# Brown CS and CCMB present

SorinFest: Phase Transitions in Computer Science and Computational Biology

a conference upon the occasion of
Sorin Istrail's seventieth birthday

Born in Târgu-Neamţ, District of Moldova, Romania, Sorin began his academic journey in 1971 where he studied computer science at Universitatea „Alexandru Ioan Cuza" (University of Iași) in Iași, Romania. Under the guidance of Professors Solomon Marcus and Sergiu Rudeanu, Professor Istrail earned his PhD from the University of Bucharest in 1979 with the thesis "Context-Sensitive Languages and Applications to Program Semantics and Number Theory". He then accepted a research position in the University of Iași, Computer Center, and "Al. Myller" Mathematical Institute (Seminarul Matematic) until 1983.
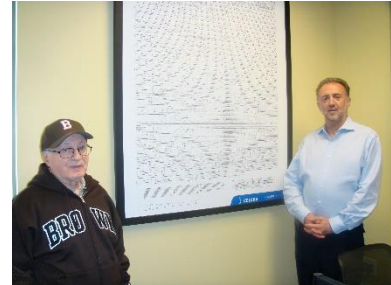
with Solomon Marcus

Professor Istrail's journey continued in America where he was an Assistant Professor of Computer Science at Wesleyan University and pursued postdoctoral studies at MIT with Professor Albert Meyer from 1984 to 1992. His research with Professor Meyer focused on axiomatizing the semantics of concurrent programs and culminated in the publication of a logical expressive power separation of Milner's Calculus of Communicating Systems and Hoare's theory of Communicating Sequential Processes (*Bisimulation can't be traced*, POPL 1988, JACM 1995).

Wesleyan colleague Professor Alan Cobham brought to him the *Universal Traversal Sequences* (UTS) problem proposed by Steve Cook, a problem of fundamental importance in complexity theory for the L vs. NL complexity classes, one of the most studied problems in computational complexity theory. Inspired by his mentor, Professor Michael Sipser (MIT), Professor Istrail provided the first construction of universal traversal sequences of 2-regular graphs in log-space, (STOC 1988, FOCS 1990). The complete proof of his algorithmic construction is long, about 70 pages. He considers this result the deepest computer science algorithm he has published. It was proved non-constructively by Aleluias, Karp, Lipton, Lovasz, and Rackoff that UTSs of length $O(d^2 n^3 \log n)$ exist (FOCS 1979). Prior constructions were "brute force" in length, which in log space are size $O(n^{\log n})$. Professor Istrail's UTS construction has size $O(n^{4.76})$, i.e., polynomial length, in log space. The general case is still open, and he continues his work there, but after 35 years, his result is still the only UTS construction for any general graph class. With Lenny Heath, inspired by VLSI circuitry design and topological graph theory, and with great interactions and mentorship support from Tom Leighton (MIT), they designed algorithms for book embedding of genus g graphs in $O(g)$ pages (on orientable and non-orientable surfaces) (STOC 1987, JACM 1992). They proved a lower bound of $O(\sqrt{g})$ and conjectured that $O(\sqrt{g})$ pages as the optimal bound. S. Malitz in his MIT PhD thesis proved non-constructively their conjecture (FOCS 1988).

In 1992, Professor Istrail joined Sandia National Labs, where he co-founded, with Ernie Brickell (Department Manager) and Dr. Fred Howes (Manager of the DOE Applied Mathematics Program, program

founded at DOE by John von Neumann), and led the Computational Biology Project until 2000. Professor Istrail's transition into computational biology began with algorithmic strategies for genome physical mapping problems, a key challenge of the Human Genome Project started at DOE by Charles DeLisi. Collaboration with Michael Waterman provided statistically rigorous foundations for the computational methods developed. Next came the protein folding problem, a grand challenge of fundamental importance in biology, biochemistry, biophysics, and biotechnology. In joint work with Bill Hart, they developed the first approximation algorithms with mathematical guaranteed error bounds for a protein folding model (STOC 1995, RECOMB 1997); the algorithm applied to one of the most studied lattice biophysical protein folding models, the HP-model (Hydrophobic-Polar model) of Ken Dill. Professor Istrail's search for an exactly solvable protein folding model continued for almost 30 years. He also continued his research on the thermodynamic hypothesis and the protein folding energy function inference using methods from voting theory and mathematical economics.

Bill Camp, Professor Istrail's division director at Sandia, brought to him the Three-Dimensional Ising Model problem and became his mentor on statistical mechanics of the 3D Ising Model. In 2000, Professor Istrail obtained the negative solution (computational intractability) of this 50-year-old unresolved problem in statistical mechanics, which asked for the derivation of the analytical closed form of the partition function of the 3D Ising Model (STOC 2000). Professor Istrail's paper on the 3D Ising Model was included in the Top 100 Most Important Discoveries of the U.S. Department of Energy (DOE)'s first 25 years, and the 7th top achievement of DOE in

Pavel Pevzner, Mike Waterman, Sorin Istrail

Advanced Scientific Computing. In 1997, with Professors Pavel Pevzner and Mike Waterman, Professor Istrail co-founded RECOMB (the International Annual Conference on Research in Computational Molecular Biology) in Santa Fe, New Mexico. RECOMB 2023, now in its 27th year, was held in Istanbul, Turkey.

After Sandia, Professor Istrail served as the Senior Director, and later Head of the Informatics Research Department, at Celera Genomics from 2000 to 2005. He was coming from Sandia Labs, which sits on an Air Force base in Albuquerque, New Mexico, where decades earlier John von Neumann and a dream team of scientists worked on the extraordinary Manhattan Project. And now he found himself at Celera Genomics in Rockville, Maryland, among a dream team of scientists pursuing another bigger-than-life scientific endeavor – the Celera Genome Project. Working closely with Celera's president Craig Venter was a life-changing experience, profound for the Informatics Research team. Craig trained and inspired the team to master the art of making the impossible possible in genomics. "Algorithms are the make-or-break of Celera," he said. How fortunate he feels to have graduated from the Craig Venter School of Leadership! Professor Istrail's group played a major role in constructing the computational assembly of the human genome and co-authored the 2001 Science paper "The Sequence of the Human Genome," which, with over 20,000 citations to date, is one of the most cited scientific papers. His group also built a powerful suite of genome-assembly-to-genome-assembly alignment algorithms that was used to compare all human genome assemblies to date in the paper "Whole Genome Shotgun Assembly and Comparison of Human Genome Assemblies" (PNAS 2004). Working with Gene Myers, leader of the Informatics Research Department, was for him most exciting – a very supportive, extremely valuable experience. In 2002, his Celera group, in collaboration with the company ClearForrest, won the ACM KDD Cup – the top

international data mining competition. The challenge then was the automatic annotation of a section of the Drosophila genome.

His Celera group also developed algorithms for SNPs and haplotype inference and analysis, with applications to population genetics, SNP assay design, and pharmacogenomics. In setting the research agenda of the SNPs and Haplotypes Team, the collaboration with Andy Clark (Cornell University), a Brown University alumnus and one of the leading researchers in human genetics was essential and inspiring. Collaboration with Michael Waterman added strength and rigor to the developed algorithmic strategies. In 2021, their Celera 2003 paper "SNPs problems, complexity and algorithms" received the 20-years Test-of-Time Award from the European Symposium of Algorithms. It pioneered the field of computational haplotyping, employing rigorous computer science methods. In collaboration with Jonathan Yewdell (NIH), his Celera group published an important viral genomics paper "Comparative Imunnopeptidomics of Humans and Their Pathogens" (PNAS 2004). In 2003, he joined the ranks of Applied Biosystems Science Fellows, Informatics.

Authors of the PNAS 2004 paper. (standing) Bjarni Halldorson, Aaron Halperin, Hagit Shatkay, Liliana Florea, Ross Lippert, Nathan Edwards, Granger Sutton, Russell Turner, Celera Genomics President Craig Venter, Sorin, Karin Remington, Ian Dew; (sitting) Clark Mobarry, Brian Walenz, Jason Miller, Merissa Henry. Sadly, our dear Celera colleagues Karin Remington and Brown University PhD Hagit Shatkay passed away. We miss them very much.

Professor Istrail joined the Department of Computer Science at Brown University in 2005, where he is the James A. and Julie N. Brown Professor of Computational and Mathematical Sciences and Professor of Computer Science. During his time at Brown, he served as the Director of Brown University's Center for Computational Molecular Biology (2006-2011) and continued as a visiting senior scientist at Caltech with his long-time mentor, and closest friend, Professor of Biological Sciences, Eric Davidson. His work with Professor Davidson over 15 years (2000-2015) made important contributions to constructing the sea urchin genome and transcriptome assemblies (Science 2006, Science 2006), and developing a mathematical information processing model for gene regulatory systems and networks, in the papers "Logic Functions of the Genomic cis-Regulatory Code" (PNAS 2005), and "The Regulatory Genome and the Computer" (Dev. Biol. 2007), a paper coinciding with the 50th anniversary of John von Neumann's seminal book "The Computer and the Brain." Written in the same style as von Neumann's book, the 2007 paper presented a comparison between every component of the electronic computer and its counterpart in the Regulatory Genome's information processing "computer" embedded in the molecular biology of the cell. Eric Davidson co-advised Professor Istrail's PhD student Ryan Tarpine, who wrote a thesis entitled "A database of causality-inferred structure-function information for genomic *cis*-regulatory architecture" (Brown University 2012).

with Eric Davidson

Sorin Istrail, Eric Davidson and Ellen Rothenberg, June 2015.

with Eric Davidson and Ryan Tarpine

Sorin Istrail in Eric Davidson's Sea Urchin lab.

Professor Istrail's legacy continues through the undergraduate, graduate, and postdoctoral students he has mentored and who have gone on to make their own contributions to their respective fields, including Dejan Živković (Serbia), Ryan Tarpine (Google), Derek Aguiar (UConn), Doug McErlean (Google), Austin Huang (Google DeepMind), Fumei Lam, Alper Uzun (Brown), Younhun Kim (MIT), Shivam Nadimpalli (Columbia), and Pinar Demetci (MIT). His work with Dr. Fumei Lam developed a new linkage disequilibrium measure that satisfied long-standing desiderata that had not been simultaneously fulfilled previously: the curse of the pairwise: conservative extension from pairwise to multiple markers, and the interpretability of intermediate values. Along with his student Dr. Derek Aguiar, he introduced the problem of assembling the haplotypes of genomes with three or more chromosomal copies and developed HapCompass, a leading tool for haplotype assembly. Most recently, Professor Istrail has developed a globally optimal algorithm for computing maximum likelihood haplotype phasing by graph homomorphism and a general optimization theory for posynomial optimization with his student Douglas McErlean and Dr. Aguiar. With his PhD student Pinar Demetci and Dr. Aguiar, the trio developed algorithmic strategies for combinatorial and statistical prediction of gene expression from haplotype sequence.
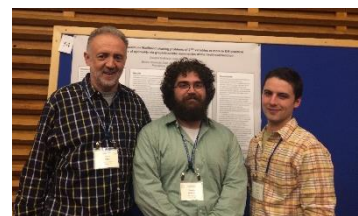

with Fumei Lam   *with Pinar Demetci*   *with Derek Aguiar at the CS Halloween party*   *with Doug McErlean and Derek Aguiar*

Professor Istrail has Erdos number 2 and postcard-Erdos number 1. In the PhD adviser genealogy, he has 9 links to Euler and 11 links to Leibniz.