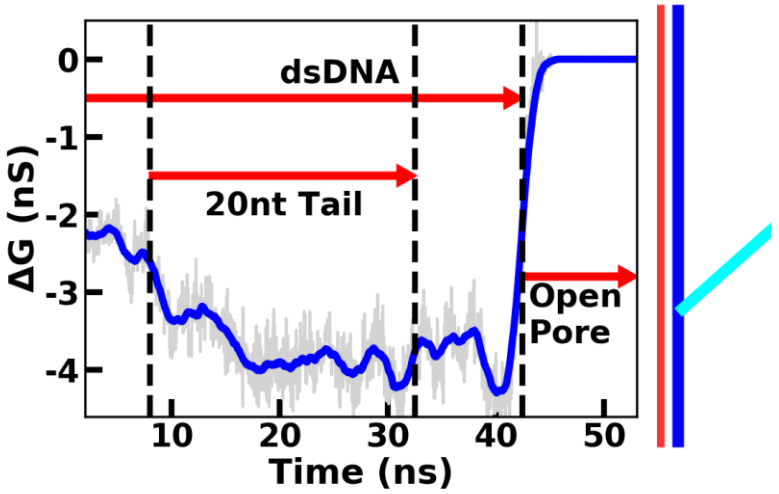
- 10 nt, 15 nt, 20 nt
- 10 bp separation
- 70 bp random sequence of dsDNA

3nm diameter Pore

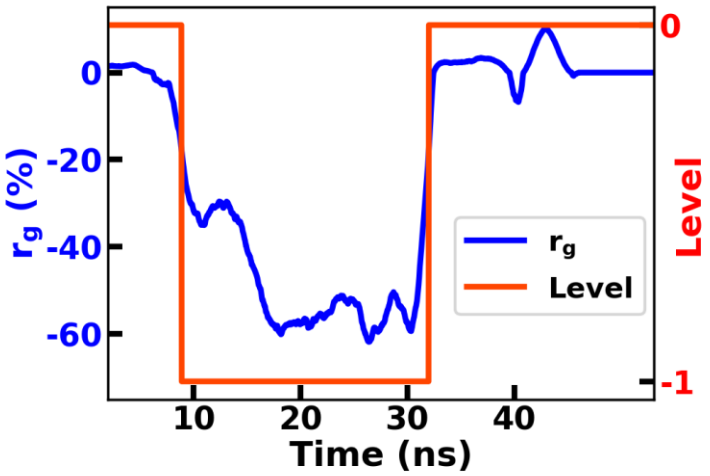
- Allow dsDNA with overlapping tails to go through
- Reduce lateral movement



$$\Delta G = G - G_{open} < 0$$



$$r_g = \frac{\Delta G - \Delta G_{DNA}}{|\Delta G_{DNA-def}|} \times 100\%$$

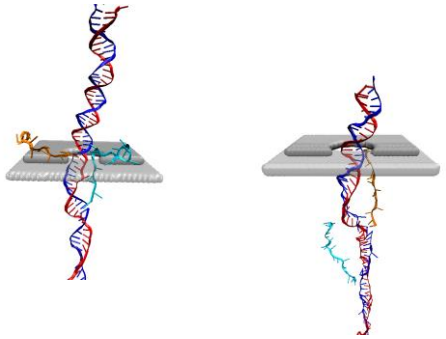


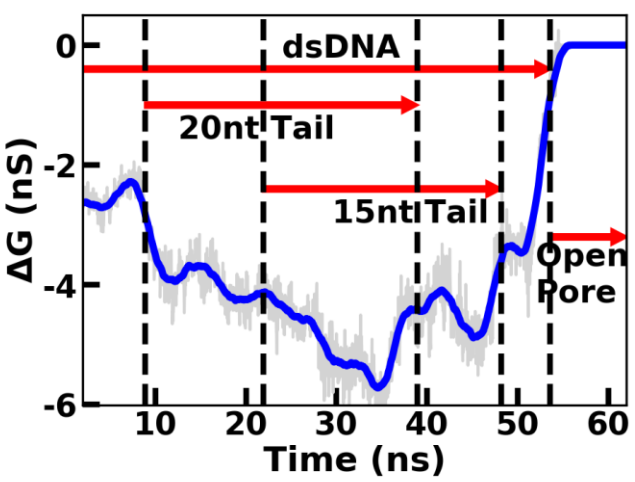
Level -1 (-20%)

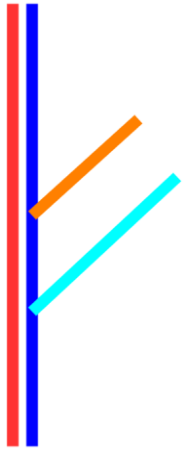
R. Chakraborty et. al., ACS Appl. Nano Mater. 2023, 6, 6, 4651–4660

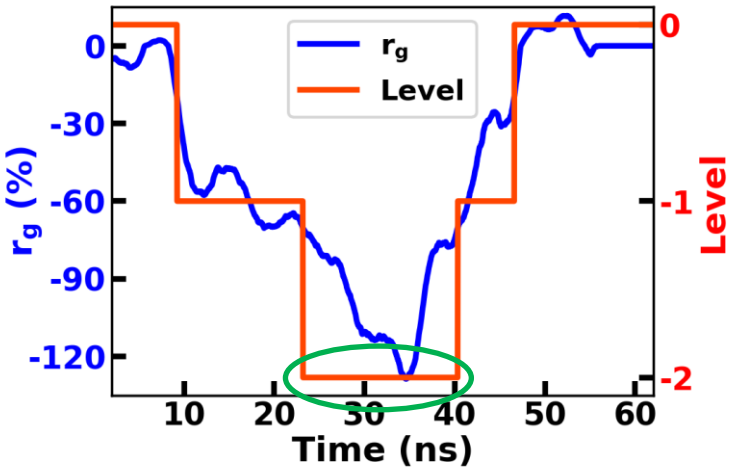
Detection of Overlapping Tails

Partial Overlapping
Tails : 20 nt & 15 nt
Separation :10 bp



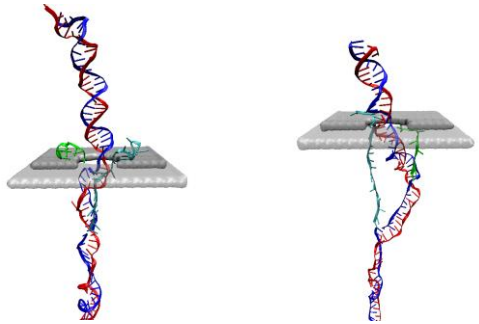


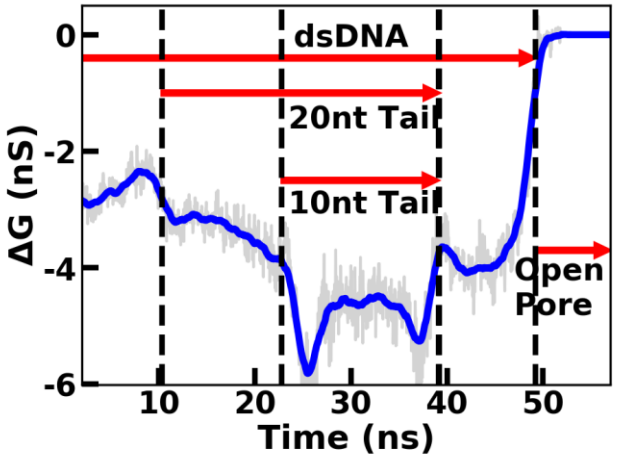


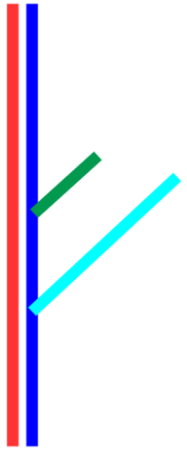


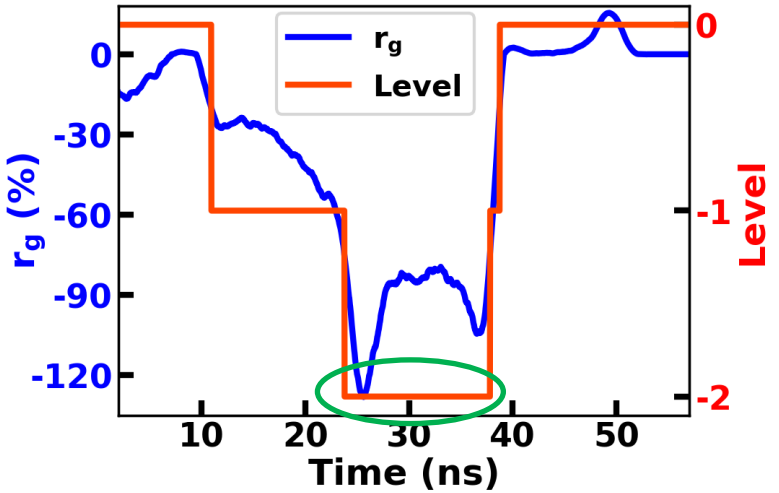
Introduce a new level -2 (-70%)

Complete Overlapping
Tails : 20 nt & 10 nt
Separation :10 bp

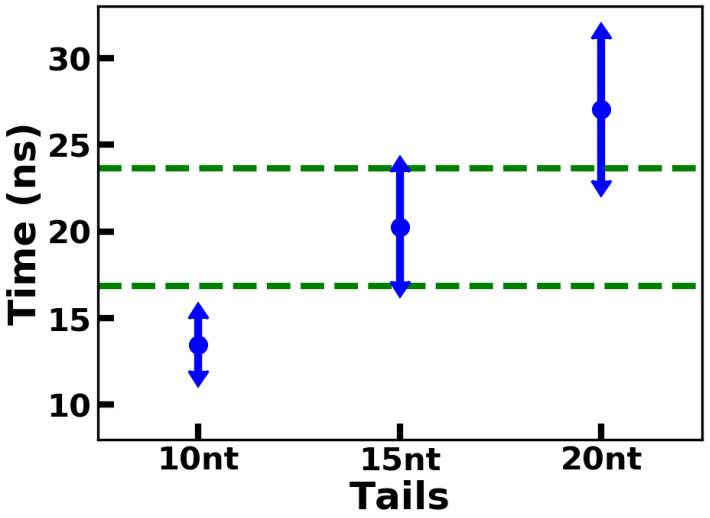


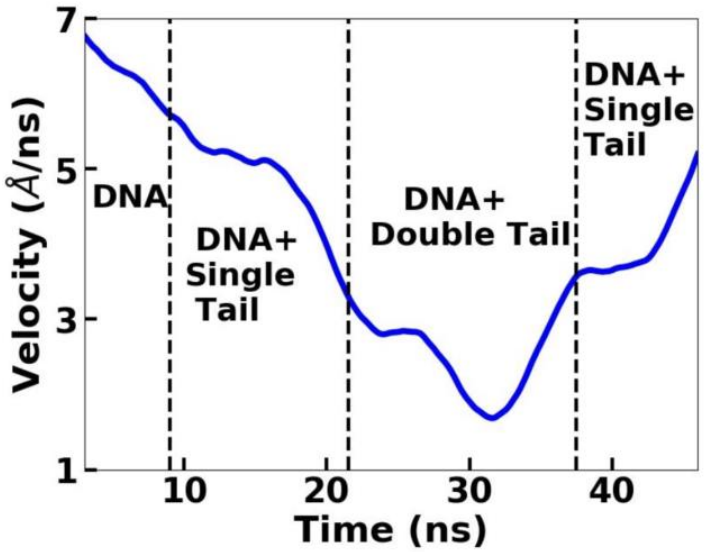


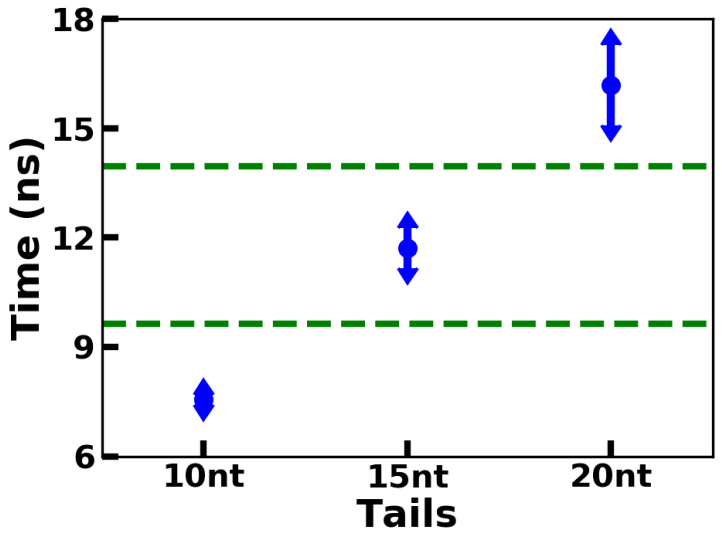




Discrimination Among Tail Lengths







Decision Boundaries :
Bisecting the center of error bar

$$t_{norm} = \frac{v_{ol}}{v_{dna}}\, t_{ol} + \frac{v_{nol}}{v_{dna}}\, t_{nol}$$

v_{dna}: 6.1 Å/ns
v_{nol}: 3.98 Å/ns
v_{ol}: 2.91 Å/ns

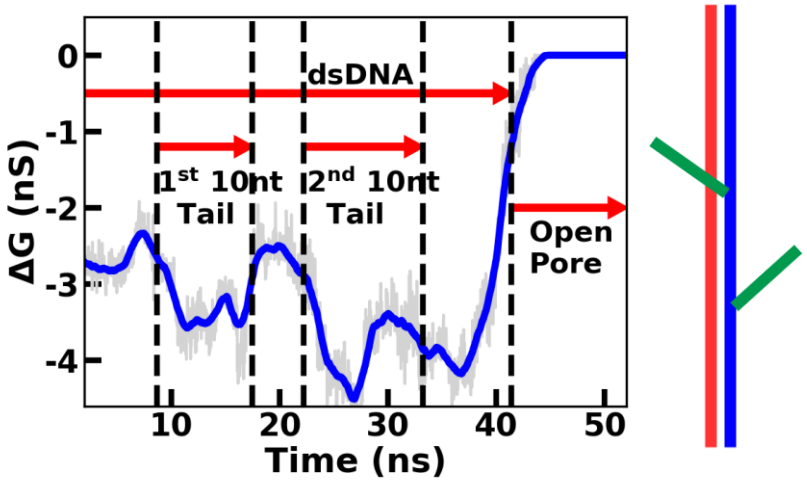
Decision Boundaries

10 nt : t_{norm} <9.64 ns

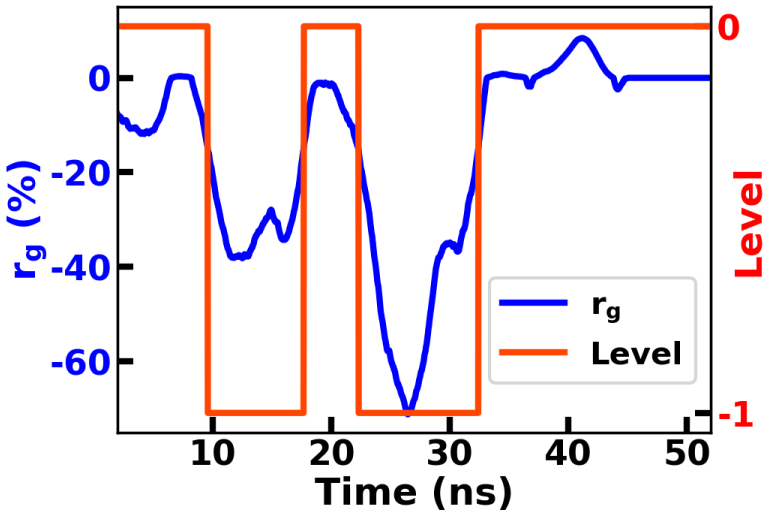
15 nt : 9.64 ns < t_{norm} <13.95 ns

20 nt : t_{norm} >13.95 ns

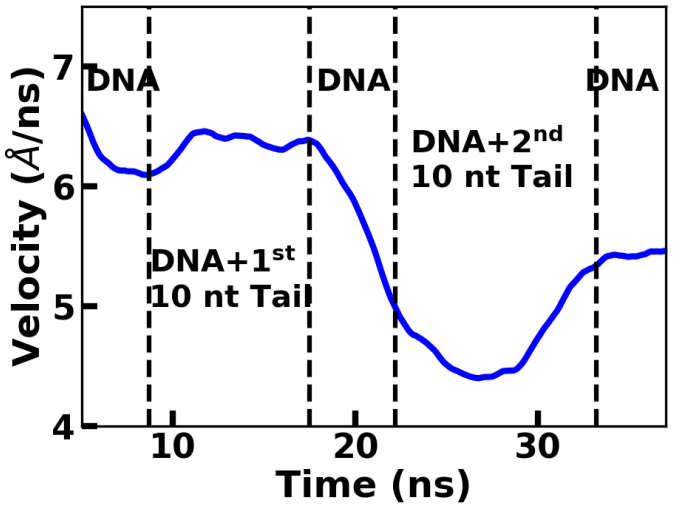
Separation Distance and Other Criteria



Two 10 nt Tails :
15 bp separation



t_{norm}=8.1 ns, 10.11 ns

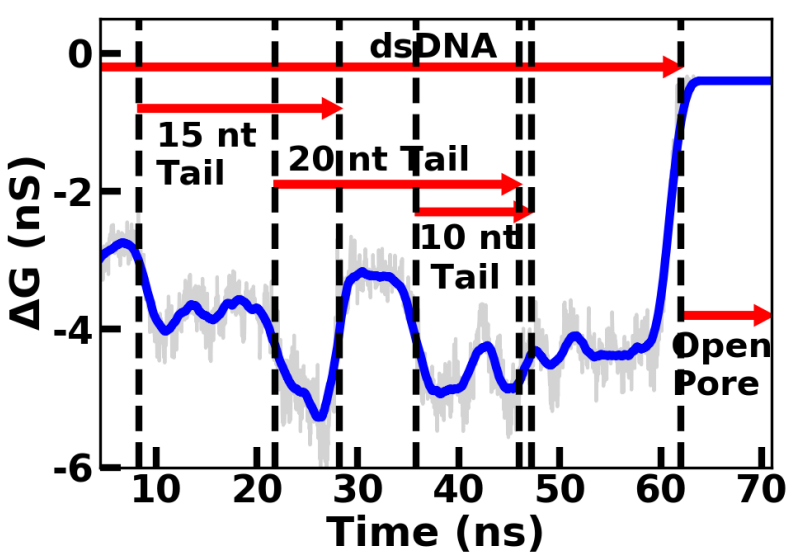


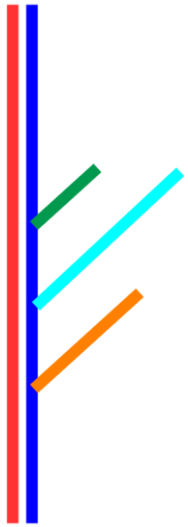
$$t_{norm} = \frac{v_{ol}}{v_{dna}} t_{ol} + \frac{v_{nol}}{v_{dna}} t_{nol}$$

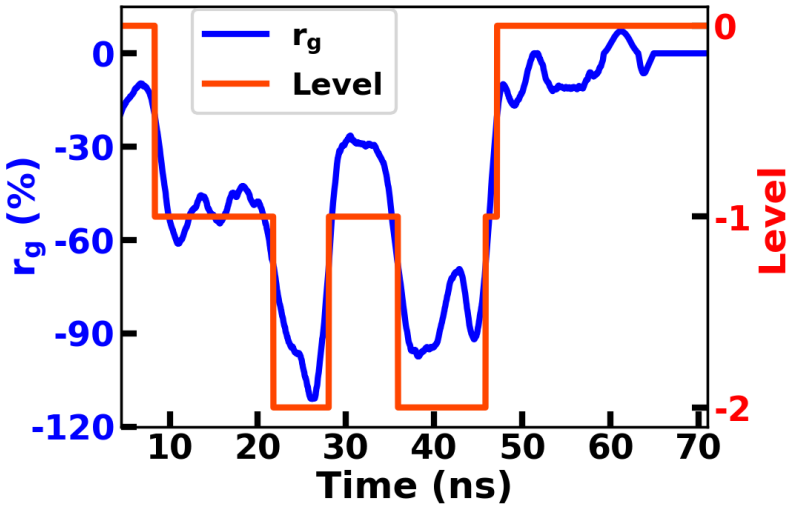
- Length difference
- Separation distance
- Velocities and level threshold values

Robustness of Algorithm

- Three Tails : 15 nt, 20 nt and 10 nt, sequentially with 10 bp separation
- 80 bp dsDNA







$$t_{norm} = \mathrm{m}\left(\frac{v_{ol}}{v_{dna}}\, t_{ol} + \frac{v_{nol}}{v_{dna}}\, t_{nol}\right)$$

m=80/70 assuming velocity change to be linear

13.47 ns : 15 nt
14.91 ns : 20 nt
6.57 ns : 10 nt

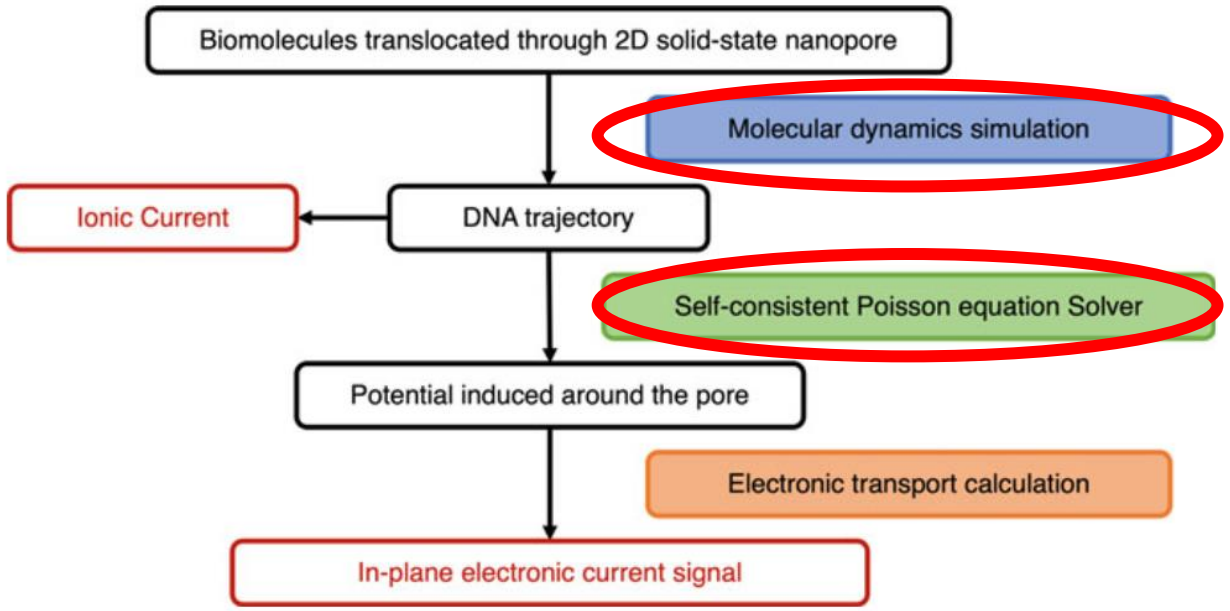
Decision Boundaries

10 nt : t_{norm} <9.64 ns

15 nt : 9.64 ns < t_{norm} <13.95 ns

20 nt : t_{norm} >13.95 ns

Use of Frontera



- Used in MD and Poisson-Boltzmann solver
- Local workstation took 20+ days to simulate a single structure

- **Frontera took ~5 days**
- **Can simulate multiple structure in parallel**
- Size of generated files : ~20 GB for some cases

- Simulation detail – 2 nodes for equilibrium simulation, 6 nodes for trajectory under applied bias
- CPU affinity and core mapping for faster simulation

Summary

Summary

- Detect presence of RNA tails on dsDNA using relative conductance
- Differentiate tail lengths from normalized dwell time
- Robustness of algorithm against tail numbers and dsDNA lengths

Future Prospects

- Improvement of tail separation using machine learning
- Expandable for inclusion of other biomolecules e.g., methylation
- Disease detection

Outcome

Paper:

Chakraborty, R., Xiong, M., Athreya, N., Tabatabaei, K., Milenkovic, O., and Leburton, J.P., “Solid-State MoS_2 Nanopore Membranes for Discriminating among the Lengths of RNA Tails on a Double-Stranded DNA: A New Simulation-Based Differentiating Algorithm,” *ACS Applied Nano Materials* 6, 6, 4651–4660 (2023)

Patent:

Leburton, J.P., Milenkovic, O., and Chakraborty, R., "Methods for Encoding Information in the Pattern of RNA Tails Attached to a Double Stranded DNA and for Reading Such Information Using Solid State Membrane Nanopores" *US Non-Provisional Patent*, Application No: 23-1712-US (2024) (pending)

The Grainger College of Engineering

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN