

How the Upper Bound Conjecture Was Proved

Richard P. Stanley

I can trace the path of the proof of the Upper Bound Conjecture (UBC) for spheres back to high school. I went to high school in Savannah, Georgia. Around the age of 13 I became interested in mathematics, but there was no one who could give me proper guidance. One person, most likely an Armstrong Junior College¹ professor named Stubbs, did suggest that I join the Mathematical Association of America and receive their journal *The American Mathematical Monthly*, so I did this. Almost all of the articles were above my head. One paper [4], however, really caught my attention. It was entitled “Preferential arrangements.” A preferential arrangement of an n -element set S would nowadays be called an ordered set partition. It is essentially a linear ordering of the elements of S , allowing ties. More precisely, it is a sequence (B_1, \dots, B_k) of pairwise disjoint nonempty subsets of S whose union is S . Thus there are thirteen preferential arrangements of $\{1, 2, 3\}$, given by 1-2-3, 1-3-2, 2-1-3, 2-3-1, 3-1-2, 3-2-1, 12-3, 13-2, 23-1, 1-23, 2-13, 3-12, 123. Let $f(n)$ denote the number of preferential arrangements of an n -element set. The paper by Gross states the result that

$$f(n) = D^n (2 - e^z)^{-1} \Big|_{z=0},$$

which is equivalent to the generating function (due to Cayley)

$$\sum_{n \geq 0} f(n) \frac{z^n}{n!} = \frac{1}{2 - e^z}. \quad (1)$$

This seemed to me to be a truly amazing fact. I did not understand why it was true, but just the statement was pure magic. Why was there this mysterious connection between a discrete counting problem and the exponential function of calculus?

The next step occurred when I was an undergraduate at Caltech and took a course in analytic number theory from Tom Apostol. I was fascinated by the Riemann zeta function $\zeta(s)$ and browsed through the book *The Zeta Function of Riemann* by E. C. Titchmarsh. On page 7 appears the following

¹Now Armstrong Atlantic State University.

result. Let $g(n)$ be the number of ways to write the integer n as an ordered product of integers all greater than 1. For instance, $g(12) = 8$, corresponding to $2 \cdot 2 \cdot 3$, $2 \cdot 3 \cdot 2$, $3 \cdot 2 \cdot 2$, $2 \cdot 6$, $6 \cdot 2$, $3 \cdot 4$, $4 \cdot 3$, 12 . Thus $g(n)$ is some kind of “multiplicative analogue” of $f(n)$. Moreover, if n is a product of k distinct primes then $g(n) = f(k)$. Equation (1.2.15) asserts that

$$\sum_{n \geq 1} \frac{g(n)}{n^s} = \frac{1}{2 - \zeta(s)}. \quad (2)$$

It was immediately clear to me that this formula was some kind of analogue of equation (1), but at that time I had no idea how to make this feeling more precise. I also did not realize that this was the kind of question that a research mathematician might work on.

When I was a junior at Caltech my adviser Marshall Hall asked me in the spring what I planned to do that summer. I told him that I was thinking of attending an NSF “math camp” for undergraduates (I can’t recall exactly what it was called), a precursor of today’s REUs, somewhere in Oklahoma. Marshall Hall said something like, “Why would you want to go all the way to Oklahoma, Richard? Wouldn’t you rather stay here and work at JPL?” JPL was the Jet Propulsion Laboratory, operated by Caltech for NASA. JPL was responsible for missions involving unmanned extraterrestrial spacecraft. I had no idea that working at JPL was even a possibility, but Marshall Hall arranged for me to be interviewed by Edward Posner, who was originally trained in ring theory and had become the head of the group that developed error-correcting codes used by the spacecraft to send and receive information. I was quite nervous going into this interview, worrying about how to answer questions like “What do you think you can contribute to our mission?” and “How do you expect to benefit from working at JPL?” However, the first question was “What are the Sylow theorems?” After this the interview went smoothly and I got the job.

It was very exciting to be at JPL when missions such as Mariner and Voyager to the moon and other planets were in operation. I continued working there every summer until one year after I completed graduate school. I was working with a very strong group of mathematicians, including (in addition to Posner) Len Baumert, Bob McEliece, Gene Rodemich, and Howard Rumsey. Consultants and/or visitors included Elwyn Berlekamp, Solomon

Golomb, Irwin Jacobs, Gus Solomon, Herb Taylor, and Andrew Viterbi. I can remember that Jacobs and Viterbi had recently started a company called Linkabit. Some of my JPL colleagues thought that this was a risky mistake in judgment. A perusal of the internet will allow the reader to judge this for himself or herself.

In some work on coding theory, Bob McEliece needed a formula that turned out to be equivalent to the computation of the Möbius function of the lattice of partitions of a set. This result, originally due to Marcel Paul Schützenberger and independently Roberto W. Frucht and Gian-Carlo Rota, appeared in a famous paper [7] of Rota (my future thesis adviser). I found this paper of Rota extremely interesting, both for the beautiful combinatorics and for the way that it unified seemingly disparate topics. In particular, I realized that if ζ denotes the ζ function of a locally finite poset P and if $s \leq t$ in P , then $(2 - \zeta)^{-1}(s, t)$ is the number of chains $s = t_0 < t_1 < \dots < t_k = t$ from s to t . The similarity to equations (1) and (2) was apparent. I realized that the theory of incidence algebras appearing in Rota's paper must be the key for unifying the two formulas. What was needed was a connection with generating functions. At that time I did not have the proper perspective and came up with an unsatisfying contrived explanation.

I don't remember exactly which summer I had the ideas described above, but most likely 1967, one year after becoming a graduate student at Harvard. When I returned to Harvard after that summer I mentioned my work to some professors. Again I cannot remember the precise details, but someone told me that Gian-Carlo Rota was at M.I.T. In fact, he had just returned from a two-year hiatus at Rockefeller University. I do remember that when Raoul Bott at some stage heard about my working for Rota, he asked me why I wanted to go into such a Mickey Mouse subject! However, Bott changed his tune later and showed great respect for Rota-style combinatorics. At any rate, I made an appointment with Rota and had a long, fruitful discussion. He in fact had been thinking along similar lines and had the key idea of looking for an isomorphism between rings of generating functions and subalgebras of incidence algebras. Very quickly I saw how to use this idea to develop the unification of various classes of generating functions. Peter Doubilet made some further contributions, eventually culminating in the paper [3].

What does all this have to do with the Upper Bound Conjecture? Because

I enjoyed algebra, I began wondering while at JPL what one could say about the structure of the incidence algebra $I(P)$ of a finite poset P (over a field K). The first question that came to mind was the classification of two-sided ideals. It wasn't difficult to show that for finite posets, the lattice of two-sided ideals of the incidence algebra $I(P)$ is isomorphic to the lattice of order ideals of the set of intervals of P , ordered by reverse inclusion. In particular, the number of two-sided ideals is finite. I wondered what this number was when P is an n -element chain, in which case $I(P)$ is isomorphic to the algebra of $n \times n$ upper triangular matrices over K . By a simple bijection I noticed that this number is the number of $2 \times n$ matrices whose entries are $1, 2, \dots, 2n$ (each number appearing exactly once), with each row and column increasing. (At that time I had no idea that I was looking at a special class of standard Young tableaux.) By a very laborious argument which I have thankfully forgotten, I was able to show that this number is the Catalan number $\frac{1}{n+1} \binom{2n}{n}$.

When I mentioned this work to Bob McEliece, he told me that these $2 \times n$ matrices reminded him of plane partitions, a subject inaugurated by P. A. MacMahon. (For the history of plane partitions, see the notes to Chapter 7 in [9].) I started looking at MacMahon's great opus [6], which I found quite fascinating. Plane partitions are a two-dimensional analogue of ordinary (linear) partitions. Thus it is natural to consider extensions to higher dimensions, as first suggested by MacMahon. MacMahon was unable to make any progress on the enumeration of three-dimensional (or *solid*) partitions, so I decided to take a look. The only work in this area was a paper of E. M. Wright [10], in which he enumerated a very special class of solid partitions.

To be more specific, let \mathbb{N} denote the infinite chain $0 < 1 < 2 < \dots$. The enumeration of d -dimensional partitions of n is equivalent to the enumeration of n -element order ideals of \mathbb{N}^{d+1} . Wright enumerated the n -element order ideals of the poset $V \times \mathbb{N}^2$, where V was the three-element order ideal of \mathbb{N}^2 with one minimal and two maximal elements. This suggested the question of enumerating n -element order ideals of $P \times \mathbb{N}^2$ for any finite poset P . Wright used an induction argument where the base case was $V \times \mathbb{N}$. Thus I first needed to enumerate n -element order ideals of $P \times \mathbb{N}$. Equivalently, one needs to count order-preserving maps $\sigma: P \rightarrow \mathbb{N}$ according to the sum $n = \sum_{t \in P} \sigma(t)$. This turned out to be a very interesting question in itself and was the inspiration for my doctoral thesis. No further progress has been

made on the original question of the exact enumeration of solid partitions, though there have been interesting asymptotic results, e.g., [2].

A couple of years after obtaining my Ph.D. in 1970² David Smith, a mathematics professor at Duke University, told me about an interesting conjecture which he thought might be related to my work. This was the *Anand-Dumir-Gupta conjecture* [1] on “magic squares.” These magic squares are much weaker creatures than those considered in recreational mathematics; they are $n \times n$ matrices of nonnegative integers whose rows and columns all have the same sum. Let $H_n(r)$ denote the number of such magic squares with row and columns sums r . Part of the conjecture of Anand-Dumir-Gupta is that for fixed n , $H_n(r)$ is a polynomial in r . For instance, it is very easy to see that $H_1(r) = 1$ and $H_2(r) = r + 1$. P. A. MacMahon [6, §407] computed in 1916 that $H_3(r) = \binom{r+5}{5} - \binom{r+2}{5}$, and David Smith himself computed $H_4(r)$ around 1970. He was right on the nose about my interest in this problem. Magic squares seemed very similar to the objects (called P -partitions, where P is a finite partially ordered set) studied in my thesis. Both magic squares and P -partitions involved nonnegative integer solutions to linear equalities

²To be precise, I had fulfilled all the requirements for obtaining a Ph.D. in 1970 but did not apply for a degree that year, so I began a postdoc at M.I.T. in fall 1970 without officially having a Ph.D. This is because the Vietnam War was in progress. I had a student deferment from the draft. On June 23, 1970, my draft eligibility would come to an end because of reaching the age of 26. That year there was also a draft lottery. Each day d of the year was randomly assigned a different number $f(d)$ from 1 to 365. If you were born on day d then you could not be drafted until the $f(d)$ th day of the year. June 23 corresponded to 109, so I would be eligible (assuming no student deferment) on April 19. My student deferment would expire when classes ended in May, so I decided not to graduate although in practice draft boards were reluctant to give a graduate student deferment for more than four years. Indeed, my draft board (located in Savannah) informed me that my request for a further year’s deferment was denied, and that I would be reclassified 1-A (available to be drafted) on April 19. Thus I had about a two month period of draft eligibility. (As an amusing aside, in April I had to take a draft physical. In the middle of this exam, which was at a nearby army base, while we were standing around in our underwear, someone phoned in a bomb threat. We had to wait outside for about 20 minutes while the area was checked. One of the examiners told us that this happened every time the group from Cambridge was examined, and that we were lucky that this was April and not January.) Although I passed the draft physical, I never heard from my draft board so once June 23 arrived I was safe. I could also mention that M.I.T. had some lawyers who were engaged in helping persons associated with M.I.T. avoid the draft. During my period of eligibility they told me not to worry, even if I were drafted they could claim that my being at M.I.T. was essential for the nation’s welfare, and that such claims had always been honored.

and inequalities.

I analyzed carefully the computation of $H_3(r)$ by MacMahon in the hope that it could be generalized. It was based on a general technique (now called the *Elliott-MacMahon algorithm*) developed by himself and E. B. Elliott which MacMahon called the “syzygetic method.” After quite a bit of work on trying to figure out exactly what syzygies were and how they behaved, I finally realized that the key to the polynomiality of $H_n(r)$ was the famous Hilbert syzygy theorem from 1890³. Moreover, I could use MacMahon’s syzygetic method to prove the other parts of the Anand-Dumir-Gupta conjecture, which concerned certain properties of the polynomial $H_n(r)$. This work resulted in the paper [8]. In this paper appears a geometric interpretation of MacMahon’s algorithm which, among other things, relates the polynomials $H_n(r)$ (and some more general polynomials) to certain triangulations of polytopes, in particular, the number f_i of i -dimensional faces of such triangulations for all i . At that time I had no interest in the f_i ’s themselves.

This situation changed when I attended a talk of Peter McMullen at Stanford, most likely in 1973. There I learned of the beautiful Upper Bound Conjecture for spheres and of McMullen’s proof for convex polytopes based on the line shellings of Bruggesser and Mani. The UBC for spheres gives an explicit upper bound $f_i(n, d)$ for the number of i -dimensional faces of a triangulation of a $(d - 1)$ -dimensional sphere with n vertices. This bound is achieved by the *cyclic polytopes*, so if true is best possible. I wondered whether the f_i ’s appearing in my paper on the Anand-Dumir-Gupta conjecture could have anything to do with the UBC and whether the algebraic machinery I had developed could give some information about the f_i ’s. At that time this thought was just wild speculation. In retrospect it is remarkable that it turned out to be so fruitful. I did not have any special insight into why commutative algebra might be related to the UBC; I simply noticed that the UBC was about the number of i -dimensional faces of some geometric object and that the machinery used to prove the Anand-Dumir-Gupta conjecture was also related to such numbers.

³The Hilbert syzygy theorem is a result in commutative algebra. I had taken a course in graduate school on commutative algebra that I did not find very interesting. It did not cover the Hilbert syzygy theorem. I had to learn quite a bit of commutative algebra from scratch in order to understand the work of Hilbert.

It was clear from the beginning that the rings I looked at in my paper on magic squares (which later turned out to be just the coordinate rings of projective toric varieties) were not going to be general enough to prove the UBC. I needed a ring whose Hilbert function had a certain relation to the f -vector of a triangulation Δ of a sphere. It wasn't long before I realized that a certain graded algebra $K[\Delta] = K[x_1, \dots, x_n]/I$ (now called the *face ring* or *Stanley-Reisner ring* of Δ), where I is an ideal of the polynomial ring $K[x_1, \dots, x_n]$ generated by certain squarefree monomials, would have the right Hilbert function. At that time I thought that this ring was too simple to be of use. In particular, the variety associated with such a ring is just a union of linear subspaces, hardly of great interest to algebraic geometers.

I had been discussing my work with Ken Baclawski. He told me about the paper [5] of Mel Hochster. In this paper Hochster shows that the rings I used in my paper on magic squares (among others) are Cohen-Macaulay rings. This was the first time I had heard of Cohen-Macaulay rings. Hochster's paper involved a lot of interesting polyhedral combinatorics, and I began to wonder whether Cohen-Macaulay rings might have something to do with the UBC. It wasn't clear to me what effect the Cohen-Macaulay property of a ring had on its Hilbert function. I asked some algebraic geometers at M.I.T. about this, including Michael Artin. Artin said something like, "Aren't Cohen-Macaulay rings free over something? This is due to Hironaka . . ." Though it was unclear to me exactly what Artin meant, I could see immediately that being "free over something" could have an influence on the Hilbert function.

At this point my memory becomes a little murky. I started having discussions with David Eisenbud, who became an invaluable resource. Eventually it became clear to me that Cohen-Macaulay graded K -algebras $R = R_0 \oplus R_1 \oplus \dots$, generated by R_1 , where $R_0 = K$ and K is infinite, were finitely-generated free modules over a polynomial subring $K[\theta_1, \dots, \theta_d]$, with $\theta_i \in R_1$. (This result is essentially equivalent to the Noether normalization applied to R .) It turns out that if Δ triangulates a sphere and if $K[\Delta]$ is Cohen-Macaulay, then the freeness result yields a bound on the f -vector of Δ which exactly coincides with the UBC — a truly astonishing "coincidence"! At this point I was starting to feel that my approach might actually lead somewhere. Could it be just be an accident that the Cohen-Macaulay property gave precisely the result I was looking for?

It remained to show that $K[\Delta]$ is Cohen-Macaulay when Δ triangulates a sphere. At that time I had nowhere enough background in commutative and homological algebra to prove this result. However, I was able to use Hochster's work to show that a very special class of triangulated spheres satisfied the UBC. These spheres may be described as follows: let \mathcal{P} be a convex d -polytope in \mathbb{R}^d with integer vertices. Suppose that \mathcal{P} has a triangulation Δ , also with integer vertices, which is *primitive* (the d -dimensional faces have volume $1/d!$, the minimum possible), and every face of Δ whose interior lies in the interior of \mathcal{P} has dimension greater than $\lfloor d/2 \rfloor$. Then the boundary of Δ (a triangulation of a $(d-1)$ -sphere) satisfies the UBC. This result is too special to be of much interest (and in fact it can be deduced from the UBC for convex polytopes proved earlier by McMullen), but it does provide some indication that it might be possible to show that $K[\Delta]$ is Cohen-Macaulay when Δ triangulates a sphere.

Although I did not have the background to prove the needed result about $K[\Delta]$, the same could not be said of Mel Hochster. He independently defined this ring, having no idea at the time of its potential applicability to combinatorics, as a kind of combinatorial analogue of the rings which he had considered in [5]. He gave to his student Gerald Reisner the problem of determining when $K[\Delta]$ was Cohen-Macaulay, or more generally, computing the depth of $K[\Delta]$.

In 1974 I attended the International Congress of Mathematicians in Vancouver. I was scheduled to give a ten-minute submitted (uninvited) talk for which I planned to discuss my partial result mentioned above. I attended an invited talk by Victor Klee in which he stated that the UBC for spheres was one of the main open problems related to polyhedra. Soon after that I ran into David Eisenbud at the meeting. He informed me that Reisner had found a complete characterization of Cohen-Macaulay face rings $K[\Delta]$. In particular, $K[\Delta]$ was Cohen-Macaulay whenever Δ triangulated a sphere. Thus the UBC for spheres was proved! This was the greatest "math high" of my career, which lasted throughout the meeting. A little after talking to David Eisenbud, I ran into Peter McMullen (who was also an invited speaker, though his talk was unrelated to the UBC) and could inform him of the great news. About a day later was my ten-minute talk, in which I could say that the UBC had been proved two days earlier.

- Morals.* 1. The shortest path may not be the best.
2. Even if you don't arrive at your destination, the journey can still be worthwhile.

References

- [1] H. Anand, V. C. Dumir, and H. Gupta, A combinatorial distribution problem, *Duke Math. J.* **33** (1966), 757–769.
- [2] S. Balakrishnan, S. Govindarajan, and N. S. Prabhakar, On the asymptotics of higher-dimensional partitions, *J. Phys. A: Math. Theor.* **45** (2012), #055001.
- [3] P. Doubilet, G.-C. Rota, and R. Stanley, On the foundations of combinatorial theory. VI. The idea of generating function, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971)*, Vol. II: Probability theory, Univ. California Press, Berkeley, Calif., 1972, pp. 267–318.
- [4] O. A. Gross, Preferential arrangements, *Amer. Math. Monthly* **69** (1962), 4–8.
- [5] M. Hochster, Rings of invariants of tori, Cohen-Macaulay rings generated by monomials, and polytopes *Ann. of Math. (2)* **96** (1972), 318–337.
- [6] P. A. MacMahon, *Combinatory Analysis*, vols. 1 and 2, Cambridge University Press, 1916; reprinted by Chelsea, New York, 1960, and by Dover, New York, 2004.
- [7] G.-C. Rota, On the foundations of combinatorial theory. I. Theory of Möbius functions. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **2** (1964), 340–368.
- [8] R. Stanley, Linear homogeneous diophantine equations and magic labelings of graphs, *Duke Math. J.* **40** (1973), 607–632.
- [9] R. Stanley, *Enumerative Combinatorics*, vol. 2, Cambridge University Press, New York/Cambridge, 1999.

- [10] E. M. Wright, The generating function of solid partitions, *Proc. Roy. Soc. Edinburgh Sect. A* **67** (1965/1967), 185-195.