Computing and Biomolecules

Alvin R. Lebeck
Duke University

Partial Goals of Talk

- Introduce you to potentially disruptive technology
  - Opportunities & Challenges
- Challenge you to think “outside the box”
  - Maintain vs. break abstractions
- Bridge the Engineering Gap
- Back to the Future: understand entire stack from chemistry/physics up through applications (hipster architect?)
- Be interdisciplinary!

Setting Context

- Computing
  - Processing and storing information
- Biomolecules
  - DNA, proteins, fluorescent molecules, etc.
  - Everyday use in the Life Sciences

1. Why put these together?
2. How do we put these together?

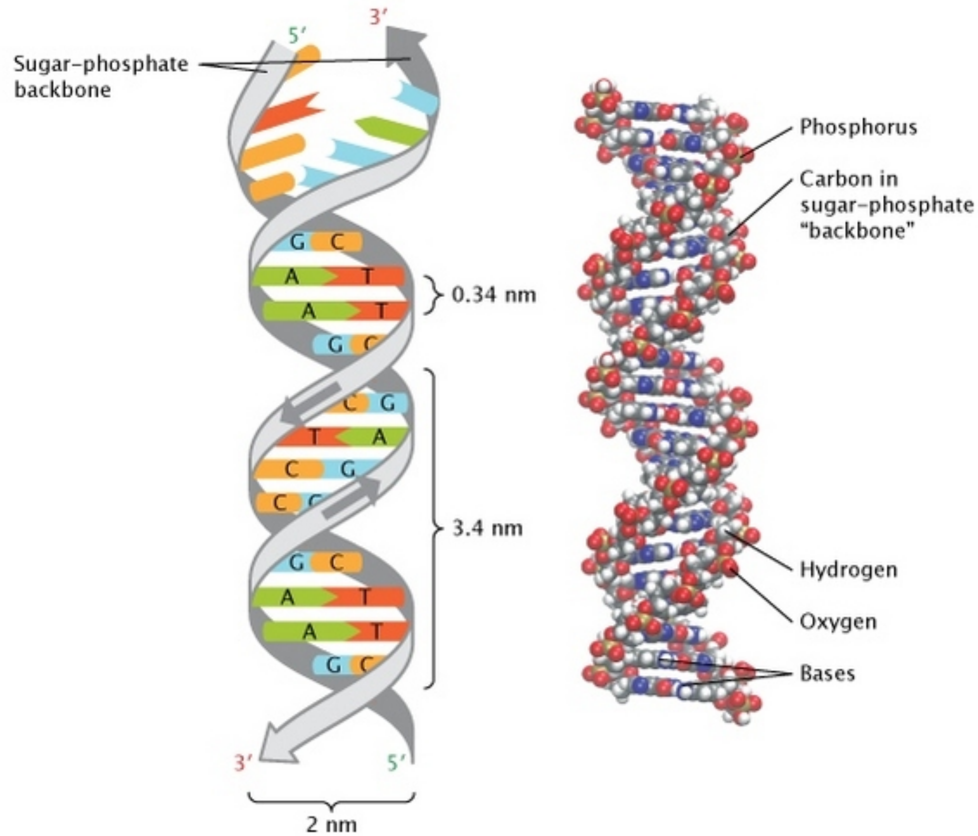
- First some background on biomolecules

Biomolecules: Synthetic DNA

- Single strand is sequence of nucleotides
- Well defined rules for base pair matching
  - Thermodynamics driven hybridization
  - Forms well-known double helix
- Molecular Scale
  - 3.4 Angstrom spacing
  - 2nm diameter
- Synthetic
  - Specify sequence of bases
  - Engineer systems

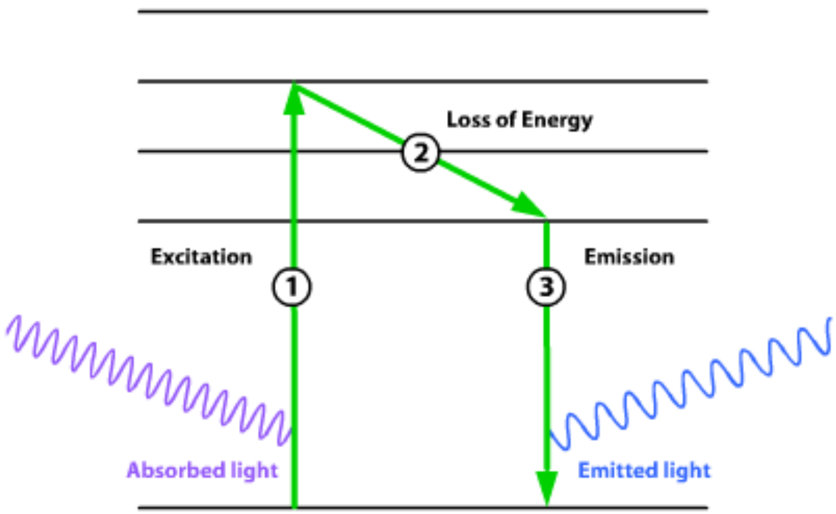
Adenine (A) ⟷ (T) Thymine

Cytosine (C) ⟷ (G) Guanine



[Figure form Pray, Nature Education, 2008]

Biomolecules: Chromophores (Fluorophores)



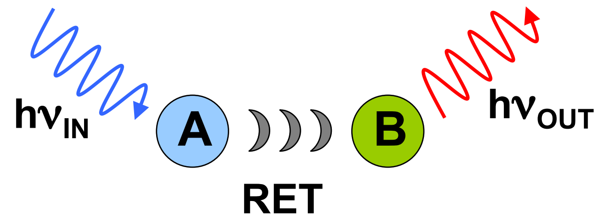
Quantum mechanical description of energy levels

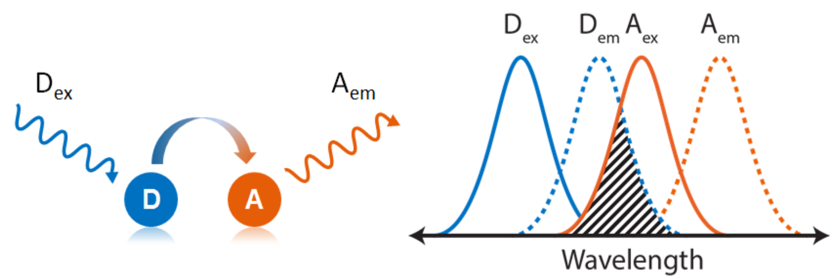
Fluorescence

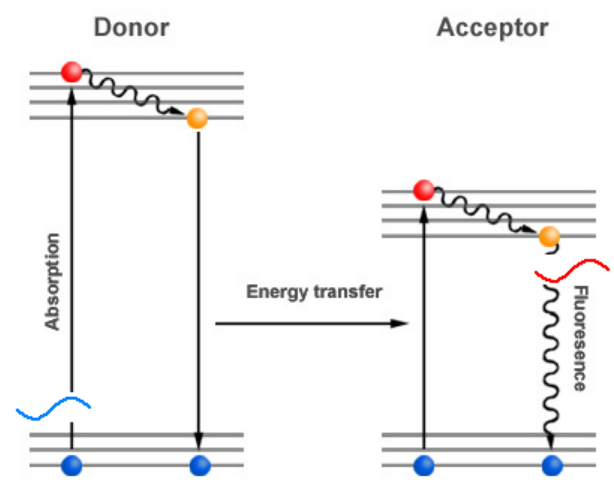
- Optically active small-molecule
- Absorb and emit photons of specific wavelengths
  - Time to fluoresce follows exponential distribution
- Size: ~20-100 atoms

Biomolecules: Resonance Energy Transfer

- Molecular Beacon or Ruler
  - E.g., detect protein folding
- Resonance Energy Transfer (RET)
  - Closely spaced (1-10nm)
  - Non-radiative dipole-dipole interaction
- Efficiency decays with 6th power of distance
- Efficiency depends on spectral overlap and dipole orientation
- Low heat generation (emits far field photon)







Why Biomolecules?

- Scale in feature size
  - DNA: 3.4 Angstroms between base pairs
  - DNA: 2nm diameter double helix
  - Chromophores: 20-100 atoms
- Scale in fabrication
  - Leverage chemical industry
  - Engineer systems at low cost and high volume
  - 1 grad student 8 hours ≈ one month of TSMC Fab 15 throughput
- Low Heat Dissipation
- Common in Life Sciences
- New Domain for computing
  - Biologically compatible
  - E.g., computing within a cell

How do we use Biomolecules?

- Exploit physical properties for

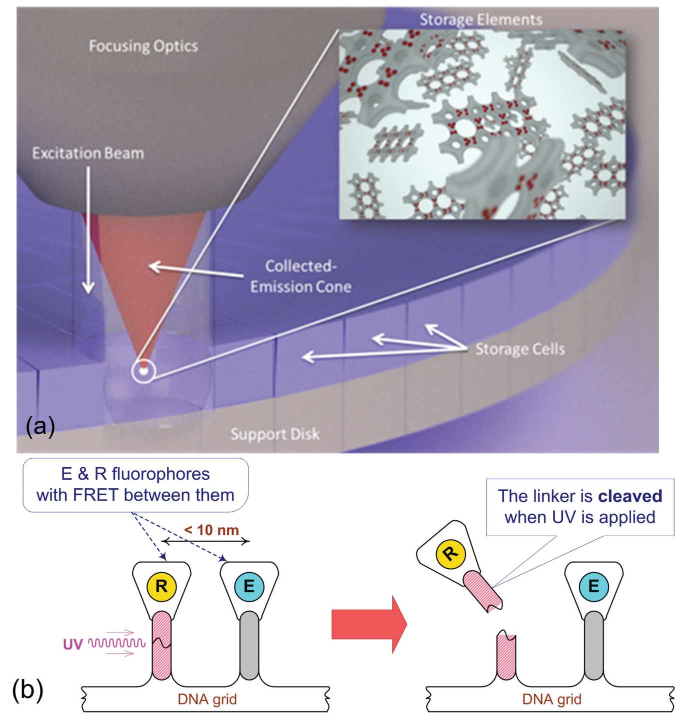
1. Storage
2. Computation
3. Fabrication
    - Place components (including other biomolecules)
    - Gates, circuits, systems

Biomolecular Storage

- Archival Storage
  - DNA base sequence as encoded data
  - Density: 10^9 GB/mm^3
  - Durability: 100s of years
  - Read Latency: DNA Sequencing
- Optical Storage
  - Photo cleavable link of Chromophore to DNA
  - Multiple bits w/in diffraction limit
  - Density: 1000x > blu-ray

Binary data	P 01010000	o 01101111	l 01101100	y 01111001	a 01100001	; 00111011
Base 3 Huffman code	12011	02110	02101	222111	01112	222021
DNA nucleotides	GCGAG	TGAGT	ATCGA	TGCTCT	AGAGC	ATGTGA

[Figure form Barnholdt, et al. ASPLOS 2016]



[Figure from Mottaghi & Dwyer, 2013].

Biomolecular Computation

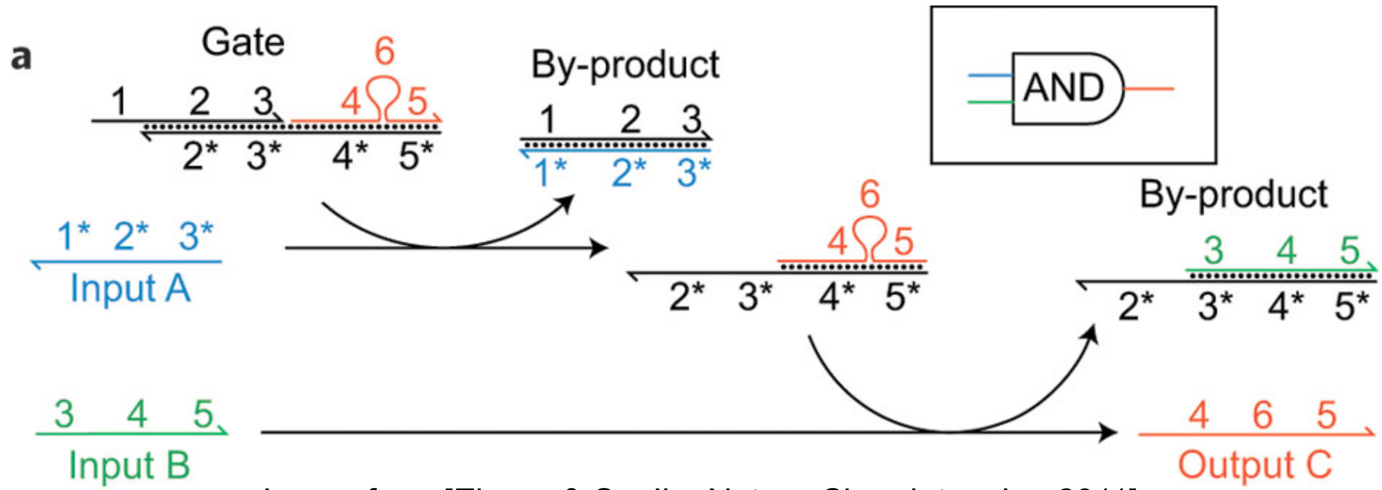


Image from [Zhang & Seelig, Nature Chemistry, Jan 2011]

- Specify sequences such that desired hybridization occurs
- DNA Computing
  - Hamiltonian Path, Tile-based computing,
  - Strand displacement (above)
  - Attach proteins (molecular recognition)
- Molecular Robotics, Synthetic Biology
- Chemical Reaction Networks
  - Careful about different input modes (e.g., concentration of disparate chemicals)

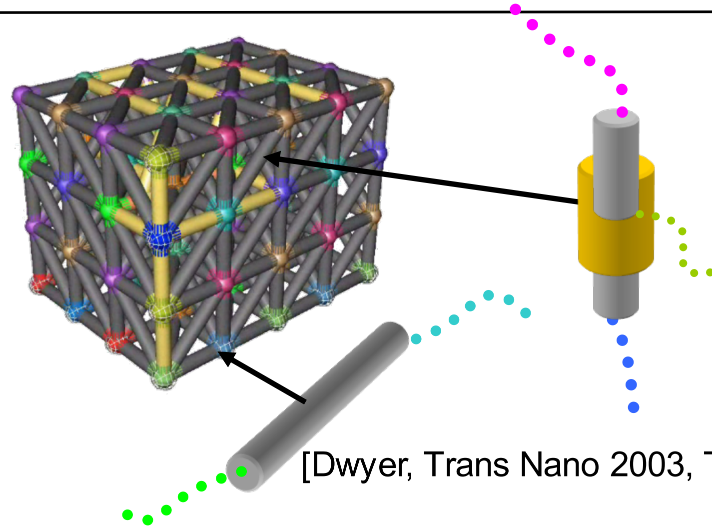
Biomolecular Fabrication

- Molecular Self-assembly
  - Molecules self-organize into stable structures

- What structures?
- What devices?
  - Nanotubes, nanorods, chormophores, etc.
- How does self-assembly affect computer system design?

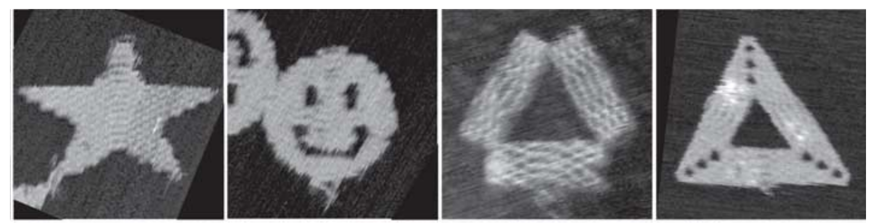
DNA for Structure

- Directed Assembly
  - Functionalize devices, etc.



[Dwyer, Trans Nano 2003, Trans VLSI 2004]

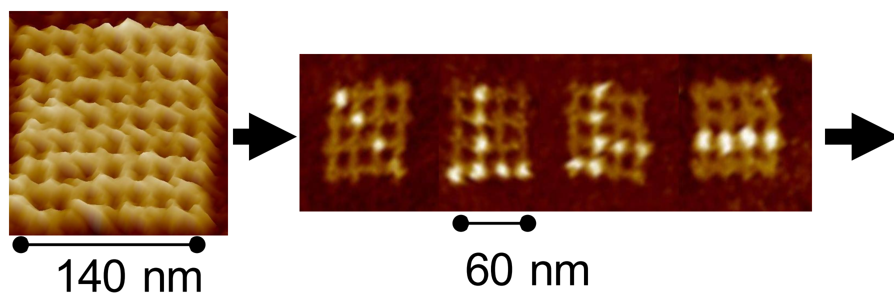
- DNA Scaffold
  - Engineered Structures
  - Origami
  - Hierarchical
  - Scale: ~10^{14} grids/mL

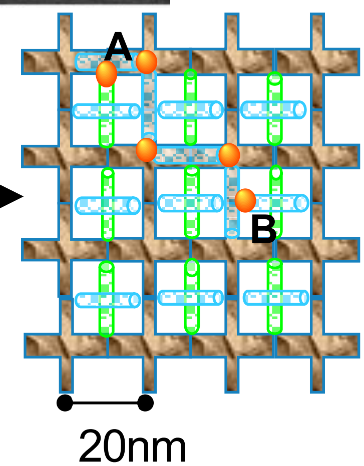


[Rothemund, Nature 2006]

- Can exploit DNA programmability
  - "at fabrication computing"

[IEEE MICRO 2005]

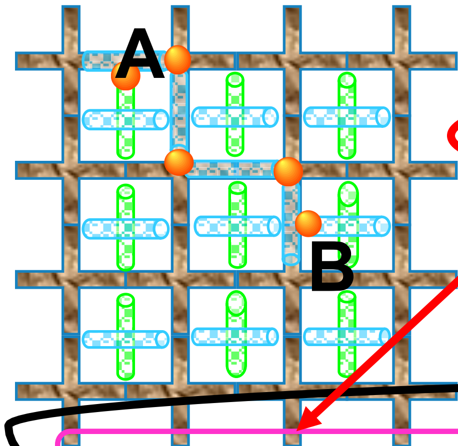




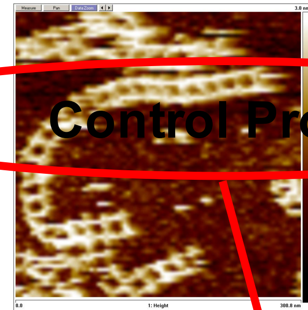
[Patwardhan, et al. 2004 & 2006; Park et al. 2006; Pistol et al. 2006,]

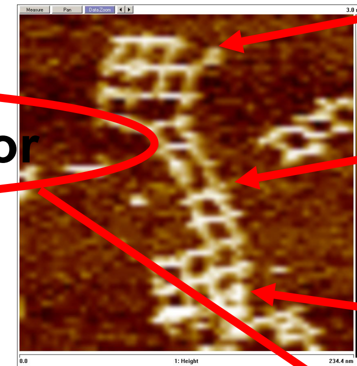
DNA Self-Assembled Parallel Processor

Node



Interconnect





Node

Wire [Yan '03] (selective metallization)

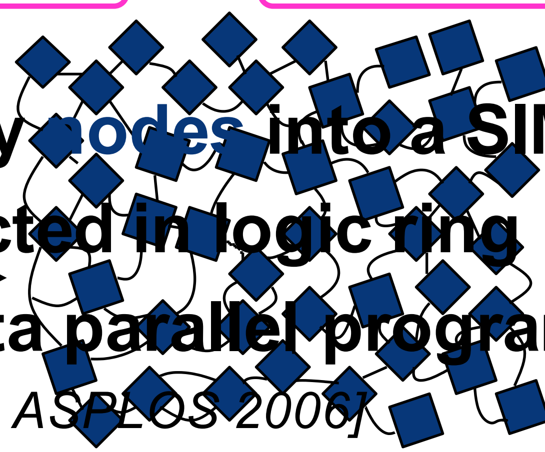
Node

Control Processor

PE

PE

- 20nm



- Self-assemble ~ 10^9 - 10^{12} simple nodes (~10K FETs)
- Potential: Tera to Peta-scale computing
- Random Graph of Small Scale Nodes
  - There will be defects
  - Scaled CMOS may (does) look similar
- How do we perform useful computation?

- **Group many nodes into a SIMD PE**
- **PEs connected in logic ring**
- **Familiar data parallel programming**

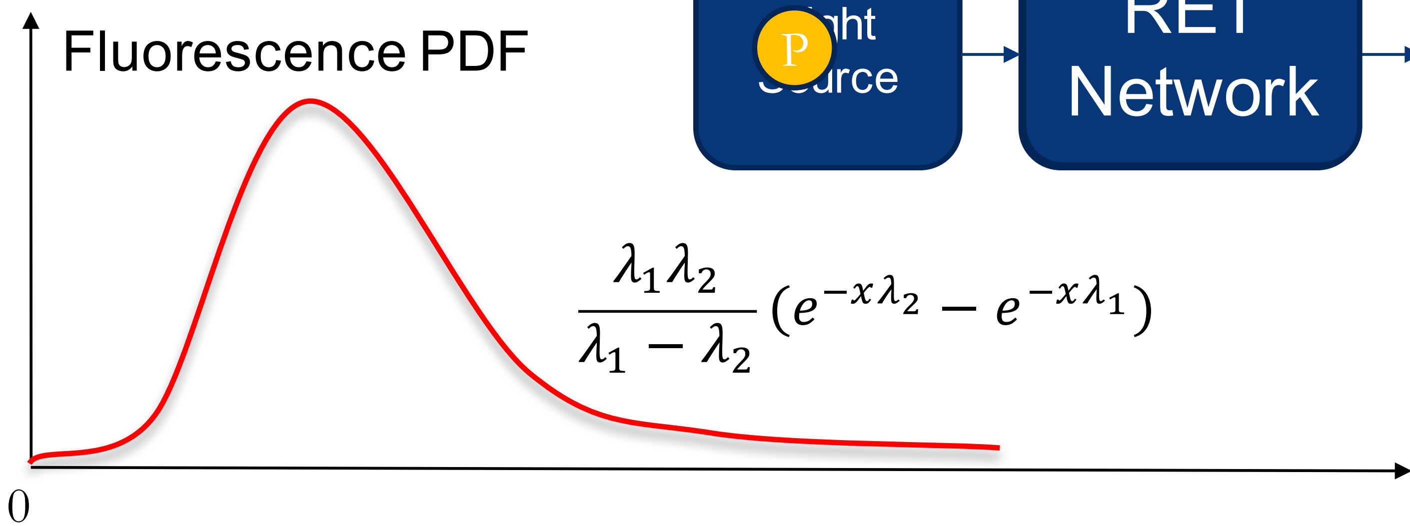
[Patwardhan, et al., ASPLOS 2006]

- **What about those chromophores?**

Duke Architecture

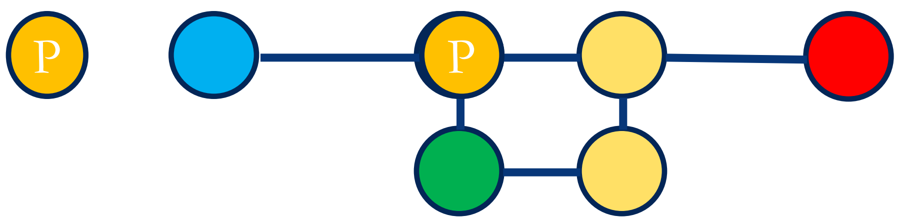
RET-based Stochastic (Probabilistic) Computing

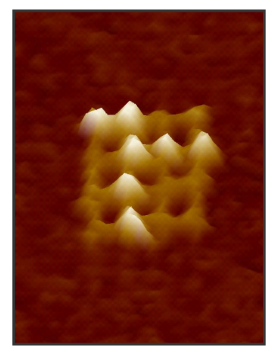
RET Circuit

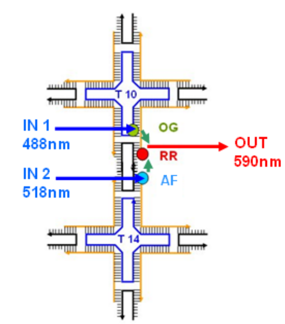


$$\frac{\lambda_1 \lambda_2}{\lambda_1 - \lambda_2}\left(e^{-x\lambda_2} - e^{-x\lambda_1}\right)$$

Resonance Energy Transfer





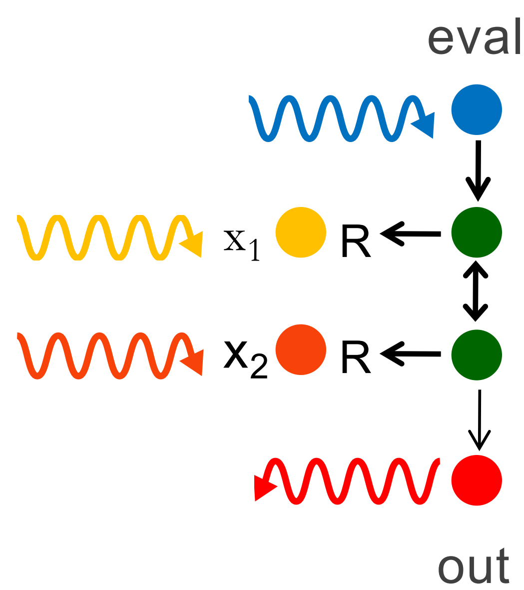


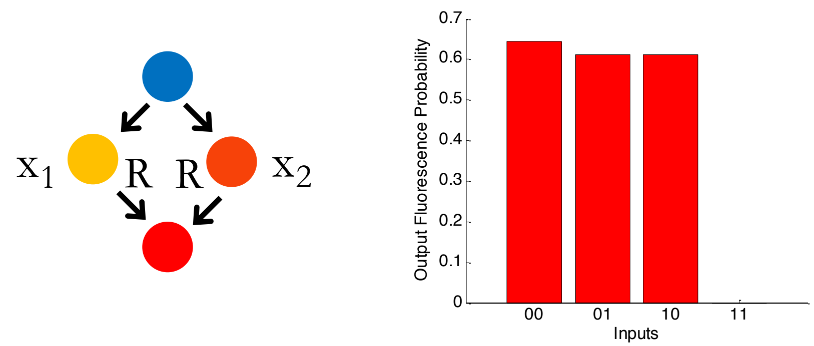
- Multi-chromophore structure: phase-type distribution [Wang et al, 2015].
- Can fit most distributions to phase-type distribution [Asmussen et al, 1996].
- New Functional Unit [Wang et al, 2016]. (Wednesday talk…)

RET-based Logic

- Chromophore types:
  1. Eval – exciton source
  2. Out – output, monitored for fluorescence
  3. Mediators – connect eval to out
  4. Inputs – x1 and x2
     - Disrupt (no RET)
     - Excitation represents applying a 1
- Multistep Cascades
- Energy and Exciton Restoration
- Biologically compatible
  - Sub-diffraction limit addressable sensing [Pistol et al. Small 2010]
  - Nanoscale Sensor Processor smaller than largest known virus [Pistol et al. ASPLOS 2009]

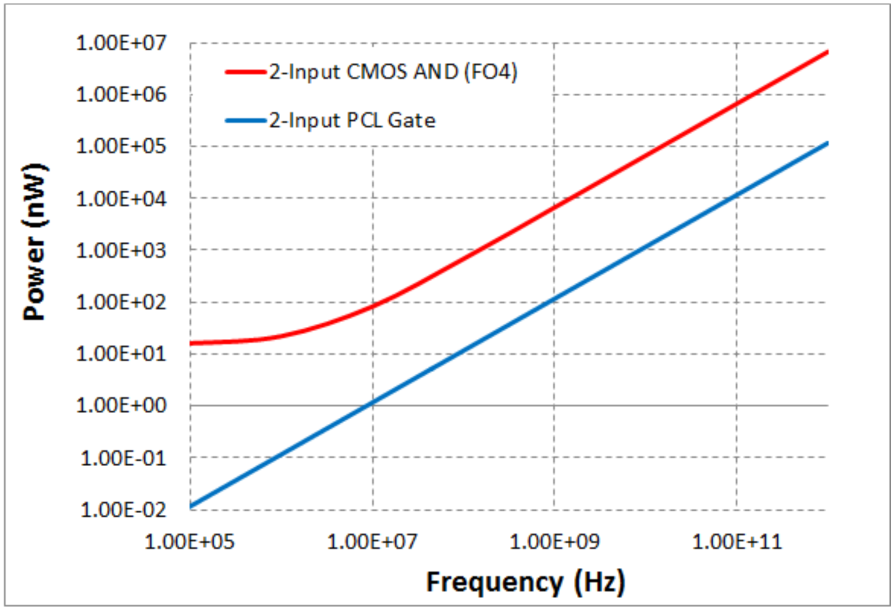
AND Gate Layout





NAND Gate Layout & Simulations

RET-based Logic Power and Area

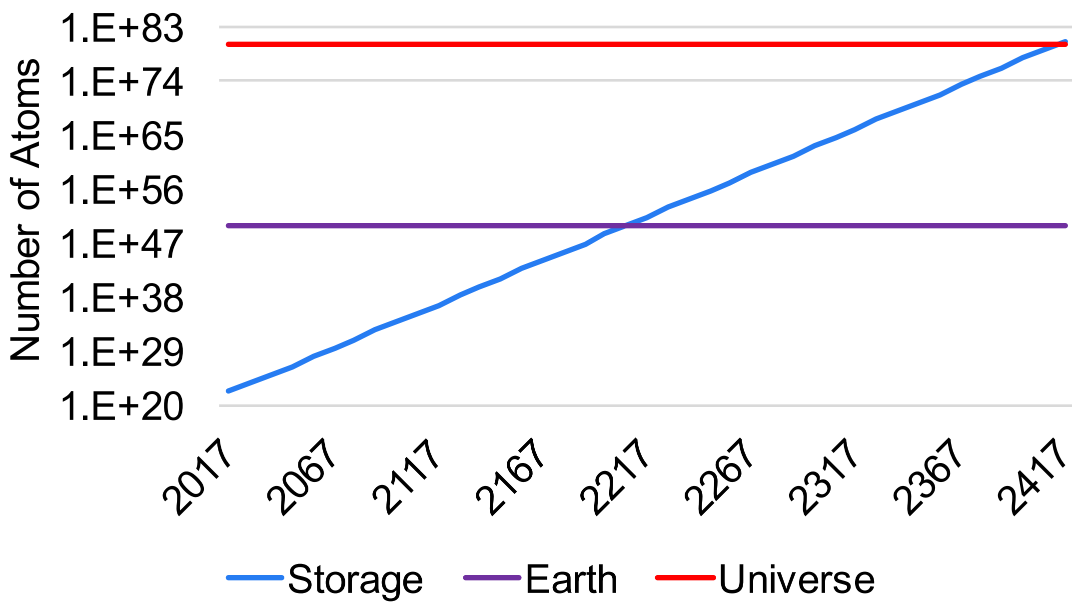


Gate	CMOS (1x) 15 nm	RET-Logic	Improvement
AND	294,912 nm^2	361 nm^2	816x
OR	294,912 nm^2	361 nm^2	816x
NAND	196,608 nm^2	361 nm^2	544x
NOR	196,608 nm^2	361 nm^2	544x

- 15nm CMOS, two-input gates
- Power: RET-Logic100x lower than CMOS
- Area: at least 500-800x smaller than CMOS
  - Conservative: Assumes two input gate occupies entire 19nm x 19nm DNA tile
- Emit far field photon –> no localized heat generation…

The Problem with Exponentials

- Desire for more compute and storage
- Biomolecular scale
- But...O(n!), $O(x^n)$, etc.
  - E.g., storage increases 40%/year
- Not Enough Atoms!
- Earth:
  - 100 years of storage
  - 42 node Hamiltonian
- Known Universe:
  - 200 years of storage
  - 60 node Hamiltonian
- Architecture 2030:
  - Still need algorithms…
  - Use atoms efficiently



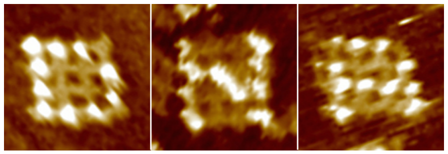
Conclusion

"*It is not the strongest of the species that survive, nor the most intelligent, but the one most responsive to change*."

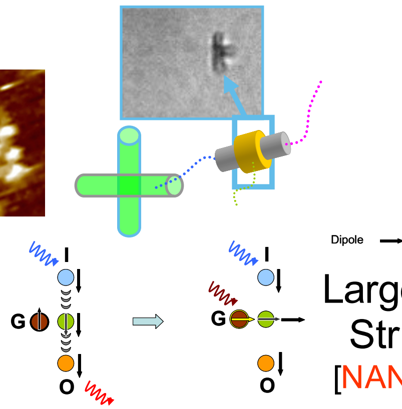
– Charles Darwin

- Technology
  - May not be a single device technology for the future
- Biomolecules
  1. Scale in feature size
  2. Scale in manufacturing
  3. Readily available
  4. New Domain for Computing
  5. Can exploit physical properties
- Interdisciplinary research teams
  - Scale up technology: from bench to processors ("engineering gap")
  - Differing goals/metrics
  - Need bus driver or shared vision
  - Publishing can be difficult…but the research is really fun!

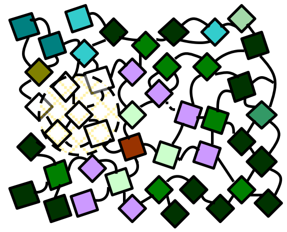
Duke Nanosystems Overview



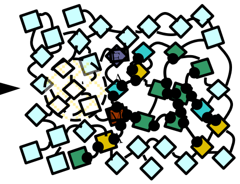
DNA Self-Assembly
[FNANO 2005, Ang. Chemie 2006, DAC 2006]



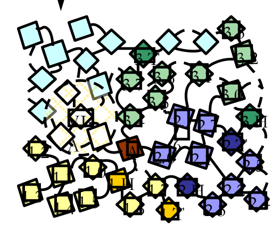
Nano Devices
Electronic, photonic, etc.
[Nanoletters 2006, IEEE MICRO 2008, Small 2010, IEEE MICRO 2015, ISCA 2016]



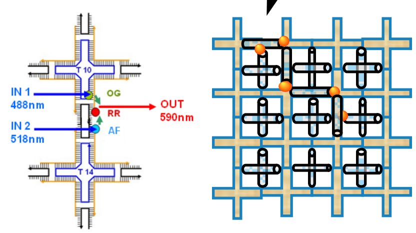
Large Scale Networks, Logical Structure & Defect Isolation
[NANOARCH 2005, 2006, Nanonets 2006, NanoCom 2009]



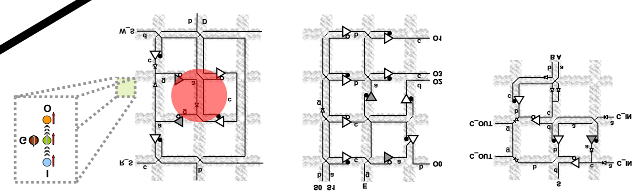
NANA - General Purpose Architecture [JETC 2006]



SOSA - Data Parallel Architecture [NANOARCH 2006, ASPLOS 2006, JETC 2007, 2009]

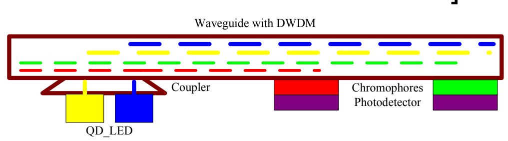


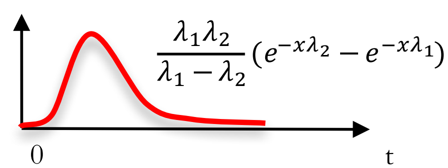
Circuit Architecture
[FNANO 2004, IEEE MICRO 2008 IEEE MICRO 2015, ISCA 2016]



Sensing & Processing
[ASPLOS 2009. IEEE MICRO 2010, Small 2010]

NoC
[NocArc 2014, JETC 2015, ASPLOS 2015]





Stochastic Computing
[IEEE MICRO 2015, ISCA 2016]