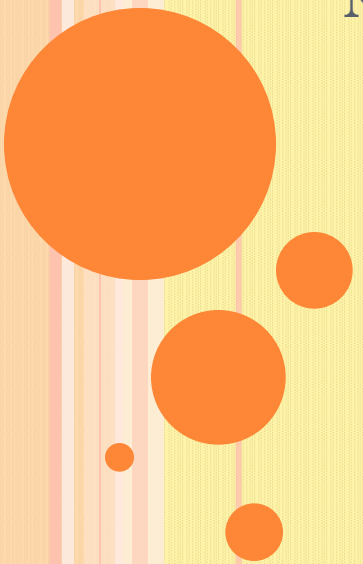
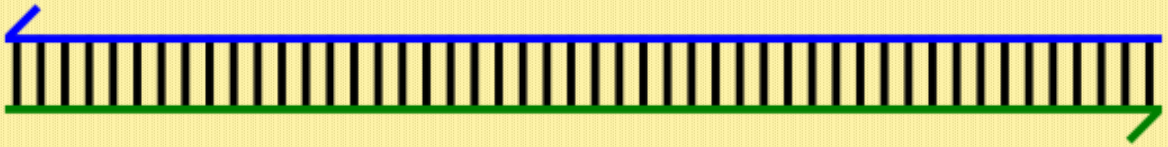
High-Fidelity DNA Hybridization using Programmable Molecular DNA Devices

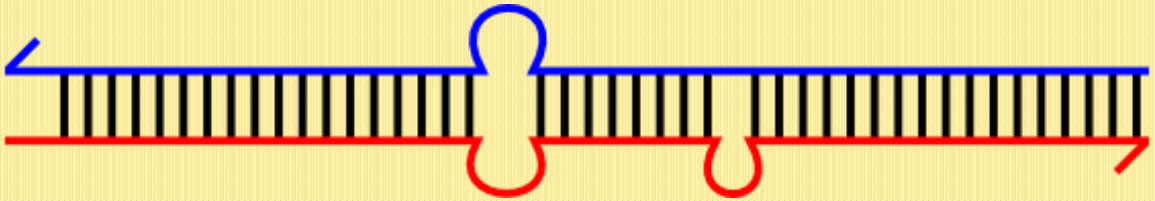
Nikhil Gopalkrishnan, Harish Chandran & John Reif



Fidelity of Hybridization



Perfect hybridization



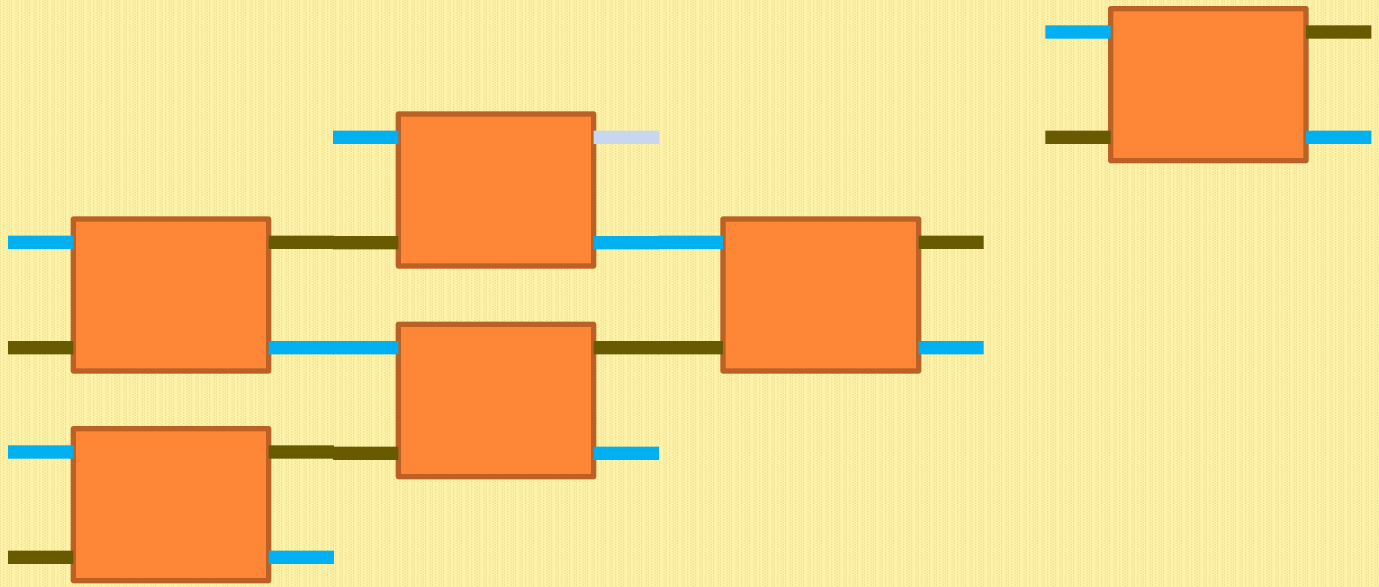
Mismatched hybridization

Difference in energy between red strand hybridization and green strand hybridization is small

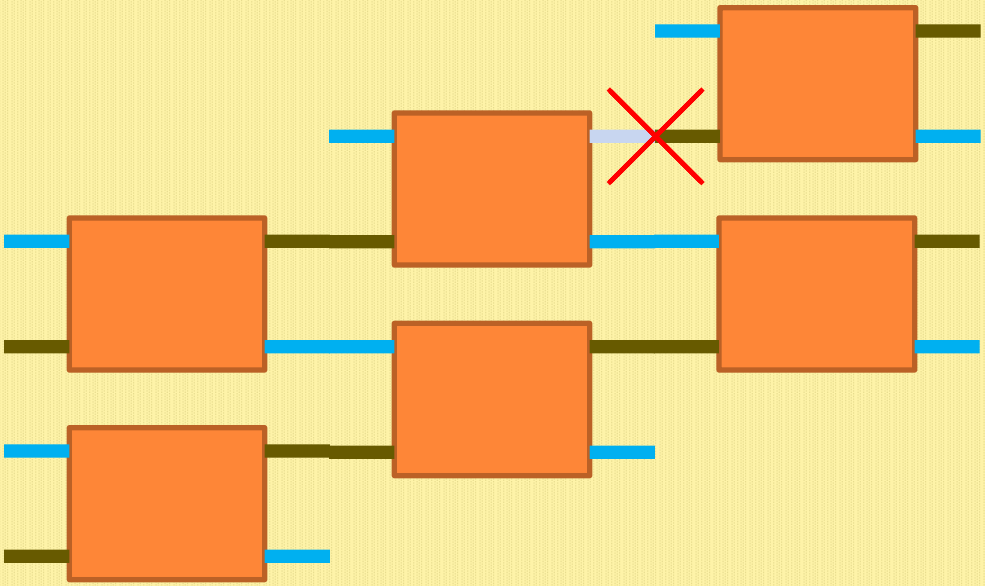
Fidelity of Hybridization

- Hybridization fidelity depends on length
- Errors in hybridization
  - Noise: Strands with sequence similar to the target

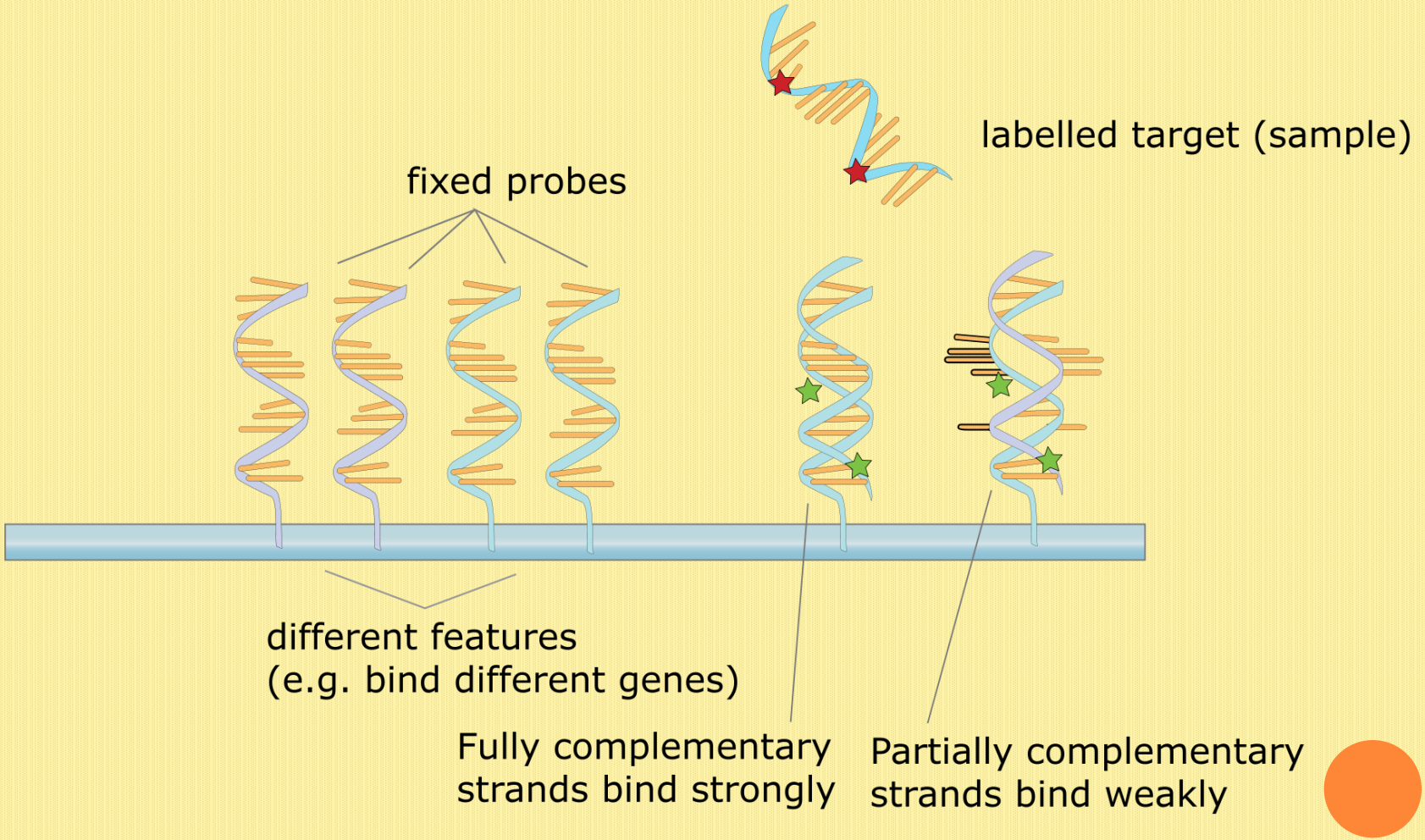
Drawbacks of Low Fidelity: Self-Assembly



Drawbacks of Low Fidelity: Self-Assembly

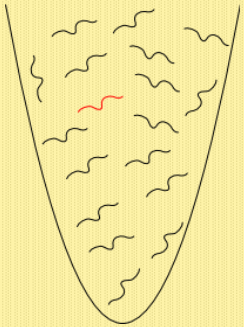


Drawbacks of Low Fidelity: DNA Microarrays

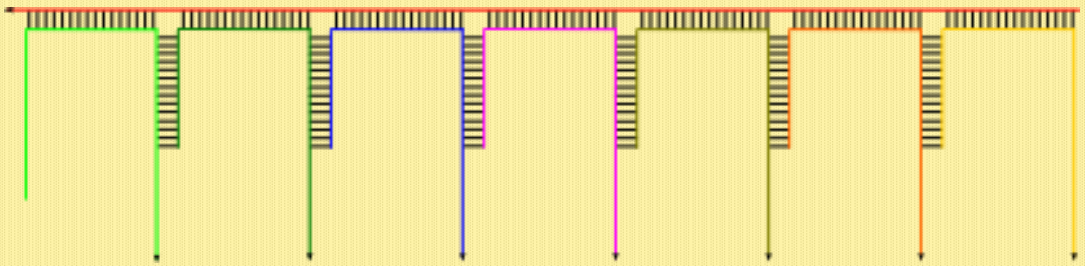


From: http://en.wikipedia.org/wiki/File:NA_hybrid.svg

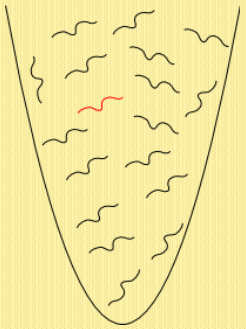
Exact High-Fidelity Hybridization



- Solution: ensemble of distinct sequences
  - Target sequence s
- Problem statement: Completely hybridize all copies of s and don't hybridize any other sequence
- Multiple strands may bind to s and cooperatively hybridize it



Exact High-Fidelity Hybridization



- Solution: ensemble of distinct sequences
  - Target sequence s
- Problem statement: Completely hybridize all copies of s and don't hybridize any other sequence
- Multiple strands may bind to s and cooperatively hybridize it
- Completion of hybridization should be detectable
  - Example: by fluoroscence

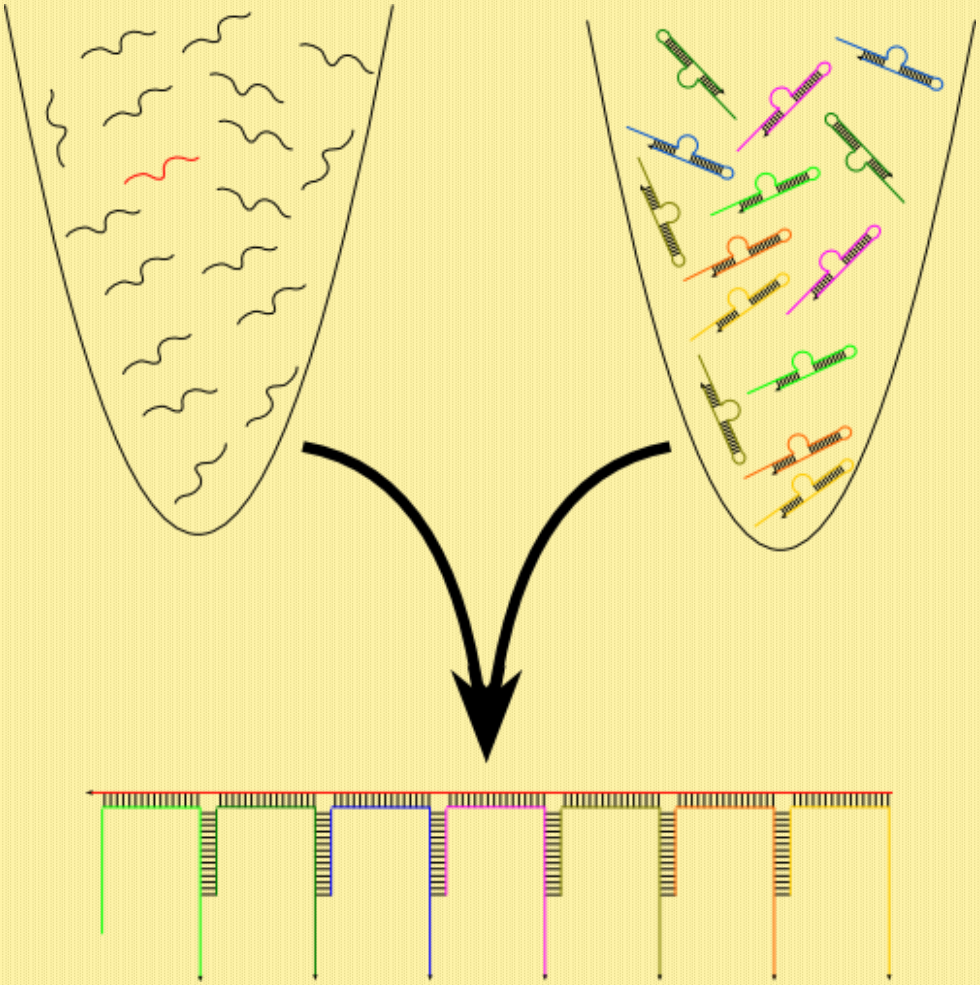
Approximate High-Fidelity Hybridization

- Hybridization Error
  - At most b bases may mismatch: b-hybridized
- Success probability
  - probability of b-hybridization at least p
- Problem statement: b-hybridize each copy of s with probability at least p and no other sequence is b-hybridized with probability greater than 1-p
- p ≈ 95% and b ≈ 1/10th of length of s

Assumptions

- Short sequences have high fidelity of hybridization
- Subsequences sequestered in short hairpins are unreactive
- Strand displacement occurs whenever possible and proceeds to completion

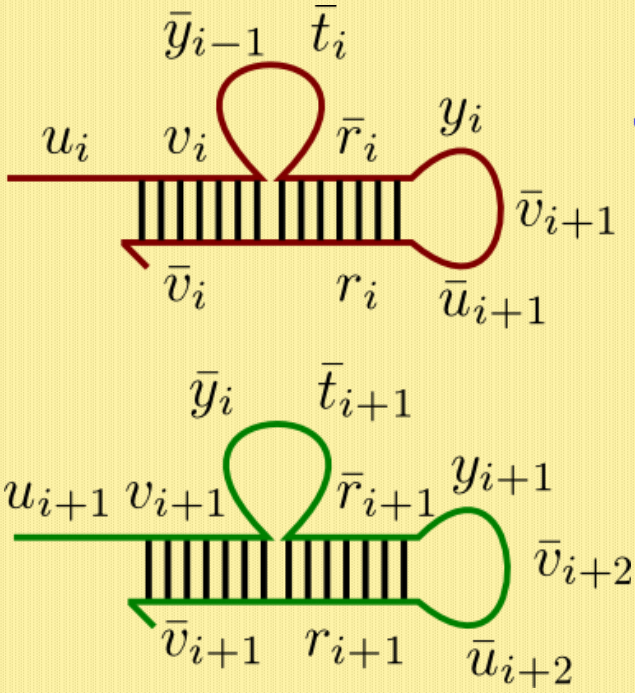
Approximate High-Fidelity Hybridization

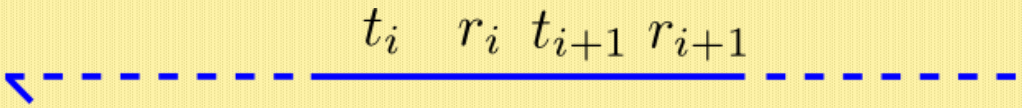


Notation

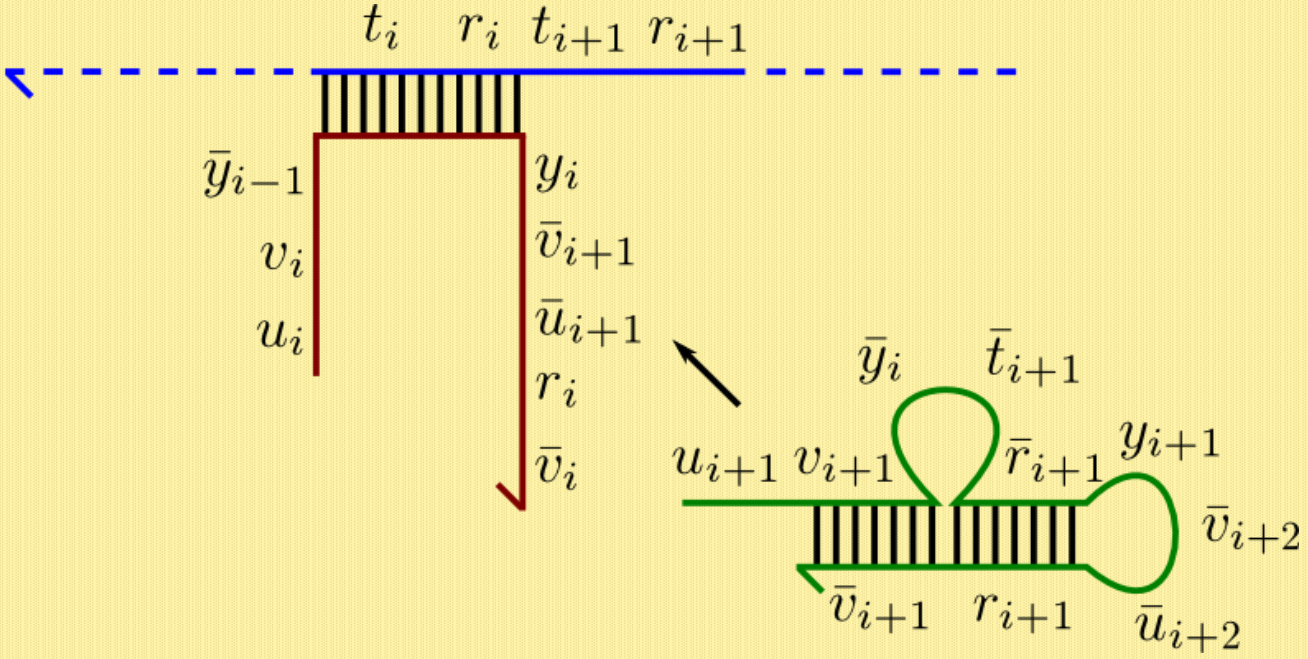
- Letters represent sequences
  - Example: c_i
- Sequences concatenate
  - $c_i = a_i\, b_i$
- Written from 5' to 3'
- Sequences differing only in the subscript are concatenations of subsequences differing only in the subscript
  - $c_i = a_{i+1}\, b_i$ implies $c_{i+1} = a_{i+2}\, b_{i+1}$
- Bar indicates reverse complement
  - $\overline{c}_i = \overline{b}_i\, \overline{a}_i$ is the reverse complement of $c_i = a_i\, b_i$

High-Fidelity Hybridization: 1st Protocol

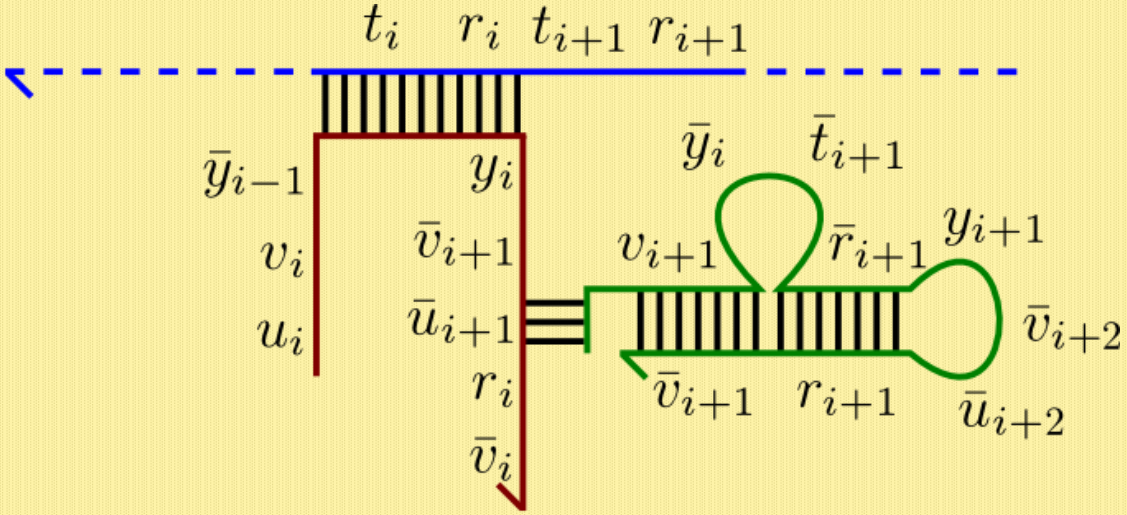




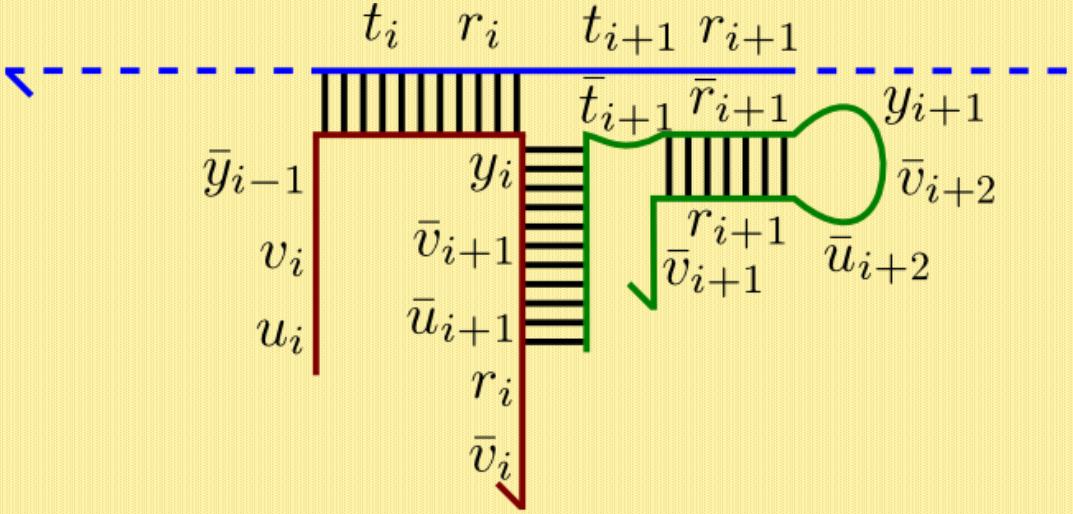
High-Fidelity Hybridization: 1st Protocol



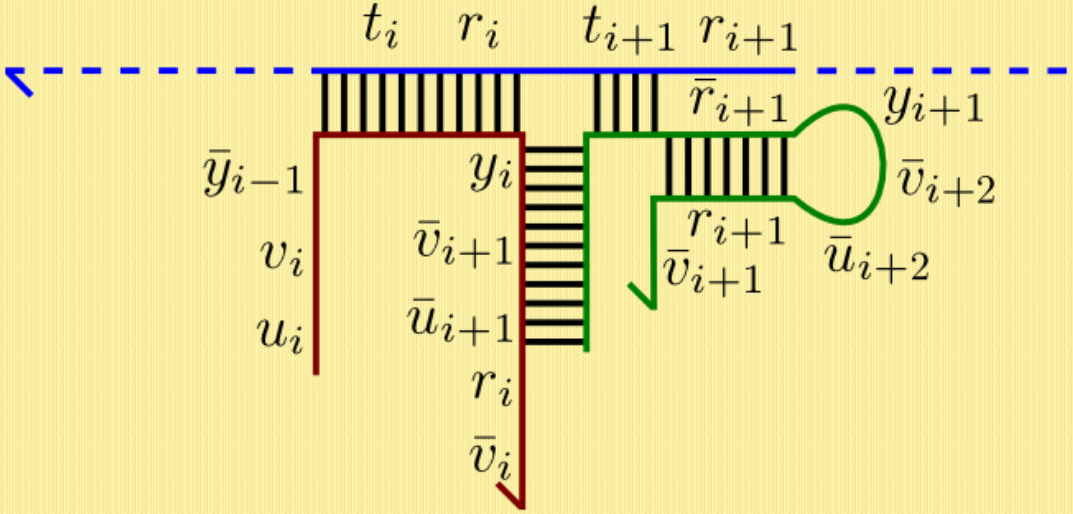
High-Fidelity Hybridization: 1st Protocol



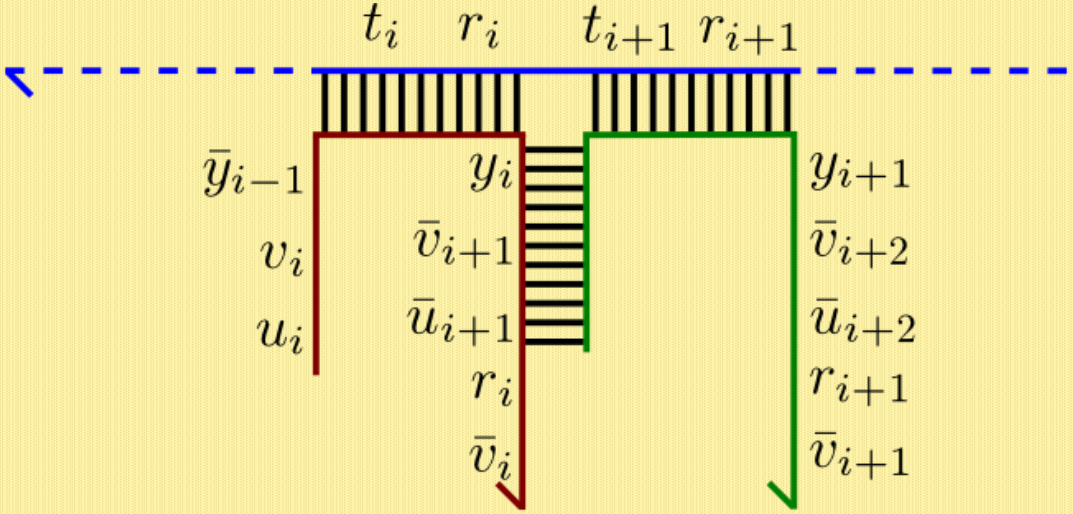
High-Fidelity Hybridization: 1st Protocol



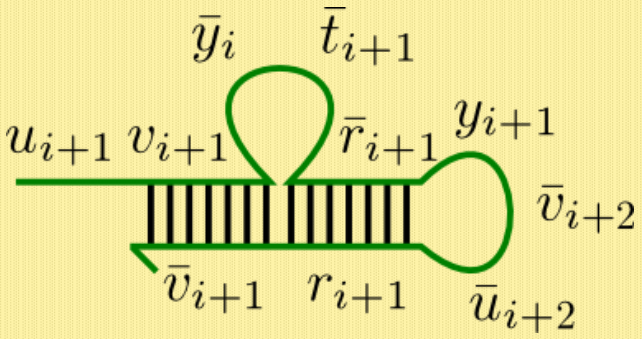
High-Fidelity Hybridization: 1st Protocol

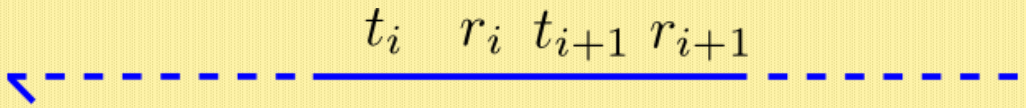


High-Fidelity Hybridization: 1st Protocol

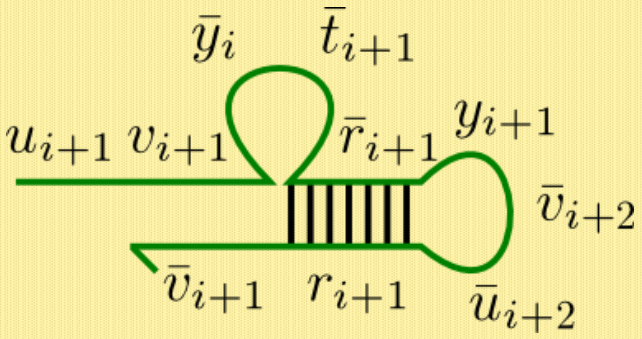


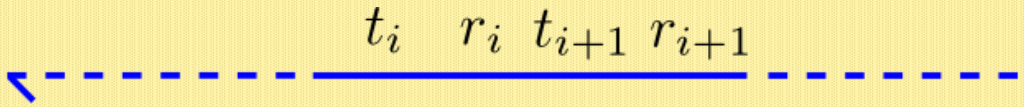
1ST PROTOCOL: POTENTIAL SOURCE OF ERROR



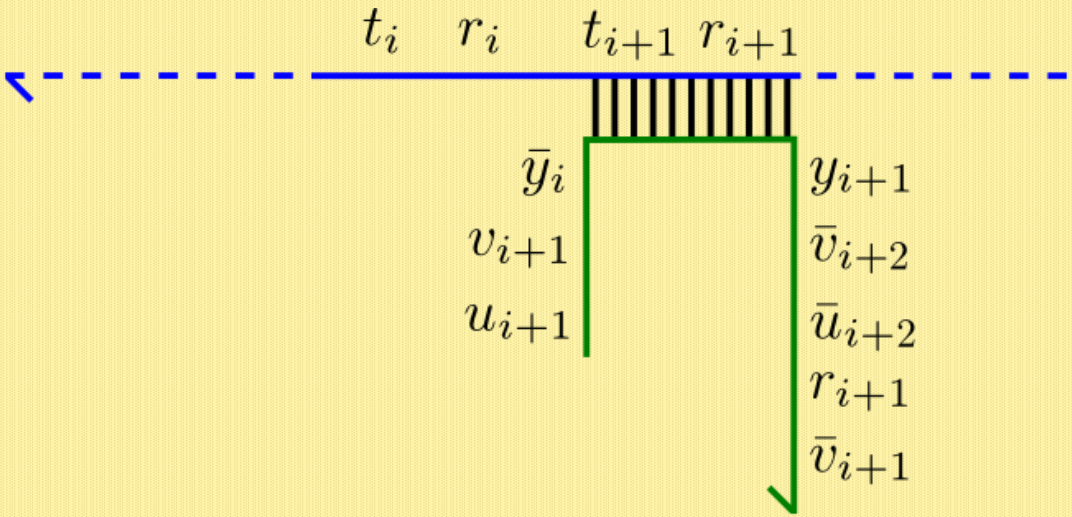


1ST PROTOCOL: POTENTIAL SOURCE OF ERROR

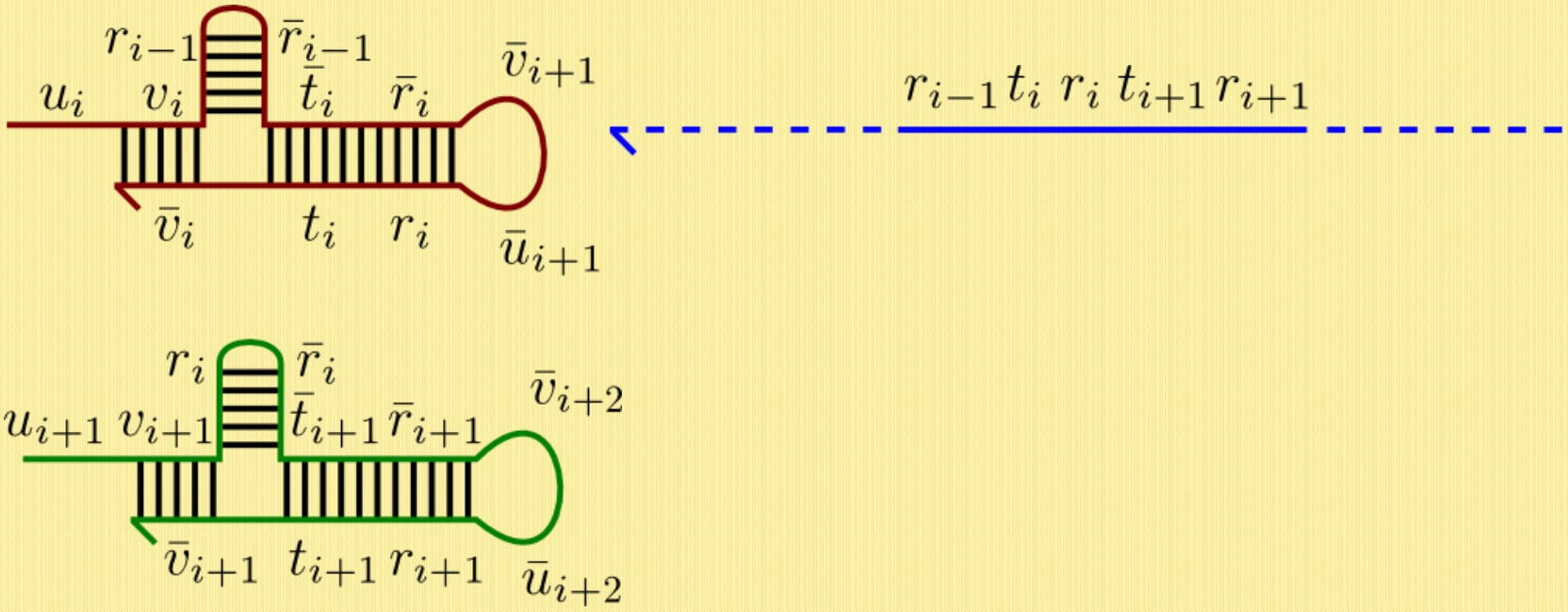




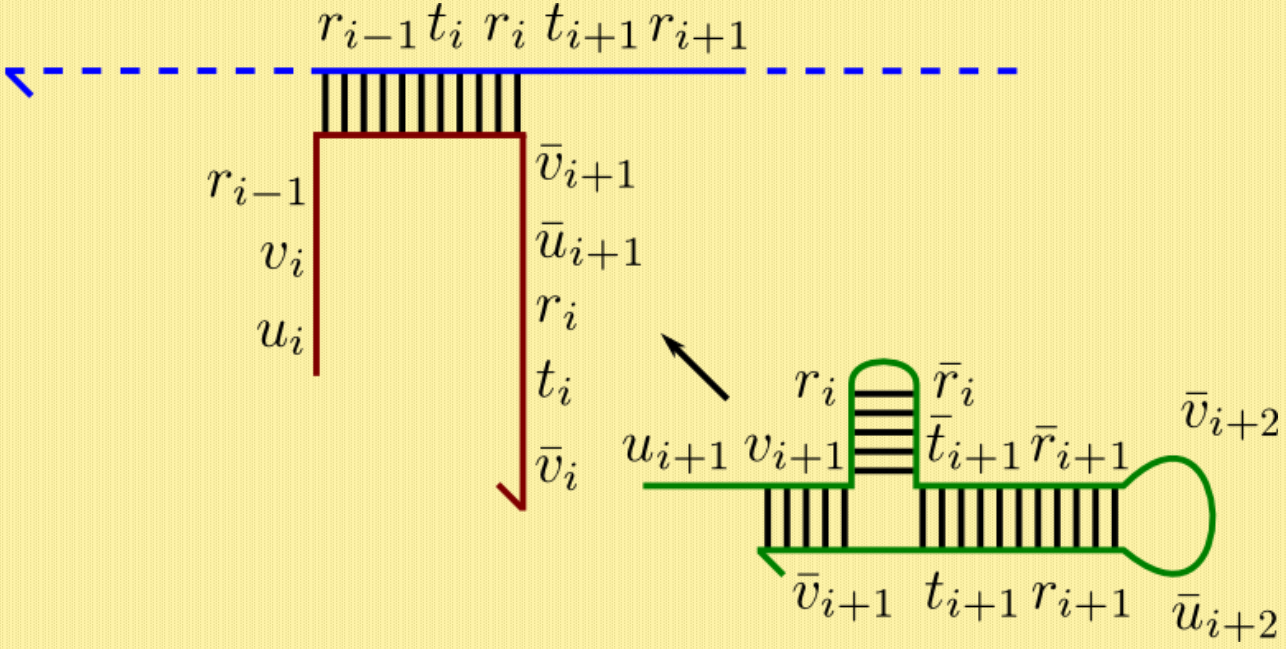
1ST PROTOCOL: POTENTIAL SOURCE OF ERROR



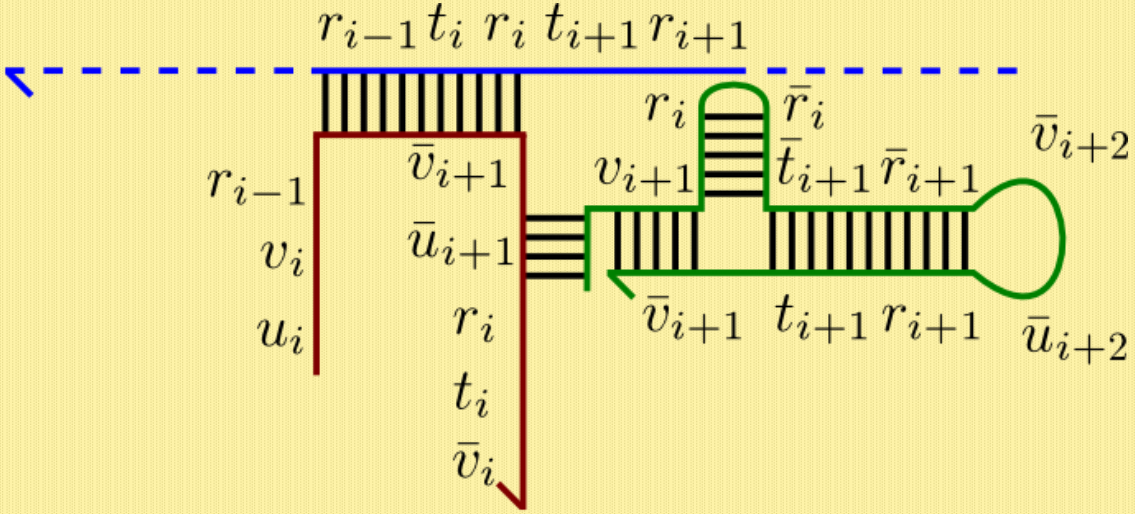
High-Fidelity Hybridization: 2nd Protocol



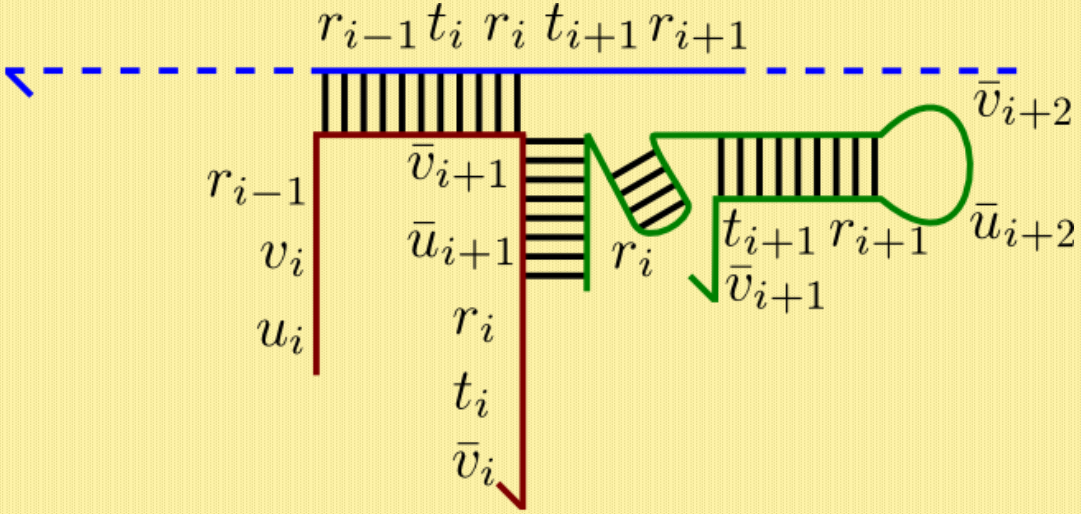
High-Fidelity Hybridization: 2nd Protocol



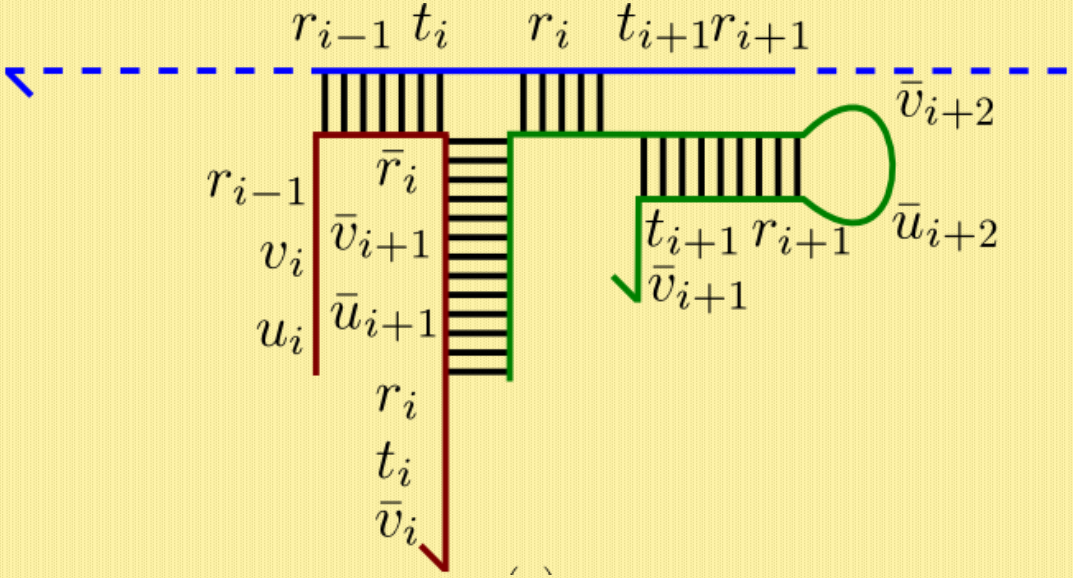
High-Fidelity Hybridization: 2nd Protocol



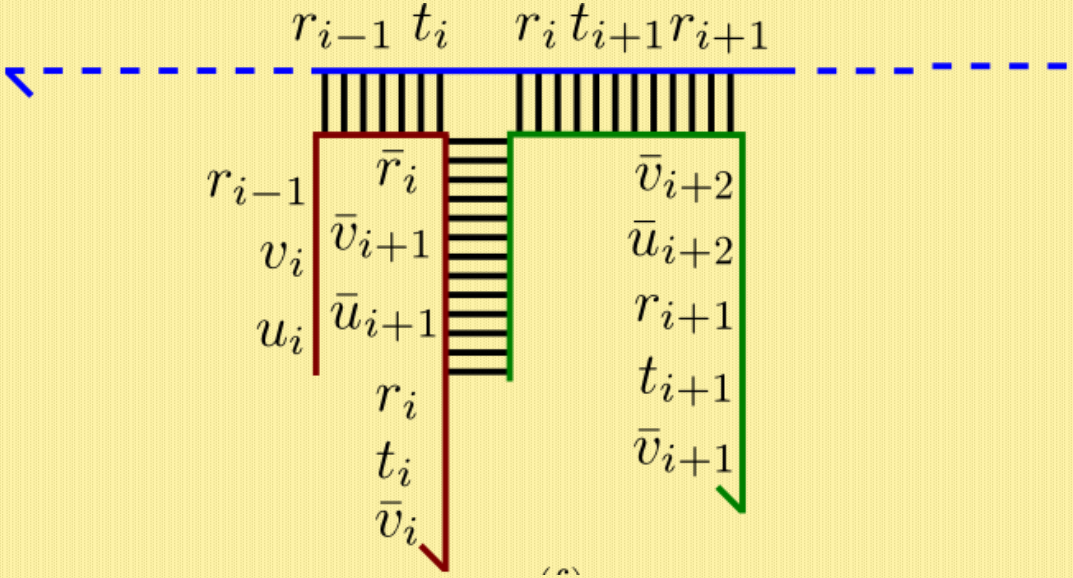
High-Fidelity Hybridization: 2nd Protocol



High-Fidelity Hybridization: 2nd Protocol



High-Fidelity Hybridization: 2nd Protocol

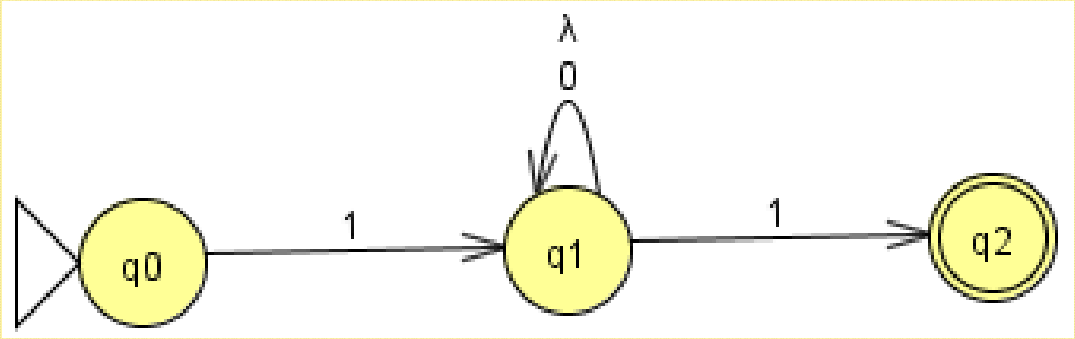


Favorable Properties of the Protocols

- Autonomous
- Fluorophore based detection

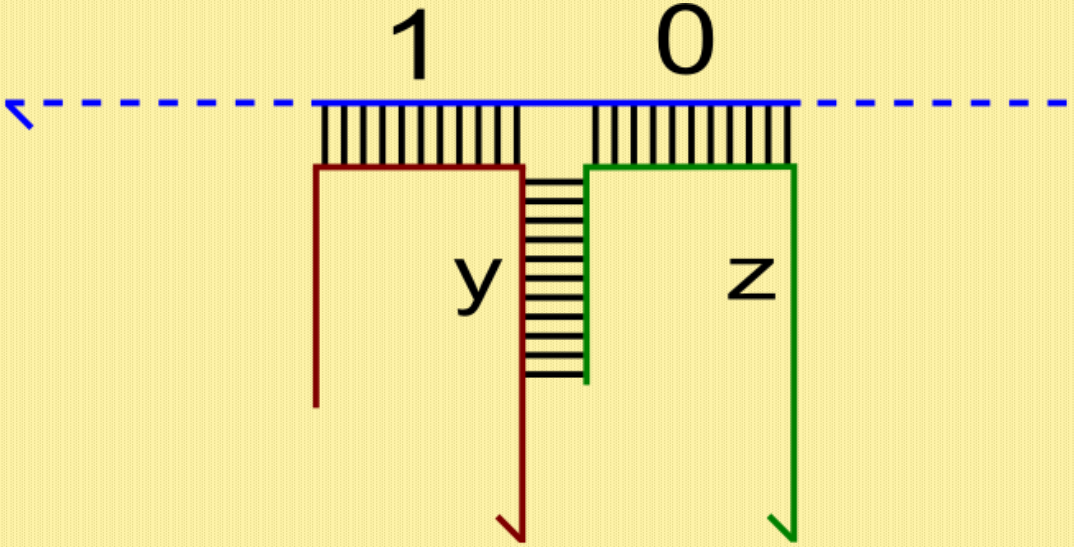
Simulation of Finite Automata

- Finite automata: Mathematical constructs that define languages
- Limited computational power
  - Memoryless



Simulation of Finite Automata

- Target strand encodes input to automata
- Checker sequences perform state transitions

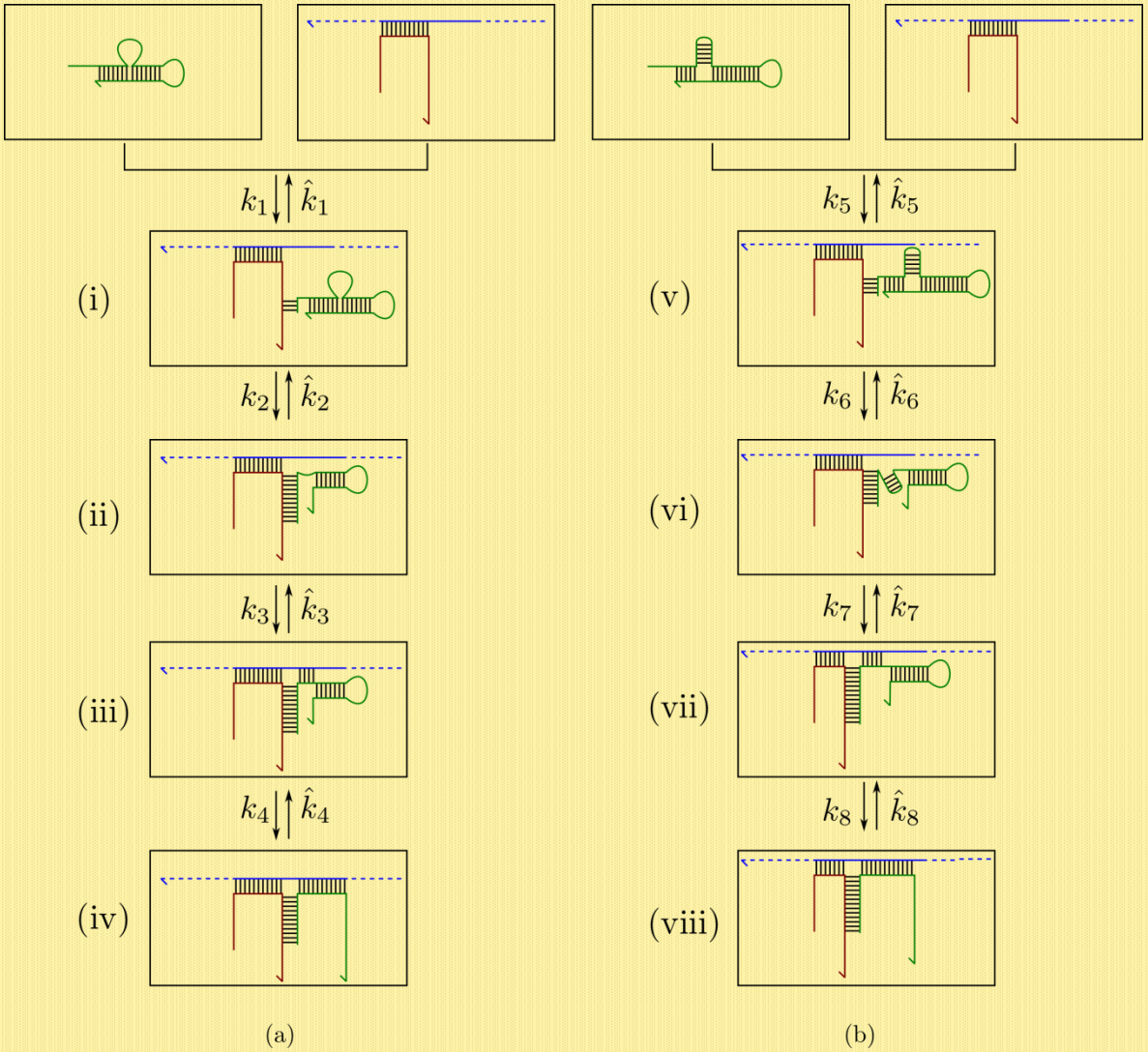


- Green sequence performs $\delta(y,0) = z$

Simulation of Finite Automata

- Incorrect checker sequence may attach
  - Further attachment is blocked as second hairpin doesn't open
- At each step, probability of correct attachment ≥ 0.5
  - Probability of successful completion ≥ $1/2^{n}$ where n=size of i/p
- Can process multiple inputs in parallel
- Number of checker sequences ≤ Twice number of edges in the transition diagram of the automata

Protocol Kinetics



Future work

- Experimental verification for a simple case with just two checker sequences
- Computer simulation to predict reaction kinetics