Mathematical Terms of Biological Information Theory

Thomas D. Schneider*

version = 1.11 of bitt.tex 2021 Jul 27

- For a basic short lesson in information theory, see [1].
- For more terms, see "A Glossary for Molecular Information Theory and the Delila System": `http://alum.mit.edu/www/toms/glossaryframes.html`.
- A PDF version of this document is at `http://alum.mit.edu/www/toms/papers/bitt/bitt.pdf`.
- An HTML version of this document is at `http://alum.mit.edu/www/toms/papers/bitt`.[1]

$B = \log_2 M$ (bits)	Number of bits for M symbols	[1]
$C = d_{\text{space}} \log_2(P/N+1)$ (bits/mmo)	Molecular machine capacity	[2, 3]
$D = 2d_{\text{space}} \ll 3n-6$	Dimensionality of a molecular machine coding space	[2]
$\frac{2R\ln 2}{\ln(\rho+1)} \leq D < \frac{2R\ln 2}{\epsilon_r}$	Lower and upper dimensionality bounds	[4]
$d_{\text{space}} = D/2$	Number of 'pins' a molecular machine uses	[2, 3]
ΔG (joules/mmo)	Free energy dissipated by a molecular machine in an operation	[2, 3]
$d.f. = 3n-6$	Degrees of freedom for n atoms	[2]
$E_{\min} = k_{\text{B}} T \ln 2$ (joules per bit)	A version of the Second Law of Thermodynamics that can be used as an ideal conversion factor between energy and bits	[5, 6]
$\epsilon_t = \frac{\ln\left(\frac{P}{N}+1\right)}{\frac{P}{N}} = \frac{\ln(\rho+1)}{\rho}$	Theoretical maximum molecular efficiency	[5, 6, 4]

*National Institutes of Health, National Cancer Institute, Center for Cancer Research, RNA Biology Laboratory, P. O. Box B, Frederick, MD 21702-1201. (301) 846-5581; schneidt@mail.nih.gov; `https://alum.mit.edu/www/toms/`

[1] The html was generated directly from the LaTeXsource using tex4ht: `http://www.tug.org/applications/tex4ht/mn.html`

$\epsilon_r \le \epsilon_t$	Real (or measured) molecular efficiency	[5, 6]
k_{B} (joules/kelvin)	Boltzmann's constant	
$\lambda = R/2$	Compressed bases: the number of bases a binding site would take up if the information of the site was compressed as small as possible.	
$M = 2^B$	Number of symbols corresponding to B bits	
mmo	Molecular machine operation	[2, 7]
μ	Mean of Gaussian distribution	
σ	Standard deviation of Gaussian distribution	
π	Circle circumference/radius, something to eat	
n	Number of atoms in a molecular machine. see $d.f.$	
N (joules/mmo)	Thermal noise flowing through a molecular machine during an opertion	[2, 8, 9]
$P = -\Delta G$ (joules/mmo)	Energy dissipated by a molecular machine in an opertion	[2, 3]
$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	Probabilty of x for a Gaussian distribution	
quincunx	Galton's Quincunx - a device that demonstrates the formation of a Gaussian distribution. See `http://tinyurl.com/GaltonQuincunx`	
R (bits/mmo)	Information gained during a molecular machine operation, often of a binding site	[10]
$R_{energy} \equiv -\Delta G^{\circ}/E_{\min}$ (bits per mmo)	The maximum bits that can be gained for the given free energy dissipation	[5, 6]
$\rho = P/N$	Energy dissipation of a molecular machine normalized by the thermal noise flowing through the machine	
T (K)	Absolute temperture, Kelvin	

x	Voltage (for a communications system) or total potental and kinetic energy for a molecular machine
y	See x

References

[1] T. D. Schneider. Information theory primer, with an appendix on logarithms. *Published on the web*, 2013, 2013. https://doi.org/10.13140/2.1.2607.2000, https://alum.mit.edu/www/toms/papers/primer/.

[2] T. D. Schneider. Theory of molecular machines. I. Channel capacity of molecular machines. *J. Theor. Biol.*, 148:83–123, 1991. https://doi.org/10.1016/S0022-5193(05)80466-7, https://alum.mit.edu/www/toms/papers/ccmm/.

[3] T. D. Schneider. Theory of molecular machines. II. Energy dissipation from molecular machines. *J. Theor. Biol.*, 148:125–137, 1991. https://doi.org/10.1016/S0022-5193(05)80467-9, https://alum.mit.edu/www/toms/papers/edmm/.

[4] T. D. Schneider and V. Jejjala. Restriction enzymes use a 24 dimensional coding space to recognize 6 base long DNA sequences. *PLoS One*, 14:e0222419, 2019. https://doi.org/10.1371/journal.pone.0222419, https://alum.mit.edu/www/toms/papers/lattice/.

[5] T. D. Schneider. 70% efficiency of bistate molecular machines explained by information theory, high dimensional geometry and evolutionary convergence. *Nucleic Acids Res.*, 38:5995–6006, 2010. https://doi.org/doi:10.1093/nar/gkq389, https://alum.mit.edu/www/toms/papers/emmgeo/.

[6] T. D. Schneider. A brief review of molecular information theory. *Nano Communication Networks*, 1:173–180, 2010. https://doi.org/10.1016/j.nancom.2010.09.002, https://alum.mit.edu/www/toms/papers/brmit/.

[7] T. D. Schneider. Sequence logos, machine/channel capacity, Maxwell's demon, and molecular computers: a review of the theory of molecular machines. *Nanotechnology*, 5:1–18, 1994. https://doi.org/10.1088/0957-4484/5/1/001, https://alum.mit.edu/www/toms/papers/nano2/.

[8] J. B. Johnson. Thermal agitation of electricity in conductors. *Physical Review*, 32:97–109, 1928. https://doi.org/10.1103/PhysRev.32.97.

[9] H. Nyquist. Thermal agitation of electric charge in conductors. *Physical Review*, 32:110–113, 1928. https://doi.org/10.1103/PhysRev.32.110.

[10] T. D. Schneider, G. D. Stormo, L. Gold, and A. Ehrenfeucht. Information content of binding sites on nucleotide sequences. *J. Mol. Biol.*, 188:415–431, 1986. https://doi.org/10.1016/0022-2836(86)90165-8, https://alum.mit.edu/www/toms/papers/schneider1986/.