Complexity and robustness

John Doyle (doyle@caltech.edu)
John G Braun Professor
Control and Dynamical Systems, BioEngineering, and Electrical Engineering
Caltech, Pasadena, CA

John Doyle's research builds on insights about the fundamental nature of complex biological and technological networks that can now be drawn from the convergence of three research themes. 1) Molecular biology has provided a detailed description of much of the components of biological networks, and with the growing attention to systems biology the organizational principles of these networks are becoming increasingly apparent([2], [3], [4], [5], [8], [9], [12], [13], [16], [17], [19], [20], [21], [22], [28], [30], [36] www.sbml.org). 2) Advanced technology has provided engineering examples of networks with complexity approaching that of biology. While the components differ from biology, we have found striking convergence at the network level of architecture and the role of layering, protocols, and feedback control in structuring complex multiscale modularity ([1], [8], [10], [16], [20], [25], [29], [31], [32], [39]). Our research is leading to new theories of the Internet and related networking technologies, and to new protocols that are being tested and deployed, particularly for high performance scientific computing ([6], [11], [13], [15], [18], [23], [26], www.hot.caltech.edu, netlab.caltech.edu). 3) Most importantly, there is a new mathematical framework for the study of complex networks that suggests that this apparent network-level evolutionary convergence both within biology and between biology and technology is not accidental, but follows necessarily from the requirements that both biology and technology be efficient, adaptive, evolvable, and robust to perturbations in their environment and component parts (www.cds.caltech.edu/sostools, [10], [13], [27], [33], [37], [38]). This theory builds on and integrates decades of research in pure and applied mathematics with engineering, and specifically with robust control theory.

Through evolution and natural selection or by deliberate design, such systems exhibit highly functional and symbiotic interactions of extremely heterogeneous components, the very essence of "complexity." At the same time this resulting organization allows, and even facilitates, severe fragility to cascading failure triggered by relatively small perturbations. Thus robustness and fragility are deeply intertwined in both biological and technological systems, and in fact the mechanisms that create their extraordinary robustness are also responsible for their greatest fragilities ([8], [16], [17], [21], [22]). Our highly regulated and efficient metabolism evolved when life was physical challenging and food was often scarce. In a modern lifestyle, this robust metabolism can contribute to obesity and diabetes ([21]). More generally, our highly controlled physiology creates an ideal ecosystem for parasites, who hijack our robust cellular machinery for their own purposes. Our immune system prevents most infections but can cause devastating autoimmune diseases, including a type of diabetes. Our complex physiology requires robust development and regenerative capacity in the adult, but this very robustness at the cellular level is turned against us in cancer. We protect ourselves in highly organized and complex societies which facilitate spread of epidemics and destruction of our ecosystems. We rely on ever advancing technologies, but these confer both benefits and horrors previously unimaginable. This universal "robust yet fragile'" (RYF) nature of complex systems is well-known to experts such as physicians and systems engineers, but has been systematically studied in any unified way only recently. It is now clear that it must be treated explicitly in any theory that hopes to explain the emergence of biological complexity, and indeed is at the heart of complexity itself.

These RYF features appear on all time and space scales, from the tiniest microbes and cellular subsystems up to global ecosystems, and also –we believe- to human social and technical systems, and from the oldest known history of the evolution of life through human evolution to our latest technological innovations. Typically, our networks protect us, which is a major reason for their existence. But in addition to cancer, epidemics, and chronic auto-immune disease, the rare but catastrophic market crashes, terrorist attacks, large power outages, computer network virus epidemics, and devastating fires, etc, remind us that our complexity always comes at a price. Statistics reveal that most dollars and lives lost in natural and technological disasters happen in just a small subset of the very largest events, while the typical event is so small as to usually go unreported. The emergence of complexity can be largely seen as a spiral of new challenges and opportunities which organisms exploit, but lead to new fragilities, often to novel perturbations. These are met with increasing complexity and robustness, which in turn creates new opportunities but also new fragilities, and so on. This is not an inexorable trend to greater complexity, however, as there are numerous examples of lineages evolving increasing simplicity in response to less uncertain environments. This is particularly true of parasites that rely on their hosts to control fluctuations in their microenvironment, thus shielding them from the larger perturbations that their hosts experience.

It is only fairly recently, and particularly the last few decades, that human technology has become focused not just on robustness, but on architectures that facilitate the evolution of new capabilities and the scaling to large system sizes.

Protocol-based multilayer modular design is permeating advanced technologies of all kinds, but the Internet remains perhaps the most well-known example. It is also particularly suitable for our purposes for several reasons. The Internet, and cyber- or netcentric technology generally, are unprecedented in the extent to which their features parallel biology. Their most salient features are often hidden from the user and thus as metaphors are often terribly misleading, yet are extremely useful when right. Only cybertechnology has the potential to rival "biotechnology" in accelerating the human/technology evolution, and the combined RYF spiral could have profound consequence. The most consistent, coherent, and salient features of all complex technologies are their protocols. To engineers, the term "protocol" is the set of rules by which components interact to create system-level functionality. Indeed, in advanced technologies, and we believe in the organization of cells and organisms, the protocols are more fundamental than the modules whose interconnection they facilitate, although they often are obscured by the overwhelming details that now characterize experimental results in biology. A central feature of efficient, protocol-based systems is that, provided they obey the protocols, modules can be exchanged and changed. The details are less important here than the consequences, which are the system-level robustness and evolvability that these protocols facilitate ([25]). New and even radically different hardware is easily incorporated at the lowest physical layers, and even more radically varying applications are enabled at the highest layers. Ironically as in biology, it is these transient elements of hardware and application software that are most visible to the user, while the far more fundamental and persistent infrastructure is the core protocols, which by design remain largely hidden from the user.

A protocol-based organization facilitates coordination and integration of function to create coherent and global adaptation to variations in their components and environments on a vast range of time scales despite implementation mechanisms that are largely decentralized and asynchronous. The parallels here between the Internet and biology are particularly striking. The TCP/IP protocol suite enables adaptation and control on time scale from the sub-microsecond changes in physical media, to the millisecond-to-second changes in traffic flow, to the daily fluctuations in user interactions, to evolving hardware and application modules over years and decades ([11], [14], [23], [25]). The remarkable robustness to changing circumstances and evolution of Internet-related technology could only have come about as the result of a highly structured and organized suite of relatively invariant and universally-shared, well-engineered protocols.

Similarly, a protocol-based architecture in biology and its control mechanisms facilitate both robustness and evolvability, despite massive impinging pressures and variation in the environment. With the most obvious example involving the table of codons, biology's universally shared set of protocols are more fundamental and invariant than the modules whose control and evolution they facilitate. Allostery, a huge suite of post-translational modifications, and the rapid changes in location of macromolecular modules enable adaptive responses to environmental signals or alterations on rapid time scales. Translational and transcriptional control and regulation of alternative splicing and editing act on somewhat longer time scales. On still longer time scales within and across generations, the sequences of the DNA itself can change, not only through random mutation, but also through highly structured and evolved mechanisms that facilitate the generation of adaptive diversity. Furthermore, as biologists dig deeper past the superficiality of sequence data into the complexity of regulation, they unearth additional layers of control that are fundamentally similar to those in advanced technologies ([2], [8], [16], [17], [22]). There is seemingly no limit to the ingenuity that biology uses in creating additional layers of sophisticated control. Now familiar examples range from RNA editing and alternative splicing to transposons, mismatch repair, and repetitive sequences to the cutting and pasting in the "arms race" of the immune system versus spirochete and trypanosome coat proteins.

Perhaps the most familiar example of lateral gene transfer in bacteria is possible because bacteria have a shared set of protocols that have even been quite appropriately described by some as the "bacterial Internet." Bacteria can simply grab DNA encoding new genes from other bacteria and incorporate it into their genome, just like computer users can buy a new computer and plug it into home or office networks. This "plug and play" modularity works because there is a shared set of protocols that allow even novel genetic material to be functional in an entirely new cellular setting. Plug and play DNA mobility and expression is further facilitated by integrons and plasmids. Thus, for example, bacteria can acquire antibiotic resistance on time-scales that would be vanishingly improbable by point mutations, an example of how rapid evolution of complexity is possible by Darwinian mechanisms. Natural selection can favor the evolution of whole protocol suites, and their interactions, which in turn massively accelerate the acquisition and sharing of functional adaptive change. Thus evolvability itself can be seen as the robustness of lineages, rather than organisms, on long time scales and to possibly large changes in the environment, indeed ones that would be lethal to organisms if they occurred rapidly. An important insight is that robustness and evolvability are generally not in conflict, and both are the product of systematic and organized control mechanisms.

The framework being developed here is radical in both its methodology and philosophical implication. Methodologically it draws on mathematics that is often not well known outside expert circles and in many cases had not traditionally been thought of as "applied." The mathematics tells us that robustness and fragility have conserved quantities ([8], [37]), and we believe these will ultimately be of as much importance to understanding biological complexity as energy and entropy were to understanding the steam engine and mitochondria. The above views of "organized complexity" motivated by biology and engineering contrast sharply with that of "emergent complexity" that is more popular within the physical sciences. Highly Optimized/Organized Tolerance/Tradeoffs (HOT) has aimed to explain the issues of organized complexity ([1], [10], [29], [30], [33], [35]), but emphasizing models and concepts such as lattices, cellular automata, spin glasses, phase transitions, criticality, chaos, fractals, scale-free networks, self-organization, and so on, that have been the inspiration for the "emergent" perspective. A side benefit of this largely pedagogical effort is it has led to apparently novel insights into RYF aspects of longstanding mysteries in physics, from coherent structures in shear flow turbulence and coupled oscillators, to the ubiquity of power laws, to the nature of quantum entanglement, to the origin of dissipation. Finally, the underlying mathematics may offer new tools to explore other problems in physics where RYF features may play a role, particularly involving multiple scales and organized structures and phenomena.

REFERENCES

[1] Doyle JC, Carlson JM, Power laws, highly optimized tolerance, and generalized source coding, Phys Rev Lett 84 (24): 5656-5659 JUN 12 2000

[2] Yi T-M, Huang Y, Simon MI, Doyle J. Robust perfect adaptation in bacterial chemotaxis through integral feedback control, Proc. Natl. Acad. Sci. USA 97, 4649-4653, 2000.

[3] Doyle J, Computational biology: Beyond the spherical cow, Nature 411, 151-2 (10 May 2001)

[4] Csete M, Walikonis J, Slawny N, Wei Y, Korsnes S , Doyle JC, Wold B, Oxygen-mediated regulation of skeletal muscle satellite cell proliferation and adipogenesis in culture, Journal of Cellular Physiology, Volume 189, Issue 2, Nov. 2001. 189-196

[5] Finney AM, Hucka M, Sauro HM, Doyle J, Kitano H, Bolouri H, The Systems Biology Markup Language Molecular Biology of the Cell 12: 708 Suppl. S NOV 2001

[6] Zhu X, Yu J, Doyle J, Heavy Tails, Generalized Coding, and Optimal Web Layout, Infocomm, 2001

[7] Zhou T, Carlson JM, Doyle J, Mutation, specialization, and hypersensitivity in highly optimized tolerance, Proc. Natl. Acad. Sci. USA 99 (4): 2049-2054 Feb 19 2002

[8] Csete M , Doyle J, Reverse Engineering of Biological Complexity, Science 295 (5560): 1664-1669 Mar 1 2002

[9] Morohashi M, Winn AE, Borisuk MT, Bolouri H, Doyle J, Kitano H, Robustness as a measure of plausibility in models of biochemical networks, Journal Of Theoretical Biology 216 (1): 19-30 MAY 7 2002

[10] Carlson JM, Doyle J, Complexity and robustness, P Natl Acad Sci USA 99: 2538-2545 Suppl. 1 FEB 19 2002

[11] Low SH, Paganini F, Doyle JC, Internet congestion control, IEEE Contr Syst Mag 22 (1): 28-43 FEB 2002

[12] Gilman AG et al., Overview of the Alliance for Cellular Signaling, NATURE 420 (6916): 703-706 DEC 12 2002

[13] Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, BIOINFORMATICS 19 (4): 524-531 MAR 1 2003

[14] Papachristodoulou, A; Li, L; Doyle, JC, 2004. Methodological frameworks for large-scale network analysis and design, COMPUTER COMMUNICATION REVIEW, JUL 2004

[15] Li, L., D. Alderson, J. Doyle, and W. Willinger. 2004. A First-Principles Approach to Understanding the Internet's Router-level Topology. Proc. ACM SIGCOMM, COMPUTER COMMUNICATION REVIEW, OCT. 2004.

[16] Csete M.E. and J.C. Doyle, 2004, Bow ties, metabolism, and disease, Trends in Biotechnology, Vol 22, Issue 9, pg. 446-450

[17] J. Stelling, U. Sauer, Z. Szallasi, F. J. Doyle III, and J. Doyle, 2004, Robustness of cellular functions, Cell, October, 2004.

[18] Willinger, W, D. Alderson, J.C. Doyle, and L.Li., 2004. More ``Normal'' Than Normal: Scaling Distributions and Complex Systems. Proceedings of the 2004 Winter Simulation Conference.

[19] Shapiro, BE; Hucka, M; Finney, A; Doyle, J. 2004. MathSBML: a package for manipulating SBML-based biological models, BIOINFORMATICS, NOV 1, 2004

[20] Doyle, J; Csete, M, 2004. Imitation of life: How biology is inspiring computing, NATURE. OCT 21 2004

[21] Kitano, H; Oda, K; Kimura, T; Matsuoka, Y; Csete, M; Doyle, J; Muramatsu, M, 2004 .Metabolic syndrome and robustness tradeoffs, DIABETES 53: S6-S15 Suppl. 3, DEC 2004

[22] H. El-Samad, H. Kurata , J.C. Doyle , C.A. Gross, and M. Khammash, 2005, Surviving Heat Shock: Control Strategies for Robustness and Performance, PNAS. PNAS (8): 2736-2741 FEB 22, 2005

[23] Jin C, Wei D, Low SH, Bunn J, Choe HD, Doyle JC,et al, FAST TCP: From theory to experiments IEEE NETWORK 19 (1): 4-11 JAN-FEB 2005

[24] Paganini F, Wang ZK, Doyle JC, et al. Congestion control for high performance, stability, and fairness in general networks , IEEE-ACM TRANSACTIONS ON NETWORKING 13 (1): 43-56 FEB 2005

[25] W. Willinger, J Doyle, Robustness and the Internet: Design and evolution, in Robust Design: A Repertoire of Biological, Ecological, and Engineering Case Studies (Santa Fe Institute Studies on the Sciences of Complexity), Erica Jen, Editor

[26] Wang JT, Li L, Low SH, Doyle JC. (2005) Cross-layer optimization in TCP/IP networks, IEEE-ACM TRANSACTIONS ON NETWORKING 13 (3): 582-595 JUN 2005

[27] Manning M, Carlson JM, Doyle J (2005) Highly optimized tolerance and power laws in dense and sparse resource regimes PHYSICAL REVIEW E 72 (1): Art. No. 016108 Part 2 JUL 2005

[28] R. Tanaka, T-M Yi, and J. Doyle (2005) Some protein interaction data do not exhibit power law statistics, FEBS letters, 579 (23): 5140-5144 SEP 26 2005

[29] Doyle et al, (2005), The "Robust Yet Fragile" Nature of the Internet, P Natl Acad Sci USA. vol. 102 no. 41, October 11, 2005

[30] Zhou T, Carlson JM, Doyle J (2005) Evolutionary dynamics and highly optimized tolerance, JOURNAL OF THEORETICAL BIOLOGY 236 (4): 438-447 OCT 21 2005

[31] L Li, D Alderson, JC Doyle, W Willinger (2005) Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications, Internet Math, 2005

[32] D Alderson, L Li, W Willinger, JC Doyle (2005) Understanding Internet Topology: Principles, Models, and Validation, IEEE/ACM TRANSACTIONS ON NETWORKING, Dec. 2005.

[33] R. Tanaka, M. Csete and J. Doyle, Highly optimised global organisation of metabolic networks, IEE Proceedings Systems Biology, 152 (4): 179-184 DEC 2005

[34] MA Moritz, ME Morais, LA Summerell, JM Carlson, J Doyle (2005) Wildfires, complexity, and highly optimized tolerance, P Natl Acad Sci USA, December 13, 2005; 102 (50) ,

[35] T Brookings, JM Carlson, and J Doyle (2005) Three mechanisms for power laws on the Cayley tree, Phys. Rev. E 72, 056120

[36] J. Doyle and M Csete (2005), Motifs, stability, and control, PLOS Biology, November 2005,Volume 3, Issue 11
[37] N Martins, M Dahleh, and J Doyle (2005) Fundamental Limitations of Disturbance Attenuation in the Presence of Side Information, 44th IEEE CDC/ECC 2005, December 12-15, 2005, Seville, Spain
[38] H El-Samad , A Papachristodoulou, S Prajna, J Doyle, and M Khammash (2006), Advanced Methods and Algorithms for Biological Networks Analysis, PROCEEDINGS OF THE IEEE, 94 (4): 832-853 APR 2006
[39] L. Chen, S. Low, M. Chiang, and J. Doyle. Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks. In Proc. of IEEE Infocom, 2006.
[40] Martins NC, Dahleh MA, Doyle JC (2006) Disturbance attenuation bounds in the presence of a remote preview, LECTURE NOTES IN CONTROL AND INFORMATION SCIENCES 331: 269-285 2006
[41] Kurata, H El-Samad, R Iwasaki, H Ohtake, JC Doyle, et al. (2006) Module-based analysis of robustness tradeoffs in the heat shock response system. PLoS Comput Biol 2(7): July 2006
[42] M Chiang, SH Low, AR Calderbank, JC. Doyle (2006) Layering As Optimization Decomposition, PROCEEDINGS OF THE IEEE, to appear