

08w5094 Analytic Tools in Computational Complexity

Paul Beame (U Washington),
Steve Cook (U Toronto),
Russell Impagliazzo (UC San Diego & IAS),
Valentine Kabanets (SFU),
Avi Wigderson (IAS)

August 3 – August 8, 2008

1 Overview of the Field

Computational Complexity Theory is the field that studies the inherent costs of algorithms for solving mathematical problems. Its major goal is to identify the limits of what is efficiently computable in natural computational models. Computational complexity ranges from quantum computing to determining the minimum size of circuits that compute basic mathematical functions to the foundations of cryptography and security.

Computational complexity emerged from the combination of logic, combinatorics, information theory, and operations research. It coalesced around the central problem of “P versus NP” (one of the seven open problems of the Clay Institute). While this problem remains open, the field has grown both in scope and sophistication. Currently, some of the most active research areas in computational complexity are

- the study of hardness of approximation of various optimization problems (using probabilistically checkable proofs), and the connections to coding theory,
- the study of the role of randomness in efficient computation, and explicit constructions of “random-like” combinatorial objects,
- the study of the power of various proof systems of logic, and the connections with circuit complexity and search heuristics,
- the study of the power of quantum computation.

2 Recent Developments

An important development in the study of computational complexity has been increased role of analytic methods. Fourier analysis has become an essential tool of the field, playing a critical role in the study of interactive proofs, the computational hardness of approximation problems, and the learnability of Boolean functions. The notion of Gowers uniformity (which was introduced by Gowers to give an analytic proof of Szemerédi’s theorem on arithmetic progressions, and whose use can be viewed as “generalized Fourier analysis”) has also been recently employed in the context of Probabilistically Checkable Proofs and hardness of approximation. A new paradigm in computational complexity is beginning to emerge, which involves reducing high dimensional discrete problems that arise in the study of Boolean functions to high dimensional continuous problems and then applying analytic methods to the resulting continuous problems.

3 Presentation Highlights

Avi Wigderson gave a tutorial on many applications of partial derivatives in complexity. Alexander Sherstov gave a tutorial on the matrix sign rank, which has found so many recent applications to communication complexity lower bounds.

In addition to these two tutorials, the workshop had a great number of excellent results presented. One of the most exciting results presented at the workshop was by Zeev Dvir, giving a simple and elegant solution to the finite-field version of Kakeya Conjecture (with an application to pseudorandomness).

However, Dvir's presentation was just one out of a large number of exciting new developments in computational complexity. The following is a summary of the presented results, grouped by topic.

3.1 PCPs

SUBHASH KHOT, **Inapproximability of NP -complete Problems, Discrete Fourier Analysis, and Geometry**

This talk is an overview of some recent results on inapproximability of NP -complete problems and their connections to discrete Fourier analysis and isoperimetric problems.

DANA MOSHKOVITZ, **Two-Query PCP with Sub-Constant Error** (based on joint work with Ran Raz)

We show that the NP -Complete language 3Sat has a PCP verifier that makes two queries to a proof of almost-linear size and achieves sub-constant probability of error $o(1)$. The verifier performs only projection tests, meaning that the answer to the first query determines at most one accepting answer to the second query.

Previously, by the parallel repetition theorem, there were PCP Theorems with two-query projection tests, but only (arbitrarily small) constant error and polynomial size. There were also PCP Theorems with sub-constant error and almost-linear size, but a constant number of queries that is larger than 2.

As a corollary, we obtain a host of new results. In particular, our theorem improves many of the hardness of approximation results that are proved using the parallel repetition theorem. A partial list includes the following:

1. 3Sat cannot be efficiently approximated to within a factor of $7/8 + o(1)$, unless $P = NP$. This holds even under almost-linear reductions. Previously, the best known NP -hardness factor was $7/8 + \epsilon$ for any constant $\epsilon > 0$, under polynomial reductions (Hastad).
2. 3Lin cannot be efficiently approximated to within a factor of $1/2 + o(1)$, unless $P = NP$. This holds even under almost-linear reductions. Previously, the best known NP -hardness factor was $1/2 + \epsilon$ for any constant $\epsilon > 0$, under polynomial reductions (Hastad).
3. A PCP Theorem with amortized query complexity $1 + o(1)$ and amortized free bit complexity $o(1)$. Previously, the best known amortized query complexity and free bit complexity were $1 + \epsilon$ and ϵ , respectively, for any constant $\epsilon > 0$ (Samorodnitsky and Trevisan).
4. Clique cannot be efficiently approximated to within a factor of $n^{1-o(1)}$, unless $ZPP = NP$. Previously, a hardness factor of $n^{1-\epsilon}$ for any constant $\epsilon > 0$ was known, under the assumption that $P = NP$ does not hold (Hastad and Zuckerman).

One of the new ideas that we use is a new technique for doing the composition step in the (classical) proof of the PCP Theorem, without increasing the number of queries to the proof. We formalize this as a composition of new objects that we call Locally Decode/Reject Codes (LDRC). The notion of LDRC was implicit in several previous works, and we make it explicit in this work. We believe that the formulation of LDRCs and their construction are of independent interest.

RAN RAZ, **A Counterexample to Strong Parallel Repetition**

I will give a short introduction to the problem of parallel repetition of two-prover games and its applications to theoretical computer science, mathematics and physics. I will then describe a recent counterexample to the strong parallel repetition conjecture (a conjecture that would have had important applications).

The parallel repetition theorem states that for any two-prover game, with value $1 - \epsilon$ (for, say, $\epsilon < 1/2$), the value of the game repeated in parallel n times is at most $(1 - \epsilon^3)^{\Omega(n/s)}$, where s is the answers' length

(of the original game). Several researchers asked whether this bound could be improved to $(1 - \epsilon)^{\Omega(n/s)}$; this question is usually referred to as the strong parallel repetition problem. A positive answer would have had important applications. We show that the answer for this question is negative.

More precisely, we consider the *odd cycle game* of size m ; a two-prover game with value $1 - 1/2m$. We show that the value of the odd cycle game repeated in parallel n times is at least $1 - (1/m) \cdot O(\sqrt{n})$. This implies that for large enough n (say, $n \geq \Omega(m^2)$), the value of the odd cycle game repeated in parallel n times is at least $(1 - 1/4m^2)^{O(n)}$.

Thus:

1. For parallel repetition of general games: the bounds of $(1 - \epsilon^c)^{\Omega(n/s)}$ given in [4, 2] are of the right form, up to determining the exact value of the constant $c \geq 2$.
2. For parallel repetition of XOR games, unique games and projection games: the bounds of $(1 - \epsilon^2)^{\Omega(n)}$ given in [1] (for XOR games) and in [3] (for unique and projection games) are tight.
3. For parallel repetition of the odd cycle game: the bound of $1 - (1/m) \cdot \tilde{\Omega}(\sqrt{n})$ given in [1] is almost tight.

A major motivation for the recent interest in the strong parallel repetition problem is that a strong parallel repetition theorem would have implied that the *unique game conjecture* is equivalent to the *NP* hardness of distinguishing between instances of Max-Cut that are at least $1 - \epsilon^2$ satisfiable from instances that are at most $1 - (2/\pi) \cdot \epsilon$ satisfiable. Our results suggest that this cannot be proved just by improving the known bounds on parallel repetition.

ANUP RAO, **Rounding Parallel Repetitions of Unique Games** (based on joint work with Boaz Barak, Moritz Hardt, Ishay Haviv, Oded Regev and David Steurer).

We show a connection between the semidefinite relaxation of a unique game and its behavior under parallel repetition. Specifically, denoting by $val(G)$ the value of a two-prover unique game G , and by $sdpval(G)$ the value of a natural semidefinite program to approximate $val(G)$, we prove that for every l , if $sdpval(G) \geq 1 - \delta$, then $val(G^l) \geq 1 - O(\sqrt{l\delta(\log k - \log \delta)})$. Here, G^l denotes the l -fold parallel repetition of G , and k denotes the alphabet size of the game. For the special case where G is an XOR game (i.e., $k = 2$), we obtain the bound $1 - O(\sqrt{l\delta})$. For games with a significant gap between the quantities $val(G)$ and $sdpval(G)$, our result implies that $val(G^l)$ may be much larger than $val(G)^l$, giving a counterexample to the strong parallel repetition conjecture.

In a recent breakthrough, Raz has shown such an example using the max-cut game on odd cycles. Our results are based on a generalization of his techniques.

RYAN O'DONNELL, **Zwick's Conjecture is implied by most of Khot's Conjectures** (based on joint work with Yi Wu)

In 1998 Zwick proved that $P = naPCP_{1,5/8}(O(\log n), 3)$ and conjectured that $naPCP_{1,5/8+\epsilon}(O(\log n), 3) = NP$ for all $\epsilon > 0$. Hastad's contemporary result $NP = naPCP_{1,3/4+\epsilon}(O(\log n), 3)$ was not improved until 2006, when Khot and Saket lowered the $3/4$ to $20/27$. We prove Zwick's $5/8$ Conjecture, assuming Khot's " d -to-1 Conjecture" for any constant d . The necessary Long Code testing analysis uses a mix of older (Hastad-type) and new (Mossel-type) ideas.

GUY KINDLER, **Can cubic tiles be sphere-like?** (based on joint work with Ryan O'Donnell, Anup Rao, and Avi Wigderson)

The unit cube tiles R^d by Z^d , in the sense that its translations by vectors from Z^d cover R^d . It is natural to ask what is the minimal surface area of a body that tiles R^d by Z^d . The volume of any such body should clearly be at least 1, and therefore its surface area must be at least that of a unit volume ball, which of order \sqrt{d} . The surface area of the cube, however, is of order d , and no better tiling was known. In this work we use a random construction to show that the optimal surface area is indeed of order \sqrt{d} , namely there exist bodies that tile R^d as a cube would, but have sphere-like surface areas.

Tiling problems were considered for well over a century, but this particular tiling problem was also recently considered in computer science because of its relations with the Unique Games conjecture and with the Parallel Repetition problem. Indeed, our current result follows from the same idea that was used recently by Raz in his counter example for the strong Parallel Repetition conjecture.

3.2 Quantum computation

ODED REGEV, **Unique Games with Entangled Provers are Easy** (based on joint work with Julia Kempe and Ben Toner)

We consider one-round games between a classical verifier and two provers who share entanglement. We show that when the constraints enforced by the verifier are “unique” constraints (i.e., permutations), the value of the game can be well approximated by a semidefinite program. Essentially the only algorithm known previously was for the special case of binary answers, as follows from the work of Tsirelson in 1980. Among other things, our result implies that the variant of the unique games conjecture where we allow the provers to share entanglement is false. Our proof is based on a novel “quantum rounding technique”, showing how to take a solution to an SDP and transform it to a strategy for entangled provers.

SCOTT AARONSON, **How To Solve Longstanding Open Problems In Quantum Computing Using Only Fourier Analysis**

I’ll discuss some simple-looking conjectures in Fourier analysis of Boolean functions which, if proved, would lead to breakthrough results in quantum complexity theory. These potential breakthroughs include an oracle relative to which BQP is not in the polynomial hierarchy, and the impossibility of a random oracle separation between BPP and BQP . I’ll also discuss the partial progress that I and others have been able to make toward resolving these conjectures.

3.3 Communication complexity

ALEXANDER SHERSTOV, **The Sign-Rank of AC^0** (based on joint work with Alexander Razborov)

We prove that $\Sigma_2^{cc} \not\subseteq UPP^{cc}$, thereby solving a long-standing open problem in communication complexity posed by Babai, Frankl, and Simon (1986).

In more detail, the sign-rank of a matrix $M = [M_{ij}]$ with $+/-1$ entries is defined as the least rank of a real matrix $A = [A_{ij}]$ with $M_{ij}A_{ij} > 0$ for all i, j . We prove a lower bound of $2^{\Omega(m)}$ on the sign-rank of the matrix $[f(x, y)]_{x, y}$, where $f(x, y) = \bigwedge_{i=1}^m \bigvee_{j=1}^{m^2} (x_{ij} \wedge y_{ij})$. This is the first exponential lower bound on the sign-rank of AC^0 , and it immediately implies the separation $\Sigma_2^{cc} \not\subseteq UPP^{cc}$.

Our result additionally implies a lower bound in learning theory. Specifically, let $\phi_1, \dots, \phi_r : \{0, 1\}^n \rightarrow \mathbb{R}$ be functions such that every DNF formula $f : \{0, 1\}^n \rightarrow \{+1, -1\}$ of polynomial size has the representation $f = \text{sign}(a_1 * \phi_1 + \dots + a_r * \phi_r)$ for some reals a_1, \dots, a_r . We prove that then $r > 2^{\Omega(n^{1/3})}$, which essentially matches an upper bound of $2^{\tilde{O}(n^{1/3})}$ due to Klivans and Servedio (2001).

Finally, our work yields the first exponential lower bound on the size of threshold-of-majority circuits computing a function in AC^0 . This substantially generalizes and strengthens the results of Krause and Pudlak (1997).

PAUL BEAME, **Multiparty Communication Complexity of AC^0** (based on joint work with Dang-Trinh Huyuh-Ngoc)

We prove non-trivial lower bounds on the multiparty communication complexity of AC^0 functions in the number-on-forehead (NOF) model for up to $\Theta(\sqrt{\log n})$ players. These are the first lower bounds for any AC^0 function for $\omega(\log \log n)$ players. In particular we show that there are families of depth 3 read-once AC^0 formulas having k -player randomized multiparty NOF communication complexity $n^{\Omega(1/k)} / 2^{O(k)}$. We show similar lower bounds for depth 4 read-once AC^0 formulas that have nondeterministic communication complexity $O(\log^2 n)$, yielding exponential separations between k -party nondeterministic and randomized communication complexity for AC^0 functions. As a consequence of the latter bound, we obtain a $2^{\Omega(\log^1 / 2n / k^{1/2}) - k}$ lower bound on the k -party NOF communication complexity of set disjointness. This is non-trivial for up to $\Theta(\log^{1/3} n)$ players which is significantly larger than the up to $\Theta(\log \log n)$ players allowed in the best previous lower bounds for multiparty set disjointness.

3.4 New directions

MADHU SUDAN, **Towards Universal Semantic Communication** (based on joint work with Brendan Juba)

Is it possible for two intelligent players to communicate meaningfully with each other, without any prior common background? What does it even mean for the two players to understand each other? In addition to being an intriguing question in its own right, we argue that this question also goes to the heart of modern communication infrastructures, where misunderstandings (mismatches in protocols) between communicating players are a major source of errors. We believe that questions like this need to be answered to set the foundations for a robust theory of (meaningful) communication.

In this talk, I will describe what computational complexity has to say about such interactions. Most of the talk will focus on how some of the nebulous notions, such as intelligence and understanding, should be defined in concrete settings. We assert that in order to communicate “successfully”, the communicating players should be explicit about their goals – what the communication should achieve. We show examples that illustrate that when goals are explicit the communicating players can achieve meaningful communication.

BOAZ BARAK, Public Key Cryptography from Different Assumptions (based on joint work with Avi Wigderson)

We construct a new public key encryption based on two assumptions:

1. One can obtain a pseudorandom generator with small locality by connecting the outputs to the inputs using any sufficiently good unbalanced expander.
2. It is hard to distinguish between a random graph that is such an expander and a random graph where a (planted) random logarithmic-sized subset S of the outputs is connected to fewer than $|S|$ inputs.

The validity and strength of the assumptions raise interesting new algorithmic and pseudorandomness questions, and we explore their relation to the current state-of-art.

3.5 Pseudorandomness and explicit combinatorial constructions

ZEEV DVIR, The finite field Kakeya conjecture and applications to the construction of mergers and extractors (joint work with Avi Wigderson)

A Kakeya set in F^n , where F is a finite field, is a set containing a line in every direction. The finite field Kakeya conjecture states that the size of such sets is bounded from below by $C_n * |F|^n$, where C_n depends only on the dimension n .

The interest in this problem came first from Mathematics as a finite field analog of the famous Euclidean Kakeya problem and later from Computer Science as a problem related to explicit constructions of mergers and extractors.

I will talk about the recent proof of this conjecture (Dvir 2008) and its application to the construction of mergers and extractors (Dvir & Wigderson 2008).

SHACHAR LOVETT, Worst case to average case reductions for polynomials (based on joint work with Tali Kaufman)

A degree- d polynomial p in n variables over a field F is equidistributed if it takes on each of its $|F|$ values close to equally often, and biased otherwise. We say that p has a low rank if it can be expressed as a bounded combination of polynomials of lower degree. Green and Tao (2007) have shown that bias implies low rank over large fields (i.e. for the case $d < |F|$). They have also conjectured that bias implies low rank over general fields. In this work we affirmatively answer their conjecture. Using this result we obtain a general worst case to average case reductions for polynomials. That is, we show that a polynomial that can be approximated by few polynomials of bounded degree, can be also exactly computed by few polynomials of bounded degree. We derive some relations between our results to the construction of pseudorandom generators.

EMANUELE VIOLA, Hardness amplification proofs require majority (based on joint work with Ronen Shaltiel)

Hardness amplification is a major line of research that mainly seeks to transform a given lower bound (e.g. a function that has correlation at most 99% with small circuits) into a strongly average-case one (i.e. a function that has negligible correlation with small circuits). Strongly average-case lower bounds are of central importance in complexity theory and in particular are necessary for most cryptography and pseudorandom generators.

In this work we show that standard techniques for proving hardness amplification against a class of circuits require that same class of circuits to compute the Majority function.

Our work is most significant when coupled with the celebrated “natural proofs” result by Razborov and Rudich (J. CSS ’97) and Naor and Reingold (J. ACM ’04), which shows that most lower-bounding techniques cannot be applied to circuits that can compute Majority. The combination of our results with theirs shows that *standard techniques for hardness amplification can only be applied to those circuit classes for which standard techniques cannot prove circuit lower bounds*. This in particular explains the lack of strong average-case lower bounds for a number of circuit classes for which we have lower bounds.

Our results also show a qualitative difference between the direct product lemma and Yao’s XOR lemma, and they give tight bounds on the number of queries needed for hardness amplification.

RONEN SHALTIEL, **Unconditional weak derandomization of weak algorithms: Explicit versions of Yao’s Lemma**

The (easy direction) of Yao’s minmax lemma says that if there is a randomized algorithm A which solves some problem (meaning that for every input, A succeeds with high probability) then there is a deterministic algorithm B of “roughly the same complexity” that solves the problem well on average (meaning that B succeeds with high probability on a random input). This can be viewed as “weak derandomization” and the statement follows by an averaging argument: there exist a fixed value r for A ’s random coins such that hardwiring r into A produces the deterministic algorithm B . Note that this averaging argument does not provide an explicit way to find r .

Recently, Zimand (building on an approach by Goldreich and Wigderson) proved an explicit version of the implication for randomized decision trees which toss “few” random coins. In this work, we consider weak derandomization of various classes of randomized algorithms.

We give a different proof of Zimand’s result. Our proof generalizes to any class of randomized algorithms as long as one can explicitly construct an appropriate randomness extractor. Using this approach we prove unconditional weak derandomization results for communication games, constant depth circuits and streaming algorithms. More precisely we show that:

1. Given a randomized communication protocol that tosses few random coins and assuming that this protocol is explicitly constructible (in the sense that players can compute their strategy in polynomial time). Then, there is an explicitly constructible deterministic communication protocol of comparable communication complexity that simulates the randomized protocol correctly on most inputs.
2. Given a randomized algorithm that can be implemented by a uniform family of poly-size constant depth circuits we construct a uniform family of deterministic poly-size constant depth circuits that succeed on most inputs. (A classic result by Nisan and Wigderson gives a deterministic circuit that succeeds on all inputs but has quasi-polynomial size).

Our techniques follow the approach of Goldreich and Wigderson in the sense that we also “extract randomness from the input”. However, in contrast to previous papers we use seedless extractors rather than seeded ones. We use extractors for bit-fixing sources (for decision trees) 2-source extractors (for communication games and streaming algorithms) and PRG based extractors (for constant depth circuits).

LUCA TREVISAN, **Dense Subsets of Pseudorandom Sets** (based on joint work with Omer Reingold, Madhur Tulsiani, and Salil Vadhan)

A theorem of Green, Tao, and Ziegler can be stated (roughly) as follows: if R is a pseudorandom set, and D is a dense subset of R , then D may be modeled by a set M that is dense in the entire domain such that D and M are indistinguishable. (The precise statement refers to “measures” or distributions rather than sets.) The proof of this theorem is very general, and it applies to notions of pseudorandomness and indistinguishability defined in terms of any family of distinguishers with some mild closure properties. The proof proceeds via iterative partitioning and an energy increment argument, in the spirit of the proof of the weak Szemerédi regularity lemma. The “reduction” involved in the proof has exponential complexity in the distinguishing probability. We present a new proof inspired by Nisan’s proof of Impagliazzo’s hardcore set theorem. The reduction in our proof has polynomial complexity in the distinguishing probability and provides a new characterization of the notion of “pseudoentropy” of a distribution. We also follow the connection between the two theorems and obtain a new proof of Impagliazzo’s hardcore set theorem via iterative partitioning and

energy increment. While our reduction has exponential complexity in some parameters, it has the advantage that the hardcore set is efficiently recognizable.

3.6 Error-correcting codes

DAVID ZUCKERMAN, **List-Decoding Reed-Muller Codes Over Small Fields** (based on joint work with Parikshit Gopalan and Adam Klivans)

We present the first local list-decoding algorithm for the r th order Reed-Muller code $RM(r, m)$ over F_2 for $r > 1$. Given an oracle for a received word $R : F_2^m \rightarrow F_2$, our randomized local list-decoding algorithm produces a list containing all degree r polynomials within relative distance $2^{-r} - \epsilon$ from R for any $\epsilon > 0$ in time $poly(m^r, \epsilon^{-r})$. The list size could be exponential in m at radius 2^{-r} , so our bound is optimal in the local setting. Since $RM(r, m)$ has relative distance 2^{-r} , our algorithm beats the Johnson bound for $r > 1$.

In the setting where we are allowed running-time polynomial in the block-length, we show that list-decoding is possible up to even larger radii, beyond the minimum distance. We give a deterministic list-decoder that works at error rate below $J(2^{1-r})$, where $J(d)$ denotes the Johnson radius for minimum distance d . This shows that $RM(2, m)$ codes are list-decodable up to radius s for any constant $s < 1/2$ in time polynomial in the block-length.

Over small fields F_q , we present list-decoding algorithms in both the global and local settings that work up to the list-decoding radius. We conjecture that the list-decoding radius approaches the minimum distance (like over F_2), and prove this when the degree is divisible by $q - 1$.

RAGESH JAISWAL, **Uniform Direct Product Theorems** (based on joint work with Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson)

Direct Product Theorems are formal statements of the intuition: “if solving one instance of a problem is hard, then solving multiple instances is even harder”. For example, a Direct Product Theorem with respect to bounded size circuits computing a function is a statement of the form: “if a function f is hard to compute on average for small size circuits, then $f^k(x_1, \dots, x_k) = f(x_1), \dots, f(x_k)$ is even harder on average for certain smaller size circuits”. The proof of the such a statement is by contradiction: we start with a circuit which computes f^k on some non-negligible fraction of the inputs and then use this circuit to construct another circuit which computes f on almost all inputs. By viewing such a constructive proof as decoding certain error-correcting code, it was independently observed by Trevisan and Impagliazzo that constructing a single circuit is not possible in general. Instead, we can only hope to construct a list of circuits such that one of them computes f on almost all inputs. This makes the list size an important parameter of the Theorem which can be minimized. We achieve optimal value of the list size which is a substantial improvement compared to previous proofs of the Theorem. In particular, this new version can be applied to uniform models of computation (e.g., randomized algorithms) whereas all previous versions applied only to nonuniform models (e.g., circuits).

3.7 Computational learning

ROCCO SERVEDIO, **Testing Fourier dimensionality and sparsity** (based on joint work with Parikshit Gopalan, Ryan O’Donnell, Amir Shpilka and Karl Wimmer)

We present a range of new results for testing properties of Boolean functions that are defined in terms of the Fourier spectrum. Broadly speaking, our results show that the property of a Boolean function having a concise Fourier representation (or any sub-property thereof) is locally testable.

We first give an efficient algorithm for testing whether the Fourier spectrum of a Boolean function is supported in a low-dimensional subspace of F_2^n (equivalently, for testing whether f is a junta over a small number of parities). We next give an efficient algorithm for testing whether a Boolean function has a sparse Fourier spectrum (small number of nonzero coefficients). In both cases we also prove lower bounds showing that any testing algorithm — even an adaptive one — must have query complexity within a polynomial factor of our algorithms, which are nonadaptive. Finally, we give an “implicit learning” algorithm that lets us test any sub-property of Fourier concision.

Our technical contributions include new structural results about sparse Boolean functions and new analysis of the pairwise independent hashing of Fourier coefficients from (Feldman et al., FOCS 2006).

ADAM KLIVANS, **Agnostically Learning Decision Trees** (based on joint work with Parikshit Gopalan and Adam Kalai)

We give a query algorithm for agnostically learning decision trees with respect to the uniform distribution on inputs. Given black-box access to an arbitrary binary function f on $\{0, 1\}^n$, our algorithm finds a function that agrees with f on almost (within an ϵ fraction) as many inputs as the best fitting size- t decision tree in time $\text{poly}(n, t, 1/\epsilon)$. This is the first polynomial-time algorithm for learning decision trees in a harsh noise model. The core of our learning algorithm is a procedure to implicitly solve an ℓ_1 minimization problem in 2^n dimensions using an approximate gradient projection method.

3.8 Cryptography

SALIL VADHAN, **Why Simple Hash Functions Work: Exploiting the Entropy in a Data Stream** (based on joint work with Michael Mitzenmacher (SODA 2008) plus improvements with Kai-Min Chung (RANDOM 2008, “Tight Bounds for Hashing Block Sources”))

Hashing is fundamental to many algorithms and data structures widely used in practice. For theoretical analysis of hashing, there have been two main approaches. First, one can assume that the hash function is truly random, mapping each data item independently and uniformly to the range. This idealized model is unrealistic because a truly random hash function requires an exponential number of bits to describe. Alternatively, one can provide rigorous bounds on performance when explicit families of hash functions are used, such as 2-universal or $O(1)$ -wise independent families. For such families, performance guarantees are often noticeably weaker than for ideal hashing.

In practice, however, it is commonly observed that simple hash functions, including 2-universal hash functions, perform as predicted by the idealized analysis for truly random hash functions. In this paper, we try to explain this phenomenon. We demonstrate that the strong performance of universal hash functions in practice can arise naturally from a combination of the randomness of the hash function and the data. Specifically, following the large body of literature on random sources and randomness extraction, we model the data as coming from a “block source,” whereby each new data item has some “entropy” given the previous ones. As long as the (Renyi) entropy per data item is sufficiently large, it turns out that the performance when choosing a hash function from a 2-universal family is essentially the same as for a truly random hash function. We describe results for several sample applications, including linear probing, balanced allocations, and Bloom filters.

3.9 Computer algebra

CHRIS UMANS, **Fast polynomial factorization and modular composition in any characteristic** (based on joint work with Kiran Kedlaya)

We give an algorithm for modular composition of degree n univariate polynomials over a finite field F_q requiring $O(n \log q)^{1+o(1)}$ bit operations. As an application, we obtain a randomized algorithm for factoring degree n polynomials over F_q requiring $O(n^{1.5} \log q + n \log^2 q)^{1+o(1)}$ bit operations, improving upon the methods of von zur Gathen & Shoup (1992) and Kaltofen & Shoup (1998). Our results also imply algorithms for irreducibility testing and computing minimal polynomials whose running times are best-possible, up to lower order terms.

The first step is to reduce modular composition to certain instances of multipoint evaluation of multivariate polynomials. We then give an algorithm that solves this problem optimally (up to lower order terms). The main idea is to lift to characteristic 0, apply a small number of rounds of multimodular reduction, and finish with a small number of multidimensional FFTs. The final evaluations are then reconstructed using the Chinese Remainder Theorem. As a bonus, we obtain a very efficient data structure supporting polynomial evaluation queries, which is of independent interest.

Our algorithm uses techniques which are commonly employed in practice, so it may be competitive for real problem sizes. This contrasts with previous asymptotically fast methods relying on fast matrix multiplication.

3.10 General complexity theory

MARIO SZEGEDY, **Long Codes and the the Dichotomy Conjecture for CSPs** (based on joint work with Gabor Kun)

The well known dichotomy conjecture states that for every family of constraints the corresponding CSP is either polynomially solvable or NP-hard. We establish a three ways connection between the conjecture, asymptotic behavior of iterated maps related to non-linear dynamical systems, and the type of Fourier analytic techniques used in the theory of Probabilistically Checkable Proofs and property testing. To demonstrate the power of our newly found connections we give an analytic proof for the Hell-Nešetřil theorem about the dichotomy for undirected graphs.

We also contribute to the theory of non-linear dynamical systems by giving a characterization of multivariate functions (maps) over a finite domain that, when iterated sufficiently many times, become arbitrarily “resilient” to changing any fixed number of inputs in an arbitrary manner.

Finally, we obtain a new characterization of algebras that avoid types 1 and 2.

MARK BRAVERMAN, **The complexity of simulating Brownian Motion**

We analyze the complexity of the Walk on Spheres algorithm for simulating Brownian Motion in a domain Ω in R^d . The algorithm, which was first proposed in the 1950s, produces samples from the hitting probability distribution of the Brownian Motion process on boundary of Ω within an error of ϵ . The algorithm is used as a building block for solving a variety of differential equations, including the Dirichlet Problem.

The WoS algorithm simulates a BM starting at a point $X_0 = x$ in a given bounded domain Ω until it gets ϵ -close to the boundary of Ω . At every step, the algorithm measures the distance d_k from its current position X_k to the boundary of Ω and jumps a distance of $d_k/2$ in a uniformly random direction from X_k to obtain X_{k+1} . The algorithm terminates when it reaches X_n that is ϵ -close to the boundary of Ω .

It is not hard to see that the algorithm requires at least $\Omega(\log 1/\epsilon)$ steps to converge. Only partial results with respect to upper bounds existed. In 1959 M. Motoo established an $O(\log 1/\epsilon)$ bound on the running time for convex domains. The results were later generalized for a wider, but still very restricted, class of planar and 3-dimensional domains by G.A. Mikhailov (1979). In our earlier work (2007), we established an upper bound of $O(\log^2 1/\epsilon)$ on the rate of convergence of WoS for arbitrary planar domains.

We introduce subharmonic energy functions to obtain very general upper bounds on the convergence of the algorithm. Special instances of the upper bounds yield the following results for bounded domains Ω :

- if Ω is a planar domain with connected exterior, the WoS converges in $O(\log 1/\epsilon)$ steps;
- if Ω is a domain in R^3 with connected exterior, the WoS converges in $O(\log^2 1/\epsilon)$ steps;
- for $d > 2$, if Ω is a domain in R^d , the WoS converges in $O((1/\epsilon)^{2-4/d})$ steps;
- for $d > 3$ if Ω is a domain in R^d with connected exterior, the WoS converges in $O((1/\epsilon)^{2-4/(d-1)})$ steps;
- for any d if Ω is a domain in R^d bounded by a smooth surface, the WoS converges in $O(\log 1/\epsilon)$ steps.

We also demonstrate that the bounds are tight, i.e. we construct a domain from each class for which the upper bound is exact. Our results give the optimal upper bound of $O(\log 1/\epsilon)$ in many cases for which only a bound polynomial in $1/\epsilon$ was previously known.

References

- [1] U. Feige, G. Kindler, and R. O’Donnell. Understanding parallel repetition requires understanding foams. In *Proceedings of the Twenty-Second Annual IEEE Conference on Computational Complexity*, pages 179–192, 2007.
- [2] T. Holenstein. Parallel repetition: Simplifications and the no-signaling case. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, pages 411–419, 2007.

- [3] A. Rao. Parallel repetition in projection games and a concentration bound. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, pages 1–10, 2008.
- [4] R. Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.