AUTOMATING COMBINATORICS: PROJECT DESCRIPTION

Doron Zeilberger

Department of Mathematics, Rutgers University (New Brunswick), Hill Center-Busch Campus, 110 Frelinghuysen Rd., Piscataway, NJ 08854-8019. zeilberg@math.rutgers.edu
http://www.math.rutgers.edu/~zeilberg/ .

Summary of Results from Previous NSF Support: DMS-9732602 and DMS 0100403

1. The current NSF award number is DMS-0100403 for the period 2001-2004, totaling $180,000 .

2. Its title is: "*Symbolic Computation and Combinatorics*".

3. Summary of the results of the completed work

(The numbered references apply to the list of papers written with the NSF support of the above grants, given at section 4. The lettered references are to papers listed in "References Cited".)

My students, collaborators, and I continued our efforts in extending and developing the so-called *Wilf-Zeilberger Algorithmic Proof Theory* (exposited in [PWZ]). This was accomplished in papers [1][3][23][28]. WZ theory, in particular the so-called *Zeilberger algorithm*, is the subject of investigation by researchers in Austria (the RISC-Linz group lead by Peter Paule), Canada (Keith Geddes and his student Ha Le), France (the INRIA group lead by Frederic Chyzak), Germany (Wolfram Koepf), Holland (lead by Tom Koornwinder), Russia (S. Abramov), and other countries. It now, along with Gosper's algorithm, has its own *Mathscinet* Subject Classification: 33F10.

In my last proposal I proposed to develop a new algorithmic theory called 'The Umbral Transfer Matrix Method' based on Gian-Carlo Rota's seminal notion of *umbra*, that realizes its full scope when interfaced with computer algebra. This resulted, so far, in five papers ([15][19][21][22][31]), that contain applications to famous combinatorial enumeration problems that arose in statistical physics, namely the counting of lattice animals and self-avoiding walks. While I did not solve these notorious problems (which most probably do not have any closed-form solution), the method was able to automatically derive what I called 'umbral schemes', which are systems of linear differential-functional equations that enable enumerations of important subfamilies of these infamous 'physical' creatures, that also shed light on the general objects themselves. Currently, my Ph.D. student, Mohamud Mohammed is applying this method to ('toy models' of) other important combinatorial families and problems: The Ising model (with a magnetic field) and percolation.

In classical combinatorics, in collaboration with Dominique Foata, we found a combinatorial proof of Bass's evaluation of the Ihara Zeta function[7]. This paper contains some 'lemmas' that turned out to be significant in knot theory. It also lead knot-theorists Stavros Garoufalidis and TQ Lê to ask me to prove their conjectured 'Quantum analog' of the celebrated MacMahon Master Theorem[29]. MacMahon was very proud of his 'master' theorem and included it in the *Combinatorial Analysis* entry that he wrote for the 11th edition (1910) of Encyclopedia Brittanica. I believe that this is but the tip of an iceberg, and will hopefully lead to 'Quantum WZ-theory' to be described below. Meanwhile, Garoufalidis[Ga] found far-reaching implications of q-WZ theory to *colored Jones polynomials.*

Other papers on classical combinatorics are [17] and [25] (with D. Foata), and [24] (with Aaron Robertson and Dan Saracino). This paper lead to interesting work by Igor Pak and Sergei Elizalde[EP].

In [10] I proved a conjecture of Clara Chan and the late David Robbins that turned out to be important in geometry, and was further elaborated by Michele Vergne and Welleda Baldoni-Silva[BV].

Paper [12] was expanded into a 'futuristic web-book' entitled 'Shalosh B. Ekhad's 2050 Geometry Textbook' that may be viewed at my website. The accompanying software is a free download.

In [18], and its sequel [30], I study the combinatorial game Chomp, as a case-study in automated proving based on the ansatz of 'ultimate-periodicity'. Inspired by [18], my student Xinyu Sun conjectured the ultimate-periodicity phenomenon, that was proved brilliantly by high-school-student Steven Byrnes (see section 7 below).

Papers [23], [26],[27], and [28] are methodological papers that attempt to develop a preliminary framework for 'rigorous experimental mathematics'.

Paper [4] solved a $100 dollar problem of Ron Graham about the maximum number of monochromatic Schur triples in a 2-coloring of the first N integers.

Paper [5] broke a twenty-year record for the lower bound for the asymptotic growth of the number of ternary square-free words. Using an extension of my method, this record was subsequently improved by Uwe Grimm, and Grimm's record, in turn, was improved by my student Xinyu Sun.

4. List of Publications resulting from the NSF awards in 1998-2003

1. (With T. Amdeberhan) *q-Apery Irrationality Proofs by q-WZ Pairs*, Adv. Appl. Math. **20**, 275-283, (1998).

2. *How Much Should a Nineteenth-Century French Bastard Inherit*, J. Difference Eq. Appl. **3**, 385-388 (1998).

3. (With Scott Ahlgren, Shalosh B. Ekhad, Ken Ono) *A Binomial Coefficient Identity Associated to a Conjecture of Beukers*, Electronic Journal of Combinatorics (`http://www.combinatorics.org`) **5**, R10 (1 page) (1998).

4. (With Aaron Robertson) *A 2-Coloring of [1,N] Can Have* $(1/22)N^2 + O(N)$ *Monochromatic Schur Triples, But Not Less!*, Electronic Journal of Combinatorics (`http://www.combinatorics.org`) **5** R19, (5 pages) (1998).

5. (With Shalosh B. Ekhad) *There Are More Than 2**(n/17) n-Lettered Ternary Square-Free Words*, J. Integer Sequences (elec.) (`http://www.research.att.com/ njas/jit/`), **98.1.9** (3 pages) (1998).

6. *Enumeration Schemes, and More Importantly, Their Automatic Generation*, Annals of Combinatorics **2**, 185-195 (1998).

7. (With D. Foata) *Combinatorial Proofs of Bass's Evaluations of the Ihara-Selberg Zeta function*

of a Graph, Trans. Amer. Math. Soc., **351**, 2257–2274 (1999).

8. *Automated Counting of LEGO Towers*, J. Difference Eq. Appl., **5**, 323-333, (1999).

9. (With J. Noonan) *The Goulden-Jackson Cluster Method: Extensions, Applications, and Implementations*, J. Difference Eq. Appl. **5**, 355-377, (1999).

10. *Proof Of A Conjecture Of Chan, Robbins, and Yuen*, ETNA, (Elec. Trans, of Numerical Analysis) (elec.) (`http://etna.mcs.kent.edu/`) **9**, 147-148, (1999).

11. (With Aaron Robertson and Herb Wilf) *Patterns and Fractions*, Elec. J. Combinatorics, **6**, (`http://www.combinatorics.org`) R38, (4 pages) (1999).

12. (With S. B. Ekhad) *PLANE GEOMETRY: An Elementary School Textbook (ca. 2050)*, Mathematical Intelligencer **21(3)**, 64-70, (1999).

13. *Symbol-Crunching with the Transfer-Matrix Method in Order to Count Skinny Physical Creatures*, INTEGERS (`http://www.integers-ejcnt.org`), **0** A9 (29 pages) (2000).

14. (With A. Edlin) *The Goulden-Jacskon Cluster Method For Cyclic Words*, Advances in Applied Mathematics **25**, 228-232, (2000).

15. *The Umbral Transfer-Matrix Method: I. Foundations*, J. Comb. Theory Ser. A **91**, 451-463, (Rota memorial issue) (2000).

16. *How Berger, Felzenbaum, and Fraenkel Revolutionized COVERING SYSTEMS The Same Way that George Boole Revolutionized LOGIC*, Elect. J. Combinatorics 8(2) (special issue in honor of Aviezri Fraenkel), A1 (`http://www.combinatorics.org`) , (9 pages), (2001) .

17. (With D. Foata) *Babson-Steingrimsson Statistics Are Indeed Mahonian (and Sometimes Even Euler-Mahonian)*, Adv. Appl. Math. **27**, 390-404, (2001).

18. *Three-Rowed CHOMP*, Adv. Appl. Math. **26**, 168-179, (2001).

19. *The Umbral Transfer-Matrix Method. III. Counting Animals*, New York J of Mathematics **7** , 223-231, (2001).

20. (With T. Amdeberhan) *Determinants Through The Looking Glass*, Adv. Appl. Math. **27**, 225-230. (2001).

21. *The Umbral Transfer-Matrix Method. IV. Counting Self-Avoiding Polygons and Walks*, Elec. J. Comb. **8(1)** , (22 pages) R28, (2001).

22. *The Umbral Transfer-Matrix Method. V. The Goulden-Jackson Cluster Method for Infinitely Many Mistakes*, INTEGERS, **2** (10 pages), A5 , (2002).

23. *Computerized Deconstruction*, Adv. Applied Math. **30** , 633-654, (2003).

24. (with Aaron Robertson, Dan Saracino) *Refined Restricted Permutations*, Annals of Combina-

torics. **6** , 427-444, (2003).

25. (with D. Foata) *The Collector's Brotherhood Problem Using the Newman-Shepp Symbolic Method*, Algebra Universalis (special Rota memorial issue), (10 pages), to appear.

26. *"Real" Analysis is a Degenerate Case of Discrete Analysis*, Proc. ICDEA 2001, Bernd Aulbach, ed., Taylor and Francis, to appear.

27. *Liebe Opa Paul, Ich Bin Auch Ein Experimental Scientist*, Adv. Applied Math., to appear.

28. *Towards a SymbolicComputational Philosophy (and Methodology!) for Mathematics*, Proc. of the BuchbergerFest, (Peter Paule, ed.), Springer, to appear.

29. *The Quantum MacMahon Master Theorem* (with Stavros Garoufalidis and Thang TQ Le), submitted.

30. *Chomp, Recurrences, and Chaos*, submitted.

31. *The Umbral Transfer Matrix Method II: Counting Plane Partitions*, `http://www.math.rutgers.edu/~zeilberg/pj.html`.

5. Many of my papers are accompanied by Maple packages that are available, free of charge, from my homepage `http://www.math.rutgers.edu/~zeilberg/`. In addition, there are quite a few packages that belong to forthcoming papers, or stand by themselves. Some of them are of a rather general scope, and should be useful to researchers in combinatorics, number theory, analysis, statistical physics, and possibly other areas.

The web-book 'Shalosh B. Ekhad's Geometry Text', written entirely in Maple, and fully illustrated, should be of interest to students of geometry at all levels, as well as students of computer algebra, since its succinct Maple procedures are good examples of efficient programming.

6. A large part of the proposed research is a direct continuation of the previous research, but there are also new directions, in which the connection is less obvious.

7. Education and Human Resources Statement

In 1999, Aaron Robertson and Melkamu Zeleke, who were each supported for one year by previous grants, graduated. Aaron accepted a tenure-track appointment at Colgate University. I am very pleased that he is continuing to do significant research, both computational and theoretical, in Ramsey theory and permutation enumeration. He will be up for promotion and tenure next year, and his prospects are very good. He just finished writing a book (coauthored with Bruce Landman) called 'Ramsey Theory on the Integers' to be published by the American Mathematical Society.

Melkamu Zeleke is currently a tenure-track assistant professor at William Patterson University, and is also up for promotion and tenure. Melkamu is collaborating very nicely with colleagues in bijective combinatorics, as well as being a first-rate teacher.

In 2000, Akalu Tefera and Anne Edlin, received their Ph.D. under my direction. Akalu holds a tenure-track assistant professorship at Grand Valley State University (in Michigan), and Anne is

assistant professor at Holy Family College (New Jersey).

Akalu Tefera developed sophisticated software for the continuous WZ method, available from his website. Anne Edlin extended and implemented the Goulden-Jackson method to cyclic words ([14]).

Currently I have four Ph.D. students, two at Rutgers: Mohamud Mohammed and Vince Vatter, and two at Temple: Xinyu Sun and Xiangdong Wen. They are all doing very well. Mohamud works on extending the Umbral Transfer-Matrix method to other combinatorial families, and Vince works on forbidden patterns and so-called Wilf classes. They each have several papers. Xinyu Sun expects to graduate in 2004, and his Ph.D. thesis includes significant results on combinatorial games and ternary square-free words. Xiangdong Wen works on computerized search and *proof* of (symbolic!, i.e. general) Symmetric Chain Decompositions for lattices, as well as a far-reaching extension of the Goulden-Jackson algorithm for *arbitrarily* large alphabets, and *arbitrarily* large set of *forbidden factors*, depending on finitely many parameters.

The research conducted by my students involves large-scale computing that should yield software of wide interest and applicability to the mathematical community. I am hence requesting student support, on the level of one student per year, that would be split between my students, and would free them from some teaching obligations, and hence will enable them to graduate sooner.

I am the local expert on computer algebra. Since 1988, I have been teaching both graduate and undergraduate courses that were very well attended, in using Maple and Mathematica to do research in mathematics. Since most of the graduate students that attend my classes are also teaching assistants, this know-how gets transmitted to the undergraduates.

More recently I have started teaching a graduate course called *Experimental Mathematics* that in addition to making them 'computer-algebra whizes' and skilled and sophisticated Maple programmers, also implicitly introduces them to the methodology of doing computer experiments to rigorously solve open problems.

Speaking of 'Experimental Mathematics', this fall a new seminar by that name has been inaugurated at Rutgers (organized by myself and Andrew Sills), that particularly encourages graduate students to participate, and we hope to instill in them the love of this new area.

In Xinyu Sun's research on the celebrated combinatorial game Chomp, he made an interesting conjecture concerning 'ultimate-periodicity', that was proved by high-school-student Steven Byrnes (who is now a Freshman at Harvard), and that won Byrnes the first prize (a $100,000 scholarship) at the Siemens-Westinghaus talent search.

Doron Zeilberger

PROPOSED RESEARCH: AUTOMATING COMBINATORICS

The Big Picture: Experimental Rigorous Mathematics

There is a delicate balance and trade-off between the *general* and the *special*, the *abstract* and the *concrete*, the *strategical* and the *tactical*, the *sacred* and the *profane*. If you just do your daily chores as a research mathematician, without a 'vision', you will probably not go very far. On the other hand, if you are a 'visionary', but are unable or unwilling to perform 'normal science', your vision will be just empty speculations and philosophizing. In other words, pardon the cliché, one must *think globally but act locally.*

My own *shtick* is *experimental mathematics*, but not in the conventionally understood way of computer-aided formulation and testing of conjectures, but rather in finding methods to 'teach' the computer to discover and *rigorously* prove new (and interesting!) theorems, that formerly needed humans to prove, and in most cases are of such complexity that only computers can handle them.

Quantum WZ-theory

While the Wilf-Zeilberger *Algorithmic Proof Theory* ([WZ][PWZ]), that has been implemented in all the major Computer Algebra systems (for example the package `SumTools` in Maple), can handle a very large class of identities, it is still a small beginning.

One possible, and I believe, promising, extension is to Quantum Algebra. This should not be confused with *q-series*, that WZ theory can already handle. Of course there is an intimate relationship, as was shown by Koelink[Ko] and others, who gave a fascinating quantum-group interpretation of the celebrated Askey-Wilson[AW] polynomials, that trickles down to many other families of orthogonal polynomials.

But, from an 'algorithmic' point of view, I believe that it would be fruitful to consider 'quantum identities' *ab initio*. What I have in mind are identities involving indeterminates that do not commute in the usual way, but rather obey quantum-commutation rules.

The simplest binomial-coefficient identity is the good old binomial theorem:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \quad ,$$

where we assume that x and y are commuting variables. But if one imposes the *quantum commutation rule* $yx = qxy$, then one has instead the *quantum binomial theorem* (not to be confused with the *q-binomial theorem*), that was stated and proved in 1953 by Marco Schützenberger,

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k}_q x^k y^{n-k} \quad ,$$

where $\binom{n}{k}_q$ is the *q-binomial* coefficient $(q)_n/((q)_k(q)_{n-k})$ (and $(q)_r = (1-q)(1-q^2)\cdots(1-q^r)$).

Once known, the proof is trivially analogous to the trivial proof of the binomial theorem, but I believe that there are lots of elegant *quantum* identities that are much deeper. One such was found

by Ludwig Faddeev[F], one of the pioneers of quantum groups and algebra ([FRT]). If one spells out Faddeev's quantum-identity one gets a q-identity, that could be proved automatically using q-WZ theory, *but* it would be nice to have an algorithmic framework that works directly in quantum algebra. There may also be identities that are not reducible to q-identities, even in principle.

What one needs is to develop an 'elimination theory' that in addition to the built-in non-commutativity of D_x and x (that satisfy $D_x x = x D_x + 1$), and the non-commutativity of the shift operator N, defined by $Nf(n) := f(n+1)$, with n (that satisfy $Nn = nN + n$), the other variables (that formerly were commuting parameters) satisfy the quantum commutation relations.

My confidence in the feasibility of Quantum WZ-theory arises from my recent article 'The Quantum MacMahon Master Theorem'[29] coauthored with S. Garoufalidis and TQ Lê.

Recall that in r-dimensional *quantum algebra* we have r indeterminates x_i ($1 \leq i \leq r$), satisfying the commutation relations $x_j x_i = q x_i x_j$ for all $1 \leq i < j \leq r$. We also consider matrices $A = (a_{ij})$ of r^2 indeterminates a_{ij}, $1 \leq i, j \leq r$, that commute with the x_i's and such that for any 2 by 2 minor of (a_{ij}), consisting of rows i and i', and columns j and j' (where $1 \leq i < i' \leq r$, and $1 \leq j < j' \leq r$), writing $a := a_{ij}, b := a_{ij'}, c := a_{i'j}, d := a_{i'j'}$, we have the *commutation relations*: $ba = qab, \quad ca = qac, \quad db = qbd, \quad dc = qcd, \quad cb = bc, \quad da = ad + (q - q^{-1})bc$. We will call such matrices A *quantum matrices*.

Also recall that the *quantum determinant*, of any (not-necessarily quantum) r by r matrix $B = (b_{ij})$ may be defined by

$$det_q(B) := \sum_{\pi \in S_r} (-q)^{-inv(\pi)} b_{\pi_1 1} b_{\pi_2 2} \cdots b_{\pi_r r} \quad ,$$

where the sum ranges over the set of permutations S_r, of $\{1, \ldots, r\}$, and for any of its members π, $inv(\pi)$ denotes the number of pairs $1 \leq i < j \leq r$ for which $\pi_i > \pi_j$.

In [29] the following theorem was proved:

Quantum MacMahon Master Theorem. Fix a quantum matrix A of size r. For $1 \leq i \leq r$, let $X_i := \sum_{j=1}^{r} a_{ij} x_j$, and for any vector $(m_1, \ldots, m_r)$ of non-negative integers let $G(m_1, \ldots, m_r)$ be the coefficient of $x_1^{m_1} x_2^{m_r} \ldots x_r^{m_r}$ in $\prod_{i=1}^{r} X_i^{m_i}$. Then

$$\sum_{m_1, m_2, \ldots, m_r = 0}^{\infty} G(m_1, \ldots, m_r) = 1/det_q(I - A).$$

Now it is possible to 'translate' this theorem to a 'q-(multisum)-binomial' identity, but the translation is artificial. Keeping the original quantum formulation enables a very elegant proof. I hope that Quantum WZ theory will discover and prove much deeper identities, In particular, it would be very interesting to discover quantum analogs of the constant term expressions featuring in my proof of the Alternating Sign Matrix Conjecture [Z1], that might also give generalizations of the refined conjecture[Z2] (see Dave Bressoud's beautiful book [Br] for a very lucid exposition).

Automated Determinant Evaluation

In [27] I developed an algorithm, based on Dodgson's *determinant condensation*, for automatically discovering, and at the same time proving, explicit (symbolic) evaluations of Hankel and Toeplitz matrices.

For example, my Maple package **CLD** discovered from scratch and *proved* the (well-known) explicit evaluation for the $n \times n$ Hilbert matrix:

$$\det\left(\frac{1}{i+j-1}\right) = \prod_{i=1}^{n-1} \frac{i!^4}{(2i+1)!(2i)!} \quad .$$

The basis for the algorithm was what I called the *hyperhypergeometric* ansatz, which for ordinary sequences $b(n)$ means that the 'ratio of ratios' $(b(n)/b(n-1))/(b(n-1)/b(n-2))$ is a rational function of n, and for double sequences $A(n,k)$ analogous conditions plus obvious compatibility conditions. Admittedly, this is a rather restricted ansatz, and hence the program is not *guaranteed* to give you an answer, that often does not exist (within the ansatz of hyperhypergeometric sequences).

However, with an appropriate relaxation of the ansatz, allowing much more general entries, it is hoped that one can device much more general automated results. Very interesting work on closed-form determinant-evaluations was done by George Andrews (making use of his amazing LDU factorization method), (e.g. [A1][A2]), Christian Krattenthaler (e.g. [Kr]), and Petkovsek and Wilf([PW]). Andrews, Krattenthaler, Petkovsek and Wilf use computer-algebra extensively, but still in its capacity, as George Andrews would put it, 'pencil with power-steering'. I hope to extract from their human methods algorithms that would make this subject fully *computer-generated* rather than merely *computer-assisted*.

Automated Symbolic Moment Calculus

The 'discrete Fubini trick' (a.k.a. changing the order of summation) is very useful for computing averages of combinatorial quantities defined over combinatorial families.

To take a very simple example, consider the set of permutations S_n on $\{1,\ldots,n\}$, and denote by $fix(\pi)$ the number of fixed points of π, i.e. $fix(\pi)$ equals the number of i such that $\pi_i = i$. For example $fix(3214) = 2$, $fix(2143) = 0$, and $fix(1234) = 4$. We are interested in the *expectation* (alias average) , $E[fix]$.

The *hard* way of computing this is by first computing for $0 \leq i \leq n$, the number of permutations on $\{1,\ldots,n\}$ with exactly i fixed points, let's call this quantity $A(i,n)$, and then evaluating the sum

$$\frac{1}{n!}\sum_{i=0}^{n} iA(i,n) \quad .$$

But a much better way is as follows. (For a statement A, [A] equals 1 or 0 according to whether it is true or false respectively).

$$E[fix] := \frac{1}{n!} \sum_{\pi \in S_n} fix(\pi) = \frac{1}{n!} \sum_{\pi \in S_n} \sum_{i=1}^{n} [\pi_i = i] \quad .$$

Now *change the order of summation!*, and get that

$$E[fix] = \frac{1}{n!} \sum_{i=1}^{n} \sum_{\pi \in S_n} [\pi[i] = i] = \frac{1}{n!} \sum_{i=1}^{n} \sum_{\substack{\pi \in S_n \\ \pi_i = i}} 1 = \frac{1}{n!} \sum_{i=1}^{n} (n-1)! = \frac{1}{n!} n(n-1)! = 1 \quad .$$

So the answer is 1.

The same 'trick' can be used to show that the average number of occurrences of the pattern 123 (i.e. the number of triples of places $1 \leq i < j < k \leq n$ with $\pi_i < \pi_j < \pi_k$) is $\binom{n}{3}/6$.

But what about the variance? and higher moments? Very soon the 'trick' loses its simplicity and one gets bogged down with complex (though still finitely-many) sigmas. Now it is time to invite the computer, and 'teach' it to do these tricks. In the process, fascinating new combinatorial structures emerge, and the task of computing the k-th moment boils down to *weighted-enumeration* of a finite but huge (exponential in k) set. It is probably infeasible to have an expression for arbitrary k, but already computing the first four moments gives lots of insight and information, in particular, it yields the *kurtosis* $(E[(X - E[x])^4]/Var(X)^4)$. Furthermore, if one is only interested in the *asymptotic* (in n) behavior, then it might be possible to keep k general (i.e. symbolic), at least in some cases of interest.

I propose to first apply this method to Wilf-class enumeration, that is for computing the moments (and all the possible covariances) for the random variables 'number of occurrences of a specified pattern'. The computer should be able to give rather precise answers. It is also hoped that one may get insight into the still open Stanley-Wilf conjecture, that for any given pattern, the number of permutations avoiding that pattern is of exponential growth. Perhaps this extra generality of considering all permutations, and computing moments, will help in its solution. Speaking of the Stanley-Wilf conjecture, Miklos Bona[B1][B2] proved it for the important family of *layered patters.* I hope to combine his ideas with the present ones. A few years ago Noga Alon and Ehud Friedgut[AF] 'almost' proved the Stanley-Wilf conjectured, but the full conjecture remains open.

Other candidates for applying the Automated Symbolic Moment Calculus are to Ramsey theory, where the random variables are 'number of monochromatic K_k' defined on the set of r-colorings of the edges of K_n, that hopefully will yield sharper (asymptotic) bounds for the Ramsey numbers $R_r(k)$, and 'number of monochromatic arithmetical sequences of length k' defined on r-colorings of the first n integers, that might yield sharper (asymptotic) bounds on the van der Waerden numbers. The powerful probabilistic method (see Alon and Spencer's classic book [AS]) only treats the expectation and occasionally variance, but shies away from higher moments, because of their immense complexity. It would be also of interest to apply this approach to *self-avoiding walks* and *square-free words.*

Enumeration Schemes

The subject of 'pattern-avoidance' pioneered by Frank Schmidt and the late Rodica Simion [SiSc], Herb Wilf (see [Wi]) and Richard Stanley[St], is currently a very active field, with far-ranging applications (e.g. Sara Billey and Greg Warrington's [BW] beautiful application to the Kazhdan-Lusztig polynomials, see also under *permutation* in the index (p. 573) of Richard Stanley's magnum opus 'Enumerative Combinatorics', vol. 2, [St]). I propose to continue the work of automating this research, that was started in [6].

Suppose that we have to find a 'formula' (in the sense of Wilf, i.e. a polynomial-time algorithm) for computing $a_n := |A_n|$, where A_n is an infinite family of finite sets, parameterized by n. Usually A_n is a natural subset of a larger set B_n, and is defined as the set of members of B_n that satisfy a certain set of conditions C_n. For example if A_n is the set of permutations on $\{1, 2, \ldots, n\}$, then B_n may be taken as the set of words of length n over the alphabet $\{1, 2, \ldots, n\}$, and C_n can be taken as the condition: 'no letter can appear twice'. A naive algorithm for enumerating A_n (that only works for numerical (and small) n) would be to actually *construct* the set, by examining the members of B_n, one by one, checking whether they satisfy C_n, and admitting those that qualify. Then a_n=Cardinality of A_n.

But a much better approach would be to find a *structure* theorem that expresses A_n, using unions, Cartesian products, and possibly complements, of well known sets. Failing this, it would be also nice to express A_n, recursively, in terms of $A_{n-1}, A_{n-2}, \ldots$, and easy-to-count sets, getting a *recurrence formula.* Going back to the permutation example, Levi Ben Gerson proved the structure theorem $A_n \equiv \{1, 2, \ldots, n\} \times A_{n-1}$, from which he deduced the *recurrence* $a_n = na_{n-1}$, enabling a polynomial-(in fact linear-) time algorithm for computing a_i, for $1 \leq i \leq n$.

Alas, this is not always easy, and for many enumeration sequences, e.g. the number of self-avoiding walks, may well be *impossible*, and who knows, perhaps one day even *provably impossible.*

It is conceivable, however, that a combinatorial family $A(n)$, does not possess a recursive structure by itself, but by *refining it*, using a suitable parameter, one can partition $A(n)$ into the disjoint union:

$$A(n) = \bigcup_{i=1}^{n} B(n,i) \quad ,$$

and try and find a *structure theorem* for the two-parameter family $B(n,i)$. This will imply a recurrence for the cardinalities $b(n,i) := |B(n,i)|$, that would enable a fast algorithm for $a(n) = \sum_{i=1}^{n} b(n,i)$.

Sometimes, not even the $B(n,i)$ suffice. Then we could try to partition $B(n,i)$ into the following disjoint union:

$$B(n,i) = \bigcup_{j=1}^{i-1} C_1(n,i,j) \cup \bigcup_{j=i+1}^{n} C_2(n,i,j) \quad ,$$

and try to express $C_1(n,i,j)$ and $C_2(n,i,j)$ in terms of $A(m)$, $B(m,i')$, $C_1(m,i',j')$, and $C_2(m,i',j')$, with $m < n$. One can keep going *indefinitely.* If this process *halts* after a finite number of refinements, then we have indeed a *formula* (in the sense of Wilf) for $a(n)$.

In [6] I showed how to enumerate Wilf classes, but the 'success rate' was only about 20 percents. I

propose to refine and enhance the notion of *enumeration schemes* to make the success rate closer to 100 percents. I also hope to use the same circle of ideas in enumerating *words*, special kinds of *graphs*, and other combinatorial families.

Recently my student Vince Vatter[V] found an interesting connection between the above approach and that of *generating trees* developed extensively by Julian West [We]. By combining the two methods he found much more efficient algorithms for the case where the enumerating series is a rational function. We believe that combining the two approaches will yield further significant new results.

Another interesting problem is that of *solving* the enumeration schemes, in terms of generating functions, automatically, of course.

Finally, I'd like to conclude with a 'long shot' favorite of mine: A 'linguistic' approach to the Four-Color-Theorem. I believe that it is important for everyone to invest *some* of their time on 'hard problems', whose prior probability of (full) success is very small, since regardless of whether or not they would achieve the purported goal, the *effort* is likely to bring forth new *ideas* whose significance might even surpass that goal, whether or not it is achieved.

A Linguistic Approach to the Four-Color Theorem

In Robin Thomas's[T] charming article on the Four Color Theorem, he describes three intriguing, deceptively simple, equivalent formulations of the Four Color Theorem. One of them is due to Yuri Matiyasevich[M], that shows that 4CT is equivalent to a certain innocent statement about divisibility of some products of binomial coefficients. Another is due to Dror Bar-Natan[Ba] and states that if a certain statement is true for $sl(2)$ than it is true for $sl(N)$. But my favorite is Lou Kauffman's[Ka], that is phrased in terms of 'associations' and the vector cross-product in R^3 with the unit vectors $\mathbf{i}, \mathbf{j}, \mathbf{k}$.

However, 'cross-product' smacks of calculus and geometry. It is easy to see that Kauffman's formulation can be rephrased in terms of context-free grammars as follows.

Consider the grammar consisting of the non-terminals **0, 1, 2**, and the terminals 0,1,2, with the starting symbol **0** and the derivations

$$\mathbf{0} \to \mathbf{12}|\mathbf{21}|0$$

$$\mathbf{1} \to \mathbf{02}|\mathbf{20}|1$$

$$\mathbf{2} \to \mathbf{01}|\mathbf{10}|2 \quad . \qquad (LOU)$$

Let L be the language generated by this grammar. For every complete binary tree T, let $\mathcal{W}(T)$ be the set of words in L whose derivation tree has shape T. It is easy to see that the labels of the internal nodes are uniquely determined by T. If the grammar would have been *unambiguous* then we would have had that for *any* pair of trees T_1 and T_2 with the same number of leaves, $\mathcal{W}(T_1) \cap \mathcal{W}(T_2) = \emptyset$, because otherwise there would be a word with non-unique parsing. Kauffman's equivalent formulation of 4CT is equivalent to the fact that the grammar (LOU) is, *totally ambiguous*, i.e. for *any* pair of complete binary trees with the same number of leaves, we have $\mathcal{W}(T_1) \cap \mathcal{W}(T_2) \neq \emptyset$.

I find the above formulation much more attractive than the original statement about the 4-colorability of plane graphs, that contains topological, non-combinatorial notions. Of course, thanks to Kuratowsky's theorem, the notion of 'plane graph' is equivalent to a purely combinatorial one, but it is a bit complicated. What can be simpler than complete binary trees? Also complete binary trees are so amenable to recursion, and so useful in so many algorithms.

I propose to introduce analogous notions to that of *configuration*, *reducibility* and *unavoidable set* in this new context, but hopefully the new proof, while still computer-assisted, will be much more natural and streamlined. In fact, I want the computer to do *everything* from scratch, once the human programmed the simple concepts.

This approach is also of interest for its own sake. I already found many *infinite* families of pairs of binary trees (defined in terms of *regular-tree grammars*) for which the statement $\mathcal{W}(T_1)\cap\mathcal{W}(T_2) \neq \emptyset$ is true, and this fact is proved automatically by computer, by automatically deriving a regular grammar for the language given by the $\mathcal{W}(T_1)$ of the left-family and the same for $\mathcal{W}(T_2)$, and then use the well-known algorithm from the theory of formal languages that computes a grammar for the intersection. Of course it is decidable whether or not it is empty.

As a bonus we actually get the regular expression defining the language $\mathcal{W}(T_1) \cap \mathcal{W}(T_2)$, for T_1 in the left-family, and T_2 in the right-family. To take the simplest example, if L_n is the 'left-comb' on n leaves, defined by $L_1 = [\quad]$ and $L_n = [L_{n-1}\,, [\quad]]$ (i.e. L_n is the tree in which every internal vertex has a right-child that is a leaf), and R_n is the 'right-comb' defined analogously, then $\mathcal{W}(L_n) \cap \mathcal{W}(R_n)$ consists of the two words $\{10^{n-2}1\,,\, 20^{n-2}2\}$ when n is odd, and the two words $\{10^{n-2}2\,,\, 20^{n-2}1\}$ when n is even.

So far it was always non-empty (luckily!, or else we would have had a counterexample to 4CT).

Conclusion: The Medium is (a large part of) the Message

The **intellectual merit** of the present proposal should be assessed on (at least) three levels. On the 'lowest' level this research contributes to combinatorics, an important field of mathematics with many applications to almost every branch of science, technology and human endeavor (the World Wide Web and telephone communication, CD players and pictures from Mars would be impossible without it, to mention just a few things that come to mind). The fact that I am using computers extensively in order to do my research in combinatorics should not be held for or against it, it is just another (legitimate) tool.

On the 'middle level', by using symbolic computation on a day-to-day basis, and trying to develop new algorithms (like in WZ theory), this research contributes, both directly and indirectly, to computer algebra, which is emerging as an indispensable tool not only in mathematics, but in all of science and technology.

Finally, on the 'highest-level', this research contributes to a new outlook and *awareness* in mathematical research. Mathematical research, until now, was *paper and pencil* and *a priori*, and people like Appel and Haken and Thomas Hales have to be apologetic and defensive about using computers. The computer is a mighty tool, go forth and use it! But we *humans* must think of *creative* ways of utilizing its immense potential, over-and-above its obvious use as a 'numerical and symbolic calculator' and 'brute force number- (and symbol-) cruncher'. We urgently need to develop new *methodologies* to enable us to make full use of computers. The potential applications are unforesee-

able, but I am sure that future computer mathematics will make all past and present mathematics look like Mickey-Mouse stuff. But these new advances will not come by themselves. The role of the human mathematician would have to change from that of 'athlete' to that of 'coach', and this would also necessitate a change in mentality. I hope that my preaching (in my papers and the opinion column of my website), courses and seminars on Experimental Mathematics, and especially research (both the research papers viewed as case studies, and the more philosophical and methodological papers, e.g. [23][26][27][28]), will form a modest, yet strictly positive, beginning. **If we build it** (a new experimental methodology for mathematics), **they will come** (present and future mathematicians will practice it.)

Similarly, the **broader impact** should also be judged on three levels. Combinatorics per se, and Computer Algebra, are both essential to science, technology, and even entertainment. But, more generally, *mathematics*, as a whole, is one of the greatest pillars of our civilization and culture, both spiritually and materially. Helping change the way we practice mathematics (for the better, I am sure), would have the broadest impact on mathematics itself. **And what's good for mathematics is good for humanity**.