ELECTRONIC RESEARCH ANNOUNCEMENTS
OF THE AMERICAN MATHEMATICAL SOCIETY
Volume 2, Number 1, August 1996

A NEW PROOF OF THE FOUR-COLOUR THEOREM

NEIL ROBERTSON, DANIEL P. SANDERS, PAUL SEYMOUR, AND ROBIN THOMAS

(Communicated by Ronald Graham)

ABSTRACT. The four-colour theorem, that every loopless planar graph admits a vertex-colouring with at most four different colours, was proved in 1976 by Appel and Haken, using a computer. Here we announce another proof, still using a computer, but simpler than Appel and Haken's in several respects.

1. INTRODUCTION

For our purposes a *graph* G consists of a finite set $V(G)$ of vertices, a finite set $E(G)$ of edges, and an incidence relation between them. Each edge is incident with two vertices, called its *ends*. An edge whose ends are equal is called a *loop*. A *plane* graph is a graph embedded in the plane (without crossings, in the usual way). The four-colour theorem (briefly, the 4CT) asserts that every loopless plane graph admits a 4-colouring, that is, a mapping $c : V(G) \to \{0, 1, 2, 3\}$ such that $c(u) \neq c(v)$ for every edge of G with ends u and v. This was conjectured by F. Guthrie in 1852, and remained open until a proof was found by Appel and Haken [3], [4], [5] in 1976.

Unfortunately, the proof by Appel and Haken (briefly, A&H) has not been fully accepted. There has remained a certain amount of doubt about its validity, basically for two reasons:

(i) part of the A&H proof uses a computer, and cannot be verified by hand, and
(ii) even the part of the proof that is supposed to be checked by hand is extraordinarily complicated and tedious, and as far as we know, no one has made a complete independent check of it.

Reason (i) may be a necessary evil, but reason (ii) is more disturbing, particularly since the 4CT has a history of incorrect "proofs". So in 1993, mainly for our own peace of mind, we resolved to convince ourselves somehow that the 4CT really was true. We began by trying to read the A&H proof, but very soon gave this up. To check that the members of their "unavoidable set" were all reducible would require

Received by the editors September 26, 1995.
1991 *Mathematics Subject Classification.* Primary 05C10, 05C15, 05C85.
Research partially performed under a consulting agreement with Bellcore, and partially supported by DIMACS Center, Rutgers University, New Brunswick, New Jersey 08903, USA.
Neil Robertson was partially supported by NSF under Grant No. DMS-8903132 and by ONR under Grant No. N00014-92-J-1965.
Daniel P. Sanders was partially supported by DIMACS and by ONR under Grant No. N00014-93-1-0325.
Robin Thomas was partially supported by NSF under Grant No. DMS-9303761 and by ONR under Grant No. N00014-93-1-0325.

a considerable amount of programming, and *also* would require us to input by hand into the computer descriptions of 1478 graphs; and this was not even the part of their proof that was most controversial. We decided it would be easier, and more fun, to make up our own proof, using the same general approach as A&H. So we did; it was a year's work, but we were able to convince ourselves that the 4CT is true and provable by this approach [14]. In addition, our proof turned out to be simpler than that of A&H in several respects.

The basic idea of the proof is the same as that of A&H. We exhibit a set of "configurations"; in our case there are 633 of them. We prove that none of these configurations can appear in a minimal counterexample to the 4CT, because if one appeared, it could be replaced by something smaller, to make a smaller counterexample to the 4CT (this is called proving "reducibility"; here we are doing exactly what A&H and several other authors did — for instance, [1], [2], [5], [9]). But every minimal counterexample is an "internally 6-connected triangulation" (defined later), and in the second part of the proof we prove that at least one of the 633 configurations appears in *every* internally 6-connected triangulation. (This is called proving "unavoidability".) Consequently, there is no minimal counterexample, and so the 4CT is true. Where our method differs from A&H is in how we prove unavoidability.

Some aspects of this difference are: we confirm Heesch's conjecture that one can prove unavoidability of some reducible set without looking beyond the second neighbourhoods of "overcharged" vertices; consequently, we avoid the problems with configurations that "wrap around and meet themselves", that were a major source of complications for A&H; our unavoidable set has size about half that of the A&H set; our "discharging procedure" for proving unavoidability (derived from an elegant method of Mayer [13]) only involves 32 discharging rules, instead of the 487 of A&H; and we obtain a quadratic time algorithm to find a 4-colouring of a planar graph, instead of the quartic algorithm of A&H.

Our proof is also somewhat easier to check, because we replace the mammoth hand-checking of unavoidability that A&H required, by another mammoth hand-checkable proof, but this time written formally so that, if desired, it can be read and checked by a computer in a few minutes. We are making the necessary programs and data available to the public for checking via "anonymous ftp" from ftp.math.gatech.edu in the directory pub/users/thomas/four.

The paper is organized as follows. In section 2 we introduce the concept of a configuration, and state the two main results. In section 3 we discuss reducibility, and in section 4 unavoidability. In section 5 we briefly discuss a quadratic algorithm to 4–colour planar graphs, and in section 6 we make some concluding remarks.

2. Configurations

The *degree* of a vertex v of a graph G is the number of edges incident with it, and is denoted by $d_G(v)$. A region of a plane graph is a *triangle* if it is incident with precisely three edges. A plane graph is a *triangulation* if it is connected, loopless and every region is a triangle. Thus, a triangulation can have parallel edges, but no circuit of length two bounds a region.

A *minimal counterexample* means a plane graph G that is not 4-colourable, such that every plane graph G' with $|V(G')|+|E(G')| < |V(G)|+|E(G)|$ is 4-colourable. To prove the 4CT we must show that there is no minimal counterexample. It is

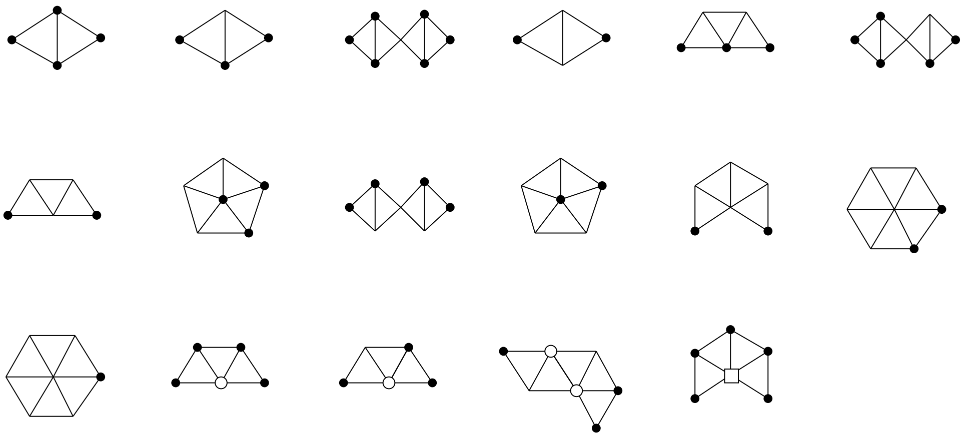

FIGURE 1. A set of configurations.

easy to show that every minimal counterexample is a triangulation, and almost is 6-connected. More precisely, let us say a graph G is *internally* 6*-connected* if G has at least six vertices, and for every set X of vertices of G such that the graph $G \backslash X$ obtained from G by deleting X is disconnected, either $|X| \geq 6$, or $|X| = 5$ and $G \backslash X$ has exactly two components, one of which has exactly one vertex. Thus every vertex of an internally 6-connected graph has degree at least five. We have (see, for example, [7])

(2.1) *Every minimal counterexample is an internally 6-connected triangulation.*

A *near-triangulation* is a non-null connected loopless plane graph G such that every finite region is a triangle. A *configuration* K consists of a near-triangulation $G(K)$ and a map $\gamma_K : V(G(K)) \to \mathbb{Z}$ with the following properties:

(i) for every vertex v, $G(K) \backslash v$ has at most two components, and if there are two, then $\gamma_K(v) = d(v) + 2$,

(ii) for every vertex v, if v is not incident with the infinite region, then $\gamma_K(v) = d(v)$, and otherwise $\gamma_K(v) > d(v)$; and in either case $\gamma_K(v) \geq 5$,

(iii) K has ring-size ≥ 2, where the *ring-size* of K is defined to be $\sum_v (\gamma_K(v) - d(v) - 1)$, summed over all vertices v incident with the infinite region such that $G(K) \backslash v$ is connected.

Two configurations are *isomorphic* if there is a homeomorphism of the plane mapping $G(K)$ to $G(L)$ and γ_K to γ_L. In [14] we exhibit a set of 633 configurations that are essential to our proof. The set is also available in electronic form by "anonymous ftp" as described earlier. Due to space limitations the full set cannot be described here, but in Figure 1 we show a small subset that will be referred to later in the paper. In Figure 1 we use a convention introduced by Heesch [11]. The shapes of vertices indicate the value of $\gamma_K(v)$ as follows: A solid black circle means $\gamma_K(v) = 5$, a dot (or what appears in the picture as no symbol at all) means $\gamma_K(v) = 6$, a hollow circle means $\gamma_K(v) = 7$, and a hollow square means $\gamma_K(v) = 8$. (We will not need vertices v with $\gamma_K(v) > 8$ in this paper.)

Any configuration isomorphic to one of the 633 configurations exhibited in [14] is called a *good* configuration. Let T be a triangulation. A configuration K *appears* in T if $G(K)$ is an induced subgraph of T, every finite region of $G(K)$ is a region of T, and $\gamma_K(v) = d_T(v)$ for every vertex $v \in V(G(K))$. We prove the following two statements.

(2.2) *If T is a minimal counterexample, then no good configuration appears in T.*

(2.3) *For every internally 6-connected triangulation T, some good configuration appears in T.*

From (2.1), (2.2) and (2.3) it follows that no minimal counterexample exists, and so the 4CT is true. We shall discuss (2.2) in section 3, and (2.3) in section 4. The first proof needs a computer. The second can be checked by hand in a few months, or, using a computer, it can be verified in a few minutes.

3. Reducibility

By a *circuit* we mean a non-null connected graph in which every vertex has degree two. Let K be a configuration. A near-triangulation S is a *free completion of K with ring R* if

(i) R is an induced circuit of S, and bounds the infinite region of S,
(ii) $G(K)$ is an induced subgraph of S, $G(K) = S \backslash V(R)$, every finite region of $G(K)$ is a finite region of S, and the infinite region of $G(K)$ includes R and the infinite region of S, and
(iii) every vertex v of S not in $V(R)$ has degree $\gamma_K(v)$ in S.

It is easy to check that every configuration has a free completion. (This is where we use the condition that ring-size ≥ 2 in the definition of a configuration — the ring-size is actually the length of the ring in the free completion, as the reader may verify.) Moreover, if S_1, S_2 are two free completions of K, there is a homeomorphism of the plane fixing $G(K)$ pointwise and mapping S_1 to S_2. (This is where condition (i) in the definition of a configuration is used.) Thus, there is essentially only one free completion, and so we may speak of "the" free completion without serious ambiguity.

Let R be a circuit. There is a concept of "consistency" of a set of 4-colourings of R which essentially dates back to Kempe [12] and Birkhoff [7]. We will not define this concept here, but instead mention the properties we need. They are:

(i) the empty set is consistent,
(ii) the union of any two consistent sets is consistent,
(iii) if T is a triangulation, R is a circuit of T, T' is the near-triangulation obtained from T by deleting all vertices of T that belong to the open disc bounded by R, and $\mathcal{C}$ is the set of restrictions of all 4-colourings of T' to R, then $\mathcal{C}$ is consistent, and
(iv) consistency is sufficiently restrictive to permit enough reducible configurations (defined later).

Let S be the free completion of a configuration K with ring R. Let $\mathcal{C}^*$ be the set of all 4-colourings of R, and let $\mathcal{C}$ be the set of all restrictions to $V(R)$ of 4-colourings of S. Let $\mathcal{C}'$ be the maximal consistent subset of $\mathcal{C}^* - \mathcal{C}$. (This is well defined by (i) and (ii).) The configuration K is said to be *D-reducible* if $\mathcal{C}' = \emptyset$, and is said to be *C-reducible* if there exists a near-triangulation S', called a *reducer*, obtained from

S by replacing $G(K)$ by a smaller graph (and possibly identifying some vertices of R) such that no member of $\mathcal{C}'$ is the restriction to $V(R)$ of a 4-colouring of S'. It is reasonably easy to show that no D-reducible configuration can appear in a minimal counterexample. (Notice that the appearance of a configuration in T does not imply that the free completion is a subgraph of T, but this difficulty is easy to take care of.) A more serious problem arises with C-reduction. There the natural proof that no C-reducible configuration can appear in a minimal counterexample would "replace" S by S' in T, but how do we know that the resulting graph is loopless? Let us call a reducer *safe* if the above process results in a triangulation for every internally 6-connected triangulation T. Checking safety was a major source of complications for A&H, and one advantage of our new proof is that we are able to avoid these difficulties. The way we handle this problem is that we only consider reducers that are obtained from K by contracting at most four edges, for which the safety check is easy.

In order to prove (2.2) we show that

(3.1) *Each of the good* 633 *configurations is either D-reducible or C-reducible with a safe reducer.*

Theorem (3.1) is proved by means of a computer program that is available to the public for examination. In fact, we used two independent programs that also computed some numerical invariants that could be compared to double-check our results.

4. Unavoidability

A configuration K *appears* in a configuration L if $G(K)$ is an induced subgraph of $G(L)$, every finite region of K is a finite region of L (and hence the infinite region of K includes the infinite region of L), and $\gamma_K(v) = \gamma_L(v)$ for every $v \in V(G(K))$.

Let T be an internally 6-connected triangulation. For a vertex v of T we define the *vicinity* of v in T to be the configuration K with $G(K)$ the subgraph of T induced by all vertices u of T such that there is a path P in T with ends u and v with at most three vertices, and such that the interior vertex of P (if there is one) has degree at most eight, and $\gamma_K(v) = d_T(v)$ for $v \in V(G(K))$. To prove (2.3) we show that a good configuration appears in the vicinity of some vertex. We now explain how we locate such a vertex.

A *rule* is a 6-tuple $(G, \alpha, \epsilon, r, s, t)$, where G is a near-triangulation, α and ϵ are mappings from $V(G)$ to $\{5, 6, 7, 8\}$ and $\{-, 0, +\}$, respectively, $r > 0$ is an integer, and s and t are distinct adjacent vertices of G. Figure 2 describes a set $\mathcal{R}$ of 32 rules as follows. The mapping α is described using the same conventions as for Figure 1, ϵ is described by placing an appropriate sign next to the corresponding vertex ("0" is omitted), $r = 2$ for the first rule and $r = 1$ for all the other rules, and $s, t \in V(G)$ are such that the unique directed edge of the rule is directed from s to t. There is some system in the first nine rules, but the other rules were chosen by trial and error, and have no particular plan or pattern.

A *pass* P is a quadruple (K, r, s, t), where K is a configuration, $r > 0$ is an integer and $s, t \in V(G(K))$ such that there exists a rule $R = (G, \alpha, \epsilon, r, s', t') \in \mathcal{R}$ and a homeomorphism h of the plane mapping $G(K)$ onto G such that

(i) $\gamma_K(v) \le \alpha(h(v))$ for every $v \in V(G(K))$ with $\epsilon(h(v)) \in \{-, 0\}$,
(ii) $\gamma_K(v) \ge \alpha(h(v))$ for every $v \in V(G(K))$ with $\epsilon(h(v)) \in \{0, +\}$, and

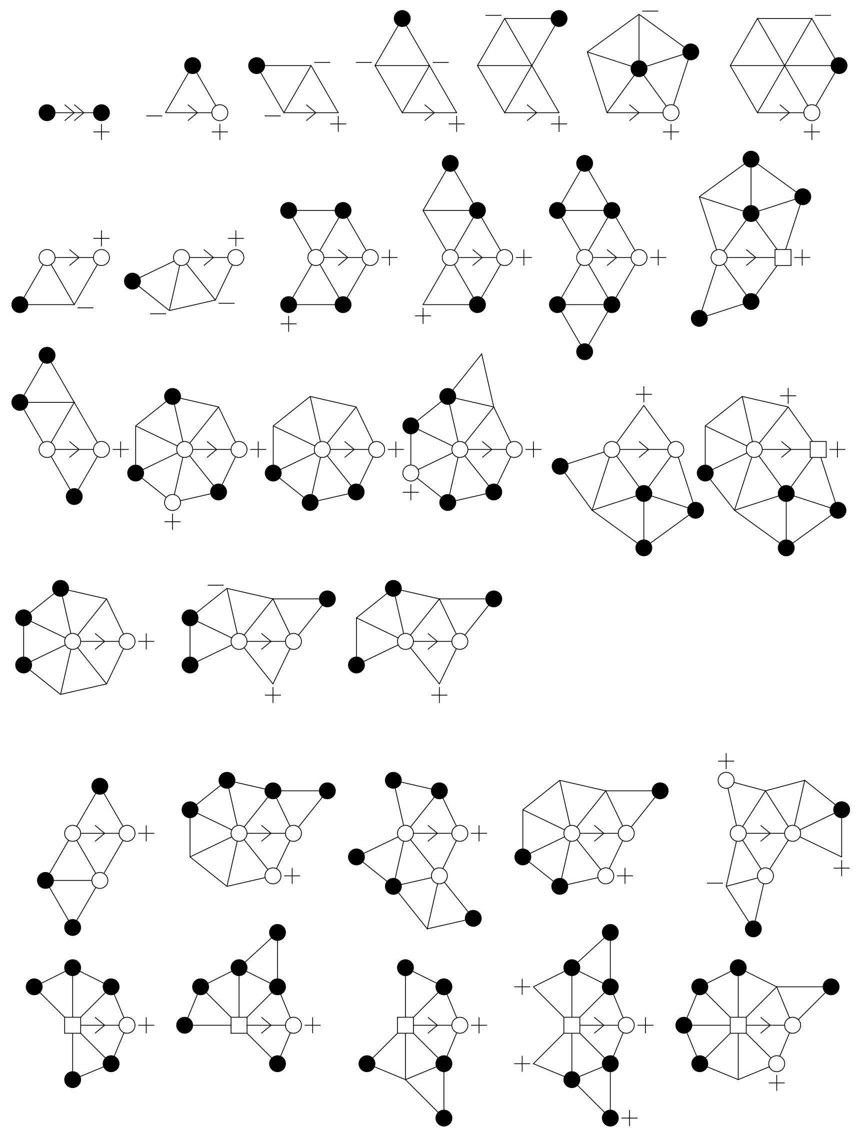

FIGURE 2. The set $\mathcal{R}$ of rules.

(iii) $h(s) = s'$ and $h(t) = t'$.

We say that P *obeys* rule R. It is straightforward to verify that every pass obeys exactly one rule. We write $r(P) = r$, $s(P) = s$, $t(P) = t$, and $K(P) = K$. We call r the *value* of the pass, s its *source*, and t its *sink*. A pass P *appears* in a triangulation T if $K(P)$ appears in T. A pass P *appears* in a configuration L if $K(P)$ appears in

L. If v is a vertex of a triangulation T, we define $\Theta_T(v)$ to be

$$10(6 - d_T(v)) + \sum (r(P) : P \text{ appears in } T\,,\ t(P) = v)$$
$$- \sum (r(P) : P \text{ appears in } T\,,\ s(P) = v)\,.$$

We remark that $\Theta_T(v)$ depends only on the vicinity of v in T.

It follows from Euler's formula that

(4.1) *For every triangulation T, $\sum_{v\in V(T)} \Theta_T(v) = 120$. In particular, there is a vertex v of T with $\Theta_T(v) > 0$.*

Thus in order to prove (2.3) it suffices to prove

(4.2) *If v is a vertex of an internally 6-connected triangulation T with $\Theta_T(v) > 0$, then a good configuration appears in the vicinity of v in T.*

Let $C_1, C_2, \ldots, C_{17}$ be the configurations pictured in Figure 1. The following can be shown without the use of a computer.

(4.3) *If v is a vertex of an internally 6-connected triangulation T with $\Theta_T(v) > 0$, and either $d_T(v) \geq 12$ or $d_T(v) \leq 6$, then a configuration isomorphic to one of $C_1, C_2, \ldots, C_{17}$ appears in the vicinity of v in T.*

Sketch of proof. Assume that $d_T(v) \geq 12$, and suppose that no such configuration appears. Let $d_T(v) = d$ and let D be the set of neighbours of v. For each $u \in D$, let $R(u)$ be the sum of $r(P)$ over all passes P appearing in T with source u and sink v. It can be shown that $R(u) \leq 5$ for every $u \in D$ (we omit the details — this is just a finite case analysis). Thus $\sum_{u\in D} R(u) \leq 5d$, and hence

$$\Theta_T(v) = 10(6-d) + \sum_{u\in D} R(u) \leq 10(6-d) + 5d = 60 - 5d \leq 0\,,$$

a contradiction. This completes the proof in the case when $d_T(v) \geq 12$. We omit the other part, because, again, it is just a finite case analysis. □

Thus in order to prove (4.2) and hence (2.3) it remains to prove the following.

(4.4) *If v is a vertex of an internally 6-connected triangulation T with $\Theta_T(v) > 0$ and $d_T(v) = 7, 8, 9, 10$ or 11, then a good configuration appears in the vicinity of v in T.*

For each of the five cases, we have a proof. Unfortunately they are very long (altogether about 13,000 lines, and a large proportion of the lines take some thought to verify), and so cannot be included in a journal article. Moreover, although any line of the proofs can be checked by hand, the proofs themselves are not "really" checkable by hand because of their length. We therefore wrote the proofs so that they are machine-readable, and in fact a computer can check these proofs in a few minutes. More information is given in [14, Section 7], and full details can be obtained by "anonymous ftp" as described earlier. Alternatively, one can write a computer program to check (4.4) directly, for it is easily seen to be a finite problem.

5. An algorithm

Our proof gives a quadratic algorithm to four-colour planar graphs, which is an improvement over the quartic algorithm of A&H. Let us clarify a possible confusion here. It might appear that there is an "obvious" quadratic algorithm as follows. Given an input plane graph G on n vertices (which may as well be assumed to be a triangulation) we find a good configuration appearing in G, replace G by a smaller graph G', and apply recursion. A reducible configuration can be found in linear time, and a 4-colouring of G can be constructed from a 4-colouring of G' in linear time. Since there are at most n recursive steps, the overall running time would be quadratic.

The problem is that (2.3) only guarantees the appearance of a good configuration if G is an internally 6-connected triangulation, and yet even if we started with a highly connected graph, the connectivity is bound to drop during the recursion. However, (4.2) permits us to quickly find either a good configuration appearing in G, in which case we proceed as suggested above, or a set X of vertices of G violating the definition of internal 6-connection, in which case we apply recursion to two carefully selected subgraphs of G, and then obtain a 4-colouring of G by piecing together 4-colourings of the two subgraphs. We refer to [14] for further details.

6. Discussion

This concludes our discussion of the proof of the 4CT and the associated algorithm. A full account can be found in [14] except for the proofs of (3.1) and (4.4). Those two theorems are just stated in [14] as having been proved by a computer. Verifying (3.1) takes about 3 hours on a Sun Sparc 20 workstation, and (4.4) takes about 20 minutes altogether. Both need less than one megabyte of RAM. The complete programs as well as documenation for them [15], [16] can be obtained for examination by "anonymous ftp" as described earlier. Thus we are making it possible for other scientists to verify all steps in our proof, including the computer programs and data.

We should mention that both our programs use only integer arithmetic, and so we need not be concerned with round-off errors and similar dangers of floating point arithmetic. However, an argument can be made that our "proof" is not a proof in the traditional sense, because it contains steps that can never be verified by humans. In particular, we have not proved the correctness of the compiler we compiled our programs on, nor have we proved the infallibility of the hardware we ran our programs on. These have to be taken on faith, and are conceivably a source of error. However, from a practical point of view, the chance of a computer error that appears consistently in exactly the same way on all runs of our programs on all the compilers under all the operating systems that our programs run on is infinitesimally small compared to the chance of a human error during the same amount of case-checking. Apart from this hypothetical possibility of a computer consistently giving an incorrect answer, the rest of our proof can be verified in the same way as traditional mathematical proofs. We concede, however, that verifying a computer program is much more difficult than checking a mathematical proof of the same length.

References

1. F. Allaire, *Another proof of the four colour theorem*-Part I, Proc. of the 7th Manitoba Conf. on Numerical Math. and Computing (1977), Congressus Numerantium XX, Utilitas Math., Winnipeg, 1978, pp. 3–72. MR **80m:**05031
2. F. Allaire and E. R. Swart, *A systematic approach to the determination of reducible configurations in the four-color conjecture*, J. Combinatorial Theory, Ser. B **25** (1978), 339–362. MR **80i:**05041
3. K. Appel and W. Haken, *Every planar map is four colorable. Part I. Discharging*, Illinois J. Math. **21** (1977), 429–490. MR **58:**27598d
4. K. Appel, W. Haken, and J. Koch, *Every planar map is four colorable. Part II. Reducibility*, Illinois J. Math **21** (1977), 491–567. MR **58:**27598d
5. K. Appel and W. Haken, *Every planar map is four colorable*, A.M.S. Contemporary Math. **98** (1989). MR **91m:**05079
6. A. Bernhart, *Another reducible edge configuration*, Amer. J. Math. **70** (1948), 144–146. MR **9:**366g
7. G. D. Birkhoff, *The reducibility of maps*, Amer. J. Math. **35** (1913), 114–128.
8. D. I. A. Cohen, *Block count consistency and the four color problem*, manuscript.
9. K. Dürre, H. Heesch, and F. Miehe, *Eine Figurenliste zur chromatischen Reduktion*, manuscript eingereicht am 15.8.1977.
10. S. J. Gismondi and E. R. Swart, *A new type of 4-colour reducibility*, Congr. Numer. **82** (1991), 33–48. MR **92j:**05069
11. H. Heesch, *Untersuchungen zum Vierfarbenproblem*, Hochschulskriptum 810/a/b, Bibliographisches Institüt, Mannheim, 1969. MR **40:**1303
12. A. B. Kempe, *On the geographical problem of the four colors*, Amer. J. Math. **2** (1879), 193–200.
13. J. Mayer, *Une propriété des graphes minimaux dans le problème des quatre couleurs*, Problèmes Combinatoires et Théorie des Graphes, Colloques internationaux CNRS No. 260, Paris, 1978. MR **80m:**05042
14. N. Robertson, D. P. Sanders, P. D. Seymour, and R. Thomas, *The four colour theorem*, manuscript.
15. ———, *Reducibility in the four colour theorem*, unpublished manuscript available from ftp://ftp.math.gatech.edu/pub/users/thomas/four.
16. ———, *Discharging cartwheels*, unpublished manuscript available from ftp://ftp.math.gatech.edu/pub/users/thomas/four.
17. P. G. Tait, *Note on a theorem in geometry of position*, Trans. Roy. Soc. Edinburgh **29** (1880), 657–660.
18. H. Whitney and W. T. Tutte, *Kempe chains and the four colour problem*, Studies in Graph Theory (D. R. Fulkerson, eds.), Part II, Math. Assoc. of America, 1975, pp. 378–413. MR **52:**13468

Department of Mathematics, Ohio State University, 231 W. 18th Ave., Columbus, Ohio 43210, USA
E-mail address: `robertso@math.ohio-state.edu`

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA
E-mail address: `dsanders@math.ohio-state.edu`

Bellcore, 445 South Street, Morristown, New Jersey 07960, USA
E-mail address: `pds@bellcore.com`

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332, USA
E-mail address: `thomas@math.gatech.edu`