CONTENTS

THE IRISH MATHEMATICAL SOCIETY

OFFICERS AND COMMITTEE MEMBERS

President	Dr. Fergus Gaines	Department of Mathematics University College Dublin
Vice-President	Dr. Richard Timoney	School of Mathematics Trinity College Dublin
Secretary	Prof. A.G. O'Farrell	Department of Mathematics Maynooth College Maynooth
Treasurer	Dr. Gerard M. Enright	Department of Mathematics Mary Immaculate College Limerick

Committee Members: P. Barry, R. Critchley, B. Goldsmith, D. Hurley, T. Hurley, R. Ryan, M. O'Reilly, M. Ó Searcóid, R. Watson.

LOCAL REPRESENTATIVES

Cork	RTC	Mr. D. Flannery
	UCC	Dr. M. Stynes
Dublin	DIAS	Prof. J. Lewis
	Kevin St.	Dr. B. Goldsmith
	NIHE	Dr. M. Clancy
	St. Patrick's	Dr. J Cosgrave
	TCD	Dr. R. Timoney
	UCD	Dr. F. Gaines
Dundalk	RTC	Dr. E. O'Riordan
Galway	UCG	Dr. R. Ryan
Limerick	MICE	Dr. G. Enright
	NIHE	Dr. R. Critchley
	Thomond	Mr. J. Leahy
Maynooth		Prof. A. O'Farrell
Waterford	RTC	Mr. T. Power
Belfast	QUB	Dr. D.W. Armitage

IMS MEMBERSHIP

Ordinary Membership

Ordinary of the Irish Mathematical Society is open to all persons interested in the activities of the Society. Application forms are available from the Treasurer and from Local Representatives. Special reciprocity rates apply to members of the Irish Mathematics Teachers Association and of the American Mathematical Society.

Institutional Membership

Institutional Membership is a valuable support to the Society. Institutional members receive two copies of each issue of the Bulletin and may nominate up to five students for free membership.

Subscriptions rates

The rates are listed below. The membership year runs from 1st October to 30th September. Members should make payments by the end of January either direct to the Treasurer or through Local Representatives. Members whose subscriptions are more than eighteen months in arrears are deemed to have resigned from the Society.

Ordinary Members	IR£5
IMS–IMTA Combined	IR£6.50
Reciprocity Members from IMTA	IR £1.50
Reciprocity Members from AMS	US$6
Institutional Members	IR£35

Note: Equivalent amounts in foreign currency will also be accepted.

IRISH MATHEMATICAL SOCIETY

Ordinary Meeting

March 22, 1989

The Irish Mathematical Society held an Ordinary Meeting at 12:30 p.m. on Wednesday, March 22, 1989.

Thirteen members were present and the President, F. Gaines, was in the chair.

1. The minutes of the meeting of December 22, 1988 were read, approved and signed.

2. **Meetings:** It was noted with satisfaction that EOLAS has undertaken to pay the airfares of two speakers at the Society's meetings in 1989, up to a limit of £400 apiece. The principal speakers at the September 1989 meeting, to be held in Maynooth on the 7th and 8th of September, will be F. Almgren, S. Donaldson and J. Lewis.

   It was suggested that a press release about the September meeting would be in order.

   A committee, consisting of R. Timoney, F. Gaines and D. Simms, was set up to select and invite a speaker for a meeting in November, to be held in Dublin.

   It was agreed to accept the offer of NIHE Dublin to host the September 1990 meeting.

3. The **Treasurer's Report**, deferred from the previous meeting, was considered and adopted, on the proposal of R. Timoney, seconded by M. Stynes.

4. R. Ryan reported on the status of the **Bulletin**. A printing schedule has been agreed with EOLAS. The Bulletin now makes a profit because the use of TEX and electronic submission have brought down production costs, while advertising revenues have grown. Proposals to improve the journal are invited. The President observed that we should try to limit the number of changes in the overall dimensions of the journal to one more.

5. **Other Business.**

   (1) D. O'Donovan tabled a report on his proposal to conduct a David-style study on Irish Mathematics. It was felt that the Society could not discuss the report without notice, and discussion was deferred to the September meeting, when a panel discussion is to be held on the subject. At the same time, it was thought desirable to initiate discussion of the report in the Colleges, and to that end it was decided to circulate the report to the local representatives. Concern was expressed that this procedure might lead people to suppose that the Society already endorses the report, but this was discounted.

   (2) The Secretary reported on the state of the EUROMATH project. When Phase II begins, there will be a need for substantial investment. A sub-committee is preparing a submission to EOLAS on the report.

   (3) A request for a further subvention from the Groups in Galway conference was referred to the Treasurer.

   (4) There was a discussion of the role of the Society in relation to the Mathematics Contest and the Olympiads. The Society initiated the Contest and National Olympiad. It initiated and is running a major Olympiad preparation programme,with centres in Limerick, Cork, Galway and Dublin. This involves a great deal of (unpaid) work by the organisers and tutors. Money is provided by EOLAS for conferencing, books and materials, but not for the travel expenses of participating students. The expenses of the team for the International Olympiad were paid by the Department of External Affairs, and the organisation of the team was carried out by a special interdepartmental committee. Apart from other benefits, the programme is believed to have had a positive effect on enrollments in Mathematics.

   (5) The Secretary reported on the programme of exchanges for the Bulletin, which is developing well.

The meeting concluded at 1:15 p.m.

Anthony G. O'Farrell,
Secretary

NEWS

Personal Items

- **Dr. Martin Mathieu** of the University of Tübingen will be visiting the Mathematics Department in University College Cork from February 26 to April 20 1990. He will deliver a series of lectures on Operator Theory and C*-algebras.
- **Professor John C. Elliott** of the Mathematics Department, University of Maine at Fort Kent, U.S.A., will visit the Mathematics Department of University College Cork for the period January to May 1990. His main interest is in Mathematics Education.
- **Professor Dan Luecking** of the University of Arkansas, Fayetteville, is a visting lecturer in Trinity College Dublin for 1989–90. Professor Luecking's research is in Complex Analysis.
- **Professor B.H. Murdoch** has retired as Erasmus Smith's Professor of Mathematics at Trinity College Dublin. He was a Senior Fellow of the College in 1988–89. Professor Murdoch will continue to teach in the Department.
- **Dr. Kirsteen Duncan** has taken up a position at Trinity College Dublin for the year 1989–90. Dr. Duncan, who holds a Ph.D. from Heriot-Watt University, works in Numerical Analysis.
- **Dr. Mark Leeney** has taken up a one-year appointment in the Mathematics Department in University College Cork.
- **Pat Crehan** has taken up a temporary lectureship in applied mathematics in the Dublin Institute of Technology at Kevin Street. Dr. Crehan was a student of John Kennedy in UCD and has just completed a doctorate in the area of quantum chaos.
- **Niall Ó Murchadha** of the Experimental Physics Department at University College Cork has been promoted to Associate Professor.

- **Professor Paddy Barry**, Head of the Mathematics Dept at University College Cork, has been appointed to the Course Committee for the Senior Cycle in Mathematics by the National Council for Curriculum and Assessment.
- **Eamonn Murphy** has been promoted to Lecturer in the Mathematics Department of the University of Limerick.
- **Finbarr Holland** is on leave of absence from University College Cork-for the academic year 1988–89, visiting universities in Australia, New Zealand and the U.S.A.
- **Robin Harte** is visiting the University of Alaska at Fairbanks for the present academic year.
- **J.J.H. Miller** is on leave of absence from Trinity College Dublin this year and is now Director of the Institute for Computational Mechanics in Propulsion at the NASA Lewis Research Center in Cleveland, Ohio.
- **Gerard Murphy** will visit the University of Tübingen for the month of May 1990.
- **Roger Dodd** has left Trinity College Dublin to take up a position at California State University at Santa Cruz, where he will join **Hedley Morris**, who resigned from Trinity College Dublin in 1986.
- **Richard Timoney** has been elected to a Fellowship of Trinity College Dublin.
- **Graham Ellis** of the Mathematics Department of University College Galway, has been invited to give a lecture at the British Mathematics Colloquium in 1990. Dr. Ellis will speak on "Algebraic models of topological spaces".
- **John Flynn**, a student at Univertsity College Cork, has won a travelling studentship from the National University of Ireland.

CONFERENCES

Groups In Galway 90

In order to mark Professor Seán Tobin's 60th birthday, the annual Groups in Galway meeting will be expanded into a four-day conference in 1990. The conference will take place in University College Galway on May 9–12. The main speakers will include:

C.K. Gupta (Manitoba)
N.D. Gupta (Manitoba)
G. Higman (Oxford)
T.J. Laffey (Dublin)
P.M. Neumann (Oxford)

Further details may be had from:

Dr. J.J. Ward
Groups in Galway 90
Department of Mathematics
University College Galway
Galway, Ireland.
(MATWARD@CS8700.UCG.IE)

Third September Meeting of the IMS
Dublin City University

The 1990 September Meeting of the Irish Mathematical Society will take place in Dublin City University. Further details will appear in the next issue.

IMACS 91
13th IMACS World Congress on Computation and Applied Mathematics

This conference will take place in Trinity College Dublin on July 22–26, 1991 under the Chairmanship of Professor John J. H. Miller. Preliminary manuscripts (original or survey papers) and proposals for the organization of sessions are invited, and may be addressed to:

Professor J.J.H. Miller
General Chairman IMACS 91
PO Box 5, Dun Laoire, Co. Dublin.

Enquiries about all other Congress related matters should be directed to:

Ms. Paulene McKeever
IMACS91
40 Millview Lawns
Malahide
Co. Dublin
Fax: (+353–1) 802523

TEX90
University College Cork — 10–12 September 1990

The 1990 TEX meeting in Cork is the fifth European TEX conference. This is also the first TEX Users Group (TUG) meeting to take place outside North America. Peter Flynn of University College Cork is Programme Co-Chairman and Local Coordinator. Further details may be obtained from:

TEX90 Office
Computer Centre
University College
Cork
(TEX90@VAX1.UCG.IE)

ARTICLES

Gödelian Incompleteness and Paraconsistent Logics

Or: why Gödel's Paradox is really a dilemma

D.H. Cohen

There are three parts to this largely expository discussion. First, Gödel's Incompleteness Theorem is differentiated from other sorts of questions in mathematics. From this a classification scheme for some of mathematics' more perplexing situations emerges[1]. The Incompleteness result is then presented, explained, and located in that scheme. Finally, in light of this, and some points of language and philosophy made in passing, a new response to Gödel's result is suggested.

1 Puzzles and other problems

Gödel's Theorem poses an altogether new *sort* of problem for mathematicians. It establishes that some very good mathematical questions do not have equally good mathematical answers. Moreover, the fault is not in what counts as a mathematically acceptable question; it is in what counts as a mathematically acceptable answer. The problem is endemic to the entire mathematical enterprise.

The idea that some good questions do not have equally good answers is troubling but hardly new. There are many other situations involving answerless questions without anyone suggesting that they represent crises for human thought. But Gödel's proof *does* pose just that sort of crisis. Identifying just what distinguishes it from similar problems is the best first step towards

framing a rational response. To that end, comparisons with other notable open questions in mathematics will be made. The first part of the strategy is to uncover the implicit theoretical framework supporting judgements that some problems *are* similar to others. The results are then applied to Gödel's Theorem in particular.

Consider "Fermat's Last Theorem." It might be thought similar to Gödel's Theorem since it may be an example of an undecidable question[2]. A good many subcases have been closed, but no comprehensive proof or disproof has yet been found. Only such (dis-)proof would qualify as a definitive answer. As a working hypothesis, however, a good assumption is that some proof or refuting counterexample will be produced eventually. Without this assumption, further progress could not reasonably be expected. It is considered a problem with no solution *yet*, and problems with no solutions yet are simply unfinished business. They are like the intellectual itch of a particularly difficult daily crossword puzzle the day before the solution is printed. Since the same community response and attitude is present, the term PUZZLE is appropriate. The assumption is that there is a unique solution[3].

The same general response seems to be called for with respect to the humanly uncheckable proof of the Four-Colour Theorem. Although it introduces an empirical component into the practice of pure mathematics, that practice can be sufficiently distanced from theory so as not to pose any conceptual problems. The theorem is either true or false; the proof is either valid or not. Human capacities are not the issue. Any faith in Platonic Realism is not really challenged.

A second sort of unanswered question does raise problems for an uncritical realism, for the belief that there is a truth of the matter out there for us to discover or not, depending on the acuteness of our powers and the blessings of the mathematical muse. Cantor's Continuum Hypothesis typifies this set. Is the cardinality of the continuum the second transfinite cardinal? What differentiates this problem from Fermat's is that we know the resources at hand are *insufficient* to decide it. The hypothesis is provably *consistent* with but *independent* of the axioms of set-theory: either it or its negation can be added to those axioms without yielding contradiction ... provided only that the original axioms were consistent. The situation is analogous to Euclid's "fifth postulate". Who knows but a range of "Non-Cantorian" set-theories are just waiting to be developed — waiting, presumably, in the same place that Non-Euclidean geometries were waiting for Riemann and Lobaschevsky, wherever that was[4].

Gödel's own position was an extreme Platonism, maintaining that the Continuum Hypothesis was either objectively true or false, depending on whether or not it accurately described the behaviour of transfinite cardinals[5]. They are out there, just like Kurdistan or Pago Pago, and it is the business of the mathematician to be a geographer of this special realm of abstract entities. Even the idea that there are alternative set-theories waiting to be discovered is a kind of Realism, albeit pluralistic about the theories themselves.

This sort of situation is a DILEMMA. Several options are available, any one of which can be developed and applied. Thus, the basis for choosing one rather than another must be based on something *external* to the different systems, say, the way the world is in its ultimate metaphysical construction, or the desiderata surrounding some specific research program or computational context. The situation would be analogous to a jigsaw puzzle that could be put together in two different ways or a single crossword grid and set of clues that could be "correctly" solved in several distinct ways[6]. In a dillema, the existing conceptual framework needs additional information. What is accepted or established may be fine as far as it goes, but it simply doesn't go far enough.

The opposite situation also arises, cases where we have, as it were, too much "information". In this case, something has to be discarded to resolve the issue rather than something having to be added to settle things. The search, then, is for a likely belief to jettison. This is a PARADOX. The common pattern is that established and accepted theses give rise to an absurdity, or even an outright contradiction. Russell's Paradox is an example of this: the set of all sets that are not members of themselves must be — but cannot be — a member of itself. This amounts to a contradiction; something has to go. We could, if we were so inclined, abandon the belief that the world is contradiction-free, or that sentences cannot be both true and false simultaneously, or that set-theory is worth pursuing. The consensus has been that the unrestricted version of the set-abstraction axiom, although "obvious", is the source of the problem. Obviousness is not always the mark of truth. Restricting its range of applicability preserves just about all of naive set-theory. Nothing so radical as an overhaul of the underlying logic is required. So, by an implicit appeal to a principle of "minimal mutilation," restricting the relevant axiom is the change that is usually made[7].

Russell's Paradox, while forcing some revision in mathematical beliefs and practice, presents no real threat to the main Platonist tenet of objective mathematical truth. It merely challenges the secondary assumption that some particular statement of the set-abstraction axiom is part of that truth. A mistake was made, big deal. Of course, at the time of its discovery it was not as easy to be so blithe about revision. We know now what Russell did not know then: effectively restricting the schema is entirely possible and not unduly burdensome.

Ironically, this sort of Realism, which is Gödel's own position, is rather less viable when confronted by Gödel's Incompleteness Theorem. The theorem presents a conceptual paradox requiring revision of very fundamental beliefs, and possibly even the logic that holds everything together.

2 Gödel's Theorem

The theorem establishes, in brief, that there is an unavoidable mismatch between mathematical truths and mathematical theorems. The two sets cannot coincide. No matter how arithmetic is packaged, there must be either some truths that elude the proof-theoretic apparatus, or else some falsehoods that sneak their way into theoremhood.

Considerations of space prevent a complete rehearsal of the details of Gödel's proof here. However, an approximation is at hand, something a bit easier, but still in the same neighbourhood. It can be proved (perhaps contrary to expectations but provable nonetheless) that no mathematician is omniscient. By "omniscient" I mean believing *all and only* the true sentences. If it were simply a matter of believing everything that is true, it would be relatively easy: believe everything. Believe that $2 + 3$ is 5 but also believe that $2 + 3$ is not 5, that it is 6 and that it is not, that cabbages are kings and that they are not, that taxes can be lowered and government revenues raised at the same time, and so on. Similarly, believing *only* truths is also relatively easy: don't believe anything. Doubt that $2 + 3$ is 6 and doubt that it is not, and so on. (This may not be such a bad idea. Descartes tried it and managed to get his co-ordinates straight.) The trick, clearly, is to manage both at the same time, to believe all the truths *and* disbelieve all the falsehoods. Now consider this sentence:

(N) Prof. N. Üllset does not believe this sentence

(letting Prof. Üllset represent an arbitrary mathematician). If he believes it, it is false and he has a false belief. If he does not, it is true and there is a truth not in his belief-set. Either way, sadly, he falls short of omniscience.

This is what Gödel managed to do for arithmetic. He showed that any language rich enough to express what arithmetic needs to express will also be able to express an equivalent of N-sentence, saying roughly, "Arithmetic cannot prove this sentence".

The proof centers on the notion of *computability*. The idea is that all functions of a certain type ought to be expressible in any language suitable for mathematics. The type in mind, *recursive functions*, is conservatively characterized: a few (rather boring) functions are taken as primitive and means for constructing new ones are provided. The given functions are (1) the constant zero-function, (2) the successor function, and (3) projection functions which simply pick out the i^{th} member of a given n-tuple. Additional functions may be built up either by (4) function-composition or by (5) recursion. That's it.

The next stage involves showing that such purely syntatic concepts as *term*, *function*, and *well-formed formula* (wff) are representable by constructible functions. The vehicle for this is an assignment of numbers to each concatenation of symbols in the language. It is then shown that the *sets of numbers* corresponding to terms, wffs, and the rest, can be defined by recursive functions. For example, a one-place recursive function can be constructed which has the value 1 if and only if its argument is a number corresponding to a well-formed formula; otherwise, it has the value 0.

In addition to concepts that are syntactic in the grammatical sense, some concepts which are syntactic in the proof-theoretic sense are similarly representable by recursive functions. These include the concepts of *axiom*, *substitution instance*, and even of *proof*! That is, whether or not a given sequence of formulas is a valid proof is the kind of question that can be answered by a Türing machine: a program can be written which will correcting answer, after a finite number of steps, "yes" or "no" to the question "Is this collection of symbols a legitimate proof?"

Theoremhood, however, is *not* recursive. Given a sequence of wffs, it can be definitely decided that it does or does not constitute a proof. If it does, then the last wff is certainly a theorem. What cannot be devised is an effective test which starts with a single formula and correctly answers yes, it *can* be the last line of a proof sequence, or not it *cannot*.

This means that arithmetic is *undecidable*. There is no alogirithm for theoremhood. Undecidability leaves it open as to whether there might be proofs that are undiscover*ed* or even recursively unrecover*able* for every mathematical truth. The proofs could be out there, alongside the undeveloped

non-Cantorian set-theories, waiting for discovery or doing whatever it is that unknown truths do. However, Gödel's Theorem actually proves something stronger than undecidability, so this picture is wrong. Gödel's Theorem is an *incompleteness* theorem, proving that at least some of those imagined proofs for each and every truth are not there — no matter where "there" is.

The way this works is that *theoremhood* is an expressible concept within the language of recursive functions, although not itself recursive. The notion of proof is recursive, so theoremhood is easily recoverable. If $ded(x, y)$ holds just in case x is the number associated with a sequence of formulas that is a proof and y is the number associated with the formula which is the last line of that proof, then $Th(y)$, defined as $(Ex)ded(x, y)$, defines theoremhood. The predicate Th holds of just those numbers associated with theorems.

The hard part is re-creating the kind of self-reference that is the N-sentence — "Prof. N. Üllset does not believe *this* sentence". The recursiveness of substitution allows that. Let $sub(x, n, a)$ represent the substitution relation. Or, more exactly, the value of the $sub(x, n, a)$ is the number associated with the formula that results from substituting the expression associated with the number n for the variable associated with the number x in the formula associated with the number a.

Now, consider the formula:

$$\mathbf{I}: \quad \neg Th(sub(k_x, x, x)),$$

where k_x is the number associated with the symbol for the variable x. This entire formula is itself associated with some number, its "Gödel number". Let it be i. Now consider the formula:

$$\mathbf{J}: \quad \neg Th(sub(k_x, i, i)).$$

This says that the result of substituting the number i for the variable x in the formula with Gödel number i is not a theorem. The formula with Gödel number i is I, so this says that the result of substituting i for x in I is a non-theorem. The result of that substition is precisely J, so J says in effect, "J is not a theorem". If it is a theorem, it's a false theorem, if it is not a theorem, it is a true non-theorem. Mathematics, no matter how axiomatized, is not "omniscient". (The notion of proof is system relative, so expanding the system by adding the unprovable sentence as a new axiom wouldn't help; a new "Gödel sentence could always be generated.)[8]

Exactly why theoremhood is expressible but not recursive concerns the quantifier in its definition: $(Ex)ded(x, y)$ — *there is* some number answering to a sequence of formulas which is a proof of the given formula. There is no upper bound that can be given on the length or complexity of the proof simply from the syntactic complexity of the putative theorem. We could, were we so determined or demented, ask if 1 is the number of a proof of a given wff y, then ask if perhaps 2 is, then 3, and so on. Practical considerations aside, this method would find a proof of y, if there were one, eventually. But if there were *no* such proof, it would continue indefinitely, never getting a negative answer. There is no point at which one could say, "There has been no proof yet so none exists".

The situation is analogous to a variation of Goldbach's Conjecture. Let us call an even number a "Goldbach Number" if it is indeed the sum of two prime numbers. For any given even number n it can be determined definitively whether or not it is a Goldbach number. Simply check all the pairs of natural numbers whose sum if n. There are only finitely many pairs to check. Suppose, in contrast, we wanted to know whether n is the *difference* of two primes instead of the sum, a "Bach-gold Number" instead of a Goldbach Number. The quantification in this version of the conjecture is unbounded — "there are two primes such that ..., " not, as was implicit in the first case, "there are two primes *less than n* ..." — so there are infinitely pairs to test.

As with Bach-gold numbers, no limit can be established beforehand on how high up the ladder of natural numbers one has to climb before one can confidently assert that some formula is *not* a theorem.

3 Gödel's Dilemma.

The impossibility of an "omniscient" axiomatization of arithmetic is no less than that of an omniscient mathematician. But just as a mathematician is given two choices — either a false belief or an unbelieved truth — so too arithmetic has two choices: *inconsistency* or *incompleteness*. That is, if we are willing to consider inconsistency as a viable possibility, then the absurdity of mathematical incompleteness is no longer paradoxical; it is more of a *dilemma*!

But is inconsistency a viable option? Classically, no. Standard truth-functional accounts of implication maintain that from a single contradiction anything whatsoever may be legitimately inferred. A set of wffs G has the wff A as a logical consequence if it is impossible for all of the members of G to be true while A is false. When G is inconsistent, i.e., when its members cannot be simultaneously true, trivially satisfied for every A. The proof-theory mirrors this semantic account. A simple proof of B from $A \& \neg A$ is:

1.	$A \& \neg A$	premiss
2.	A	1, &-Elim ("Simplification")
3.	$A \vee B$	2, ∨-In ("Addition")
4.	$\neg A$	1, &-Elim
5.	B	3, 4 ∨-Elim ("Disjunctive Syllogism")

Both semantically and proof-theoretically, the admission of any inconsistency annihilates the theory.

Recent work in "Paraconsistent Logics," however, has shown that inconsistent systems can be viable. Paraconsistent logics are logics that can tolerate contradiction without degenerating into triviality. On the proof-theoretic side, this involves putting some restrictions on the patterns of proof permitted. The semantic innovation is to abandon the idea that the implication connective is entirely a truth-functional one.

Paraconsistent logics have been motivated in a variety of ways. Often, the motivation is the failure of the truth-functional analysis of if-then sentences. It fails as a model for the use of such conditionals in ordinary discourse[9] and it fails to provide the necessary conceptual framework for non-trivial reasoning from inconsistent premisses, i.e., for *reductio ad absurdum* reasoning. Moreover, if a logic is to be an information processing tool, the *pre-requisite* of consistency is self-defeating: it is doubtful how many human intellectual endeavours are consistent. And, needless to say, managing to prove consistency prior to *any* use of logic would be a great acomplishment.

The two important questions to address are the *how* and the *why* of inconsistent arithmetics. First the why: Why even consider a system that is known beforehand to have at least this one big flaw? The answer, in part, is that this might not be a flaw at all. It is important to keep in mind that part of the task of describing the world consists in *devising a language* with which to do so. Not even the most extreme Realist could deny it. Objective world or not, the choice of a vocabulary is a determinant of the shape of the resultant theory. The preliminary task of choosing or designing a language is not trivial. It is something that can be done well or poorly. It is a task using skills and criteria for success quite apart from those used in the subsequent descriptive operation. Also, there is nothing at the outset of the enterprise that guarantees success. On the contrary, Gödel's Theorem can be read as guaranteeing at least partial failure!

Success or failure aside, it is a mistake to minimize the contribution that the language makes to description. The language used makes its own appreciable contribution. Certain sentences may be certified as true *by the lanaguage itself* regardless of how the world is. "If it is raining, then it is indeed raining" is certainly true and just as certainly independent of the weather. No object can be in two places at the same time. Of course, but is this really a profound and *a priori* fact about objects or is it a perfectly natural consequence of the way we use the word "object", of the way we count objects, and of the way we decide what is to count as a single object? Even if an object could be in two places at once, we wouldn't count it as one object in two places but as two objects. Likewise, certain sentences are certified as false by the language itself. Anyone seriously asserting "It's raining but it isn't" or "Santa Clause does not exist although I sincerely believe he really does" would be guilty of a kind of linguistic incompetence.

Could the set of sentences certified as true by the language and the set certified as false by that language intersect? Nothing rules it out. This may be what the Liar's Paradox is all about. "This sentence is false" is both true and false according to the rules of language (And also neither true nor false, by those same rules)[10]. But what consequences about the world should one be able to draw from that?

This might be crucial to understanding Gödel's Theorem. Incompleteness and inconsistency represent genuine alternatives for arithmetic. Inconsistency can be accepted, if only as a pathological consequence of any language that permits self-reference. It is indeed an avenue worthy of further exploration. The underlying logic would have to be adapted accordingly, but that can be done. Localise the inconsistency; contain its effects. If it turns out that sentence J and its negation are both provable, what follows? Well, it follows that J is a theorem; it also follows that $\neg J$ is a theorem. So are both $J \vee A$ and $\neg J \vee A$ although A might or might not be deducible. There is no reason to suppose that the consequences extend to Fermat's conjecture or the next general election or why the sea is boiling hot and whether pigs have wings.

But *how* can an inconsistency be localized? Isn't the proof above incontrovertible? Two philosophers means three opinions, so of course it is not beyond dispute. Indeed, I think the proof is demonstrably fallacious — a case of "Begging the Question." Specifically, the disjunctive syllogism (DS) is the last line is illegitimate. Ordinarily, the reasoning from $A \vee B$ and $\neg A$ to B is quite unexceptionable: if it is either A or B, and it isn't A, then it simply must be B. And if someone asks why it has to be B the answer is because it isn't A. We already know A is false. But in this particular case we also know that it *is* A! That is, if what we are given is that A is both true and false, then we cannot later appeal to the principle that nothing can be both true and false. Taking $A \& \neg A$ as a premiss requires that we suspend the principle of non-contradiction — and with it, the principle of disjunctive syllogism that relies on it.

The situation is analogous to this. Suppose I said that I think if Gauss were alive and doing his work today, he would not only be more famous than any other living mathematician, he would even be more famous than Michael Tyson, Michael Jackson, or Mikhail Gorbachev. You might respond that, sure, he's a great mathematician and would deserve it, but it's hard to believe that society would suddenly and at long last give mathematicians their due. To that I'd say, "But don't forget, if Gauss were living today, he'd be *212 years old* and how many 212 years olds are any good at mathematics at all?" The joke is obvious. Conversational implicatures demand the incorporation of certain beliefs and the suspension of others. Violating the implicit rules can have comic effects. If we are asked to suppose that Gauss were alive, we are generally meant to suppose, among other things, that this is possible and that it involves a minimal change of the result of our beliefs about the way the world is. Some beliefs must change, such as that there is no one quite like Gauss around, but most other beliefs need not be put aside, including the belief that the world just doesn't have 212 year old mathematicians in it. But it *could* have someone now pretty much like Gauss was 200 years ago, even though we may believe it does not.

The same general sort of thing is going on in the proof. If we are asked to suppose that $A \& \neg A$ then we are also asked to suppose that it is possible. That requires suspension of many other beliefs, including the universal applicability of DS. Implicitly, there is an appeal to "It can't be both A and $\neg A$" — in spite of the fact that $A \& \neg A$ is exactly what we were asked to suppose! If the information that A is true was used to get $A \vee B$, that forestalls using the information that A is *not* true, even though the negation of A is also supposed to be true. DS might or might not work for $B \vee C$ and $\neg B$, but it has been set aside for A.

Would the reformulation of arithmetic with a paraconsistent logic as its basis be workable? How much of mathematics could be recovered or reconstructed with additional restrictions on allowable methods of proof? Largely that depends on the exact restrictions adopted. Intuitionist mathematics is a relevant precedent. At the heart of Intuitionistic thought is a rejection of Platonism and an embrace of "constructivism". If mathematics is "identical with the exact part of our thought" then the process of human thought is integral to the subject. This includes, notably, the fact of human limitations and thought's temporal unfolding. The infinite is tolerated only insofar as it is exactly specified. Infinite sets are countenanced, for instance, only as potentialities and only if rules for construction can be given. Existence proofs are accepted only if they include a method for constructing that mathematical entity; a *reductio ad absurdum* of a non-existence claim would not suffice.

The Intuitionistic program has had its share of success in both logic and philosophy. For example, much work has been done on formalizing and researching Intuitionistic logics, demonstrating at least that the philosophical program can be given and exact and coherent formal basis[11]. Further, the same sorts of considerations that led to some of the negative reactions against Zermelo's original use of the axiom of choice in 1904 to prove the well-ordering theorem continue to play an important part in current debates in the philopophy of language[12]. Within mathematics proper, however, Intuitionism has had mixed results. On the one hand, the restrictions Intuitionists impose on non-constructive existence proofs undermine the whole of Cantorian transfinite arithmetic — as was desired. And much of classical mathematics can be recovered within their guidelines. On the other hand, Intuitionistic proofs can be unwieldy, and rejecting the axion of choice and its equivalents means forswearing perhaps more of set-theory and analysis than would be desired[13].

The Intuitionists' penchant for constructivism (alternatively: their squeamishness about the infinite) entails rejecting the law of the excluded middle: undecidable propositions are neither true nor false. Their logic has truth-value "gluts" or inconsistencies — sentences taking both truth- values — so is not really a paraconsistent logic. It is these logics, logics that deny the principle of non-contradiction, that are relevant here.

The best developed paraconsistent logics are from the family of systems known collectively as "Relevance Logics". Although some work has been done on Relevant arithmetics using Robinson's and Peano's axiomatisations of arithmetic and the logic systems R, RM, and E, the work has been neither as systematic nor as institutionalized as the Intuitionists'[14]. In part, this may be due to dissension in their ranks as to the appropriate system to use;[15] in part it may be simply due to the absence of a Brouwer-Heyting calibre combination of mathematician and logician. Nevertheless, there has certainly been

enough success to warrant further exploration of the "Inconsistency option" with respect to Gödel's Incompleteness Theorem.

Notes

1. This taxonomy of puzzles-paradoxes-dilemmas is developed in greater detail in D.H. Cohen 1988.

2. It has been suggested (by Martin Gardner 1989, p. 26n) that if the "theorem" really is undecidable, then it must be true: if it were false, there would be a counterexample which would decide things. This assumes the system is "omega-consistent", which amounts to Realism of a sort.

3. This is similar to the use of "puzzle" in Thomas Kuhn 1970. Kuhn suggests that "puzzle-solving" is the mark of a "normal" science.

4. See. P.J. Cohen and R. Hersh 1967.

5. See Kurt Gödel 1964.

6. See D.H. Cohen 1985.

7. W.V. Quine offers an excellent general discussion of paradoxes in the title essay Quine 1966.

8. E. Nagel and J.R. Newman 1959 is an excellent introduction to the notion of incompleteness. A good, more technical account is in J.W. Robbin 1969, pp.90–119.

9. Routley 1982 contains many counterexamples and an extended polemic against the truth-functional model for conditionals. Other discussions abound.

10. See Jennings and Johnston 1983.

11. Arend Heyting's formalization of Intuitionist logic has attracted the most attention. See Haack 1974, pp.91-103. The *Journal of Philosophical Logic* special issue on Intuitionism (v. 12, no. 2, May 1983) addresses a spectrum of the logical questions.

12. M.A.E. Dummett is perhaps the leading proponent of Intuitionisti thought in logic, philosophy of mathematics, and philosophy of language Dummett 1977 offers a good introduction to Intuitionism in mathemat ics; Dummett 1978 considers it in other contexts.

13. The consistency and independence proofs for the axiom of choice hav defused the issue somewhat. See Dummett 1977 or Heyting 1972.

14. See Anderson and Belnap 1975 for a quasi-official statement of the pro gram; see Routley 1982 for dissenting voices. Volume II of *Entailment*, edited by N. Belnap and J.M. Dunn, is due out shortly. Relevantly developed systems of mathematics will be included.

References

[1] A. Anderson and N. Belnap, *Entailment*, Vol. 1, Princeton, 1975.

[2] D.H. Cohen, *Putting Paradoxes to Pedagogical Use in Philosophy*, Teaching Philosophy 8 No. 4 (1985), 309–317.

[3] P. J. Cohen and R. Hersh, *Non-Cantorian Set-Theory*, Scientific American **217** No. 6, 104–116.

[4] D.H. Cohen, *Stalking the Wild Paradox*, Metaphilosophy 19 (1988), 25–31.

[5] M. Dummett, *Truth and Other Enigmas*, Harvard University Press, 1977.

[6] M. Dummett, *Elements of Intuitionism*, Claredon Press, Oxford, 1977.

[7] Martin Gardner in The New York Review of Books, 16 March, 1989, p. 26n.

[8] Kurt Gödel , *What is Cantor's Continuum Hypothesis*, reprinted in P. Benaceraff and H. Putnam, eds., *Philosophy of Mathematics*, Prentice-Hall, Englewood Cliffs, N. J., 1964.

[9] Susan Haack, *Deviant Logic*, Cambridge University Press, 1974.

[10] A Heyting, *Intuitionism: An Introduction*, North-Holland Publishing Company, Amsterdam, 1972.

[11] R.E. Jennings and D.K. Johnston, *Paradox-tolerant Logic*, Logigue et Analyse **26** Nos. 103–104 (1983), 291–307.

[12] Thomas Kuhn, *The Structure of Scientific Revolutions* 2nd ed., University of Chicago Press, 1970.

[13] E. Nagel and J.R. Newman, *Gödel's Proof*, Routledge and Kegan Paul, 1959.

[14] W.V. Quine, *Ways of Paradox*, Random House, New York, 1966.

[15] J.W. Robbin, *Mathematical Logic*, W.A. Benjamin, Inc., Amsterdam, 1969.

[16] R. Routley, R.K. Meyer, V. Plumwood and R.T. Brady, *Relevant Logics and Their Rivals*, Ridgeview Publishing Company, Atascadero, CA, 1982.

Department of Philosophy
Colby College
Waterville, ME 04901, USA

Numerical Analysis of Semiconductor Devices

John Carroll

1 Introduction

Although the device modeling problem is perhaps one of the most difficult computational problems in current research, there is a dramatic increase in reliance on process and device simulation tools for technology development and device optimization. The characteristic feature of early modeling was the separation of the interior into different regions, whose treatment could be simplified by various assumptions like special doping profiles, completion regions and quasi-neutrality. These separately treated regions were simply connected to produce the overall solution.

Fully numerical device modeling based on partial differential equations which describe all different regions of semiconductor devices in a single unified manner was first suggested by Gummel [11] for the one-dimensional bipolar transistor. This approach was further developed and applied to pn-junction theory by De Mari [7] and to IMPATT diodes by Sharfetter and Gummel [15]. A two-dimensional numerical analysis of a semiconductor device was first undertaken by Kennedy and O'Brien [12] who investigated the junction field effect transistor. Since then, two-dimensional modeling has been applied to almost all important semiconductor devices.

It is now universally accepted that device simulation tools provide a worthwhile alternative to the conventional experimental approach of running wafer lots through a process line. We present a brief overview of the numerical techniques which are being employed to solve the coupled system of highly nonlinear partial differential equations which model the behaviour of electron and holes in a semiconductor structure. We begin in Section 2 with an introduction to the basic semiconductor equations in order to define the relevant physical variables. The scaling procedures and dependent variable alternatives are considered briefly in Section 3 while Section 4 concentrates mainly on the discrete form of the mathematical equations. The nonlinear and linear solu-

tion strategies as well as certain algorithmic factors are discussed in Section 5.

2 The Semiconductor Equations

The partial differential equations which model the steady-state and transient behaviour of carriers under the influence of external fields can be derived, in a semiclassical framework, from the Boltzmann Transport Equation. In this way, carrier motion is considered as a series of acceleration events (described by classical mechanics) and scattering events (described by quantum mechanics). If we assume that the response of carriers to a change in the electric field is considerably faster than the rate of change of the field itself, we can write the basic equations of semiconductor transport in the most commonly used form [16] as follows.

The Poisson equation

$$\epsilon_s \nabla \cdot \mathbf{E} = -\epsilon_s \nabla^2 \psi = \rho \tag{1}$$

relates the total space charge ρ to the divergence of the electric field $\mathbf{E}$, which defines the electrostatic potential ψ as

$$\mathbf{E} = -\nabla \psi \tag{2}$$

Under the assumption of total ionization, the total space charge ρ is given as

$$\rho = -q(n - p + \Gamma) \tag{3}$$

where $\Gamma = N_D^+ - N_A^-$ is the total electrically active net impurity concentration, q is the electric charge, and n and p are the electron and hole densities respectively. The connection between the behaviour of the carrier densities and the electric field is given by the current equations for electrons and holes

$$\mathbf{J}_n = q\mu_n n\mathbf{E} - qD_n \nabla n \tag{4}$$

$$\mathbf{J}_p = q\mu_p p\mathbf{E} + qD_p \nabla p \tag{5}$$

where μ_n and μ_p are the electron and hole mobilities and D_n and D_p are the corresponding diffusion coefficients. Both mobilities and diffusion coefficients depend on the temperature, the doping level and the electric field.

The electron and hole concentrations may be written as

$$n = n_{ie} \exp \frac{q(\psi - \phi_n)}{kT} \tag{6}$$

$$p = n_{ie} \exp \frac{q(\phi_p - \psi)}{kT} \tag{7}$$

where we have defined the *quasi-Fermi potentials* ϕ_n and ϕ_p [16]. The factor n_{ie} is the effective carrier concentration. For low doping, n_{ie} approaches the intrinsic carrier concentration n_i. If we assume the *Einstein relation* [17] for both electrons and holes

$$D = \mu \frac{kT}{q} \tag{8}$$

then equations (4) and (5) can be re-written using (6) and (7) as

$$\mathbf{J}_n = q\mu_n n \nabla \phi_n \tag{9}$$

$$\mathbf{J}_p = q\mu_p p \nabla \phi_p \tag{10}$$

The continuity equations for electrons and holes are given by

$$\frac{\partial n}{\partial t} = -\frac{1}{q}\nabla \cdot \mathbf{J}_n - R + G \tag{11}$$

$$\frac{\partial p}{\partial t} = \frac{1}{q}\nabla \cdot \mathbf{J}_p - R + G \tag{12}$$

where G and R represent generation and recombination processes respectively. In some applications, generation is ignored and the Shockley-Read-Hall steady state recombination is adopted, namely

$$R = \frac{pn - n_i^2}{\tau_p(n + n_i) + \tau_n(p + n_i)} \tag{13}$$

where τ_n and τ_p are respectively the electron and hole lifetimes.

In the time dependent case, the equation of total current continuity couples the change in electric field strength to the current densities

$$\nabla \cdot \mathbf{J}_T = \nabla \cdot \left(\epsilon \frac{\partial \mathbf{E}}{\partial t} + \mathbf{J}_n + \mathbf{J}_p \right) \tag{14}$$

where J_T is the total current, which consists of both the conduction components J_n and J_p and the displacement current $\epsilon\, \partial \mathrm{E}/\partial t$.

The boundary conditions for semiconductor devices are given by neutrality and equilibrium conditions, namely

$$pn = n_i^2 \tag{15}$$

$$n - p + \Gamma = 0 \tag{16}$$

Thus at a Dirichlet contact, the three potentials are

$$\psi = V_c(t) + V_{bi} \tag{17}$$

$$\phi_n = \phi_p = V_c(t) \tag{18}$$

where V_{bi} and $V_c(t)$ are the built-in and contact voltages respectively. When external circuit elements are applied to the device, the contact voltage becomes an unknown and is given by

$$\frac{1}{R}[V_c(t) - V_a(t) + C\frac{d}{dt}[V_c(t) - V_a(t)] + \int_\Gamma \nabla \cdot \mathbf{J}\, dl = 0 \tag{19}$$

where R, C and $V_a(t)$ are the resistance, capacitance and applied voltage respectively and Γ is an appropriate contour surrounding the contact.

The complete set of semiconductor equations is given by (1), (9), (10), (11) and (12) together with appropriate initial and boundary conditions and, as we will see in the next section, this set of coupled nonlinear partial differential equations are usually written in dimensionless form appropriate for numerical simulation.

3 Problem Formulation

Before proceeding to the numerical solution, there are a number of factors which must be considered, notably the choice of dependent variables and an appropriate scaling of the equations. The choice of variables can crucially affect the linearity of the equations as well as the symmetry of the iteration matrix. Scaling is important as the dependent variables can be of different order of magnitudes and show a strongly different behaviour in regions with small and large space charge.

For the system with dependent variables $\{\psi, n, p\}$, a standard approach to scaling was proposed by de Mari, see [8], where ψ is scaled by the thermal voltage $V_t = \frac{kT}{q}$; n and p are scaled by n_i and the independent variables are scaled such that all multiplying constants in Poisson's equation become unity — all spatial quantities are scaled by the intrinsic *Debye* length $L_i = \sqrt{\epsilon_s kT/qn_i}$. This approach may be physically reasonable but suffers in that the variables n and p are still several orders of magnitude larger than ψ. An alternative "singular perturbation approach" was proposed in [1] which effectively reduces the variables ψ, n and p to the same order of magnitude. In this case, the variables n and p are scaled by the maximum absolute value of the net doping Γ and the independent variables are scaled by the characteristic length of the device.

Following [8] for example, the basic equations can be written in normalized form as

$$\begin{aligned} g_1(\psi, n, p) &= -\nabla^2\psi + n - p - k_1 = 0 \\ g_2(\psi, n, p) &= \frac{\partial n}{\partial t} + \nabla \cdot J_n + k_2 = 0 \\ g_3(\psi, n, p) &= \frac{\partial p}{\partial t} - \nabla \cdot J_p + k_2 = 0 \end{aligned} \qquad (20)$$

where ψ, n, p, J_n and J_p are the normalized electrostatic potential, carrier densities and current densities respectively, and k_1 and k_2 represent the normalized impurity concentration and generation-recombination terms. We could also write the system in terms of the normalized quasi-Fermi potentials ϕ_n and ϕ_p as follows:

$$\begin{aligned} g_1(\psi, \phi_n, \phi_p) &= -\nabla^2\psi + e^{\psi-\phi_n} - e^{\phi_p-\psi} - k_1 = 0 \\ g_2(\psi, \phi_n, \phi_p) &= \frac{\partial e^{\psi-\phi_n}}{\partial t} + \nabla \cdot (\mu_n e^{\psi-\phi_n}\nabla\phi_n) + k_2 = 0 \\ g_3(\psi, \phi_n, \phi_p) &= \frac{\partial e^{\phi_p-\psi}}{\partial t} - \nabla \cdot (\mu_p e^{\phi_p-\psi}\nabla\phi_p) + k_2 = 0 \end{aligned} \qquad (21)$$

There are many choices for the set of dependent variables and, in what follows, we will refer to the arbitrary choice u, v, w as including such possibilities as $\{\psi, n, p\}$, $\{\psi, \phi_n, \phi_p\}$, $\{\psi, \psi - \phi_n, \psi - \phi_p\}$ or perhaps $\{\psi, \Phi_n, \Phi_p\}$ where $\Phi_n = e^{-\phi_n}$ and $\Phi_p = e^{\phi_p}$ all of which have appeared in recent publications. The first set of variables used in device simulation was the potential and carrier concentrations but, because the continuity matrix took on a positive definite form, many workers switched to exponentials of the quasi-Fermi potentials. The quasi-Fermi potentials themselves are now quite popular and offer the advantage of reducing the numerical range of the dependent variables.

4 Discretization

Software for device analysis could roughly be classified in two categories. One of these involves codes for analysis of specified device types. These codes often utilize a regular mesh, finite difference discretization and iterative solution methods. Since the device type is known, some behaviour of the structure can be predicted. This information can be used to select appropriate physical parameters and to improve the mesh generation and the equation solution methods. The second category contains codes for analysis of an arbitrary semiconductor structure. They are characterized by a high degree of flexibility which makes the user more responsible for the final results. The finite element method (fem) has many properties which motivate its usage in this category.

As an illustration, we will consider the steady state case, i.e. where $\frac{\partial n}{\partial t} = \frac{\partial p}{\partial t} = 0$ with exponentials of the quasi-Fermi levels, $\Phi_n = \exp(-\phi_n)$ and $\Phi_p = \exp(\phi_p)$, as dependent variables. Ignoring the generation term, we consider the one-dimensional form

$$\begin{aligned} -\frac{d^2\psi}{dx^2} &= \Phi_p e^{-\psi} - \Phi_n e^{\psi} + \Gamma(x) \\ \frac{dJ_p}{dx} &= R \\ \frac{dJ_n}{dx} &= -R \end{aligned} \qquad (22)$$

where

$$J_p = \frac{d\Phi_p}{dx} e^{-\psi}$$

$$J_n = -\frac{d\Phi_n}{dx} e^{\psi}$$

A constant space charge density results in a parabolic function for ψ and we therefore use a standard Taylor's expansion to discretize (22). First we define the non-uniform mesh on [0, 1]

$$x_0 = 0$$

$$\begin{aligned} x_{j+1} &= x_j + h_j, 0 \le j \le N-1 \\ x_N &= 1 \end{aligned}$$

where the h_j are suitably chosen to model rapidly varying solution behaviour and are typically constrained by the inequality

$$\max_j\left(\frac{h_{j+1}}{h_j}, \frac{h_j}{h_{j+1}}\right) \le 2$$

for all $0 \le j \le N-1$. Before linearization, the difference scheme corresponding to Poisson's equation has the form

$$-D_+D_-\psi_j = \Phi_{p,j}e^{-\psi_j} - \Phi_{n,j}e^{\psi_j} + \Gamma_j \tag{23}$$

where the difference operator D_+D_- is the standard three-point difference operator on a non-uniform mesh

$$D_+D_-y_i = \frac{2}{h_ih_{i-1}(h_i+h_{i-1})}[h_iy_{i-1} - (h_i+h_{i-1})y_i + h_{i-1}y_{i+1}]$$

The standard Taylor expansion has proved inadequate however for the continuity equations since this approach, for the quasi-Fermi potentials for example, would suggest exponential current density profiles. In reality current densities are known not to vary very rapidly with respect to the spatial coordinate. Hence we employ the Sharfetter-Gummel discretization (see [15]) which incorporates approximate integrals of the basic equations into the formulation of the difference equations and yields constant current densities between adjacent mesh points. Taking the hole current for example and recalling that

$$\frac{d\Phi_p}{dx} = J_pe^{\psi}$$

we integrate over the interval $[x_j, x_{j+1}]$ on which we assume that the current density is constant and that the electrostatic potential is linear. Thus we get

$$\begin{aligned} \Phi_{p,j+1} - \Phi_{p,j} &= \int_{x_j}^{x_{j+1}} J_pe^{\psi}\,dx \\ &= J_{p,j+\frac{1}{2}}\int_{x_j}^{x_{j+1}} \frac{d(e^{\psi})}{\frac{d\psi}{dx}} \\ &= h_jJ_{p,j+\frac{1}{2}}\frac{e^{\psi_{j+1}} - e^{\psi_j}}{\psi_{j+1}-\psi_j} \end{aligned}$$

or equivalently we write

$$J_{p.j+\frac{1}{2}} = e^{-\psi_j}D_+\Phi_{p,j}B(\Delta\psi_j)$$

where we introduce the notation

$$D_+z_i = \frac{z_{i+1}-z_i}{h_i}$$

$$\Delta z_i = z_{i+1} - z_i$$

$$B(x) = \frac{x}{e^x - 1}$$

The corresponding difference expression over the interval $[x_{j-1}, x_j]$ may be found analogously. The current density approximations may be summarised as follows

$$J_{p,j+\frac{1}{2}} = e^{-\psi_j}B(\Delta\psi_j)D_+\Phi_{p,j} \tag{24}$$

$$J_{p,j-\frac{1}{2}} = e^{-\psi_j}B(-\Delta\psi_{j-1})D_+\Phi_{p,j-1} \tag{25}$$

$$J_{n,j+\frac{1}{2}} = -e^{\psi_j}B(-\Delta\psi_j)D_+\Phi_{n,j} \tag{26}$$

$$J_{n,j-\frac{1}{2}} = -e^{\psi_j}B(\Delta\psi_{j-1})D_+\Phi_{n,j-1} \tag{27}$$

Using a standard centered difference approximation, the continuity equations can be discretized as follows

$$\frac{2}{h_{j-1}+h_j}[J_{p,j+\frac{1}{2}} - J_{p,j-\frac{1}{2}}] = R_j \tag{28}$$

$$\frac{2}{h_{j-1}+h_j}[J_{n,j+\frac{1}{2}} - J_{n,j-\frac{1}{2}}] = -R_j \tag{29}$$

and, on using equations (24) to (27), these become

$$B(-\Delta\psi_{j-1})D_+\Phi_{p,j-1} - B(\Delta\psi_j)D_+\Phi_{p,j} = -\frac{(h_{j-1}-h_j)}{2}e^{\psi_j}R_j \tag{30}$$

$$B(\Delta\psi_{j-1})D_+\Phi_{n,j-1} - B(-\Delta\psi_j)D_+\Phi_{n,j} = -\frac{(h_{j-1}+h_j)}{2}e^{-\psi_j}R_j \tag{31}$$

The difference equations (23), (30) and (31), subject to appropriate boundary conditions provide the basis for the numerical solution procedures in one-dimensional steady state simulations.

Any discretization scheme for the semiconductor device equations should possess certain desirable properties. In particular, it should

- function in arbitrary geometries
- be conservative
- provide adequate treatment of drift-diffusion terms
- allow convenient enforcement of boundary and interface conditions
- permit adaptive mesh construction
- be free of dimensional restrictions

There are a number of discretization strategies in current use — the "finite difference method", of which the foregoing is an example, the "finite box method" which is just a more general finite difference method and the "finite element method". No attempt is made to provide a serious mathematical preference for one method or the other. The finite difference method and the finite element method are frequently considered to be mutually independent from the beginning. However it is often a matter of interpretation only and one can sometimes obtain the exact same discrete equations from either a finite difference approach or a finite element approach. It should be noted however that finite difference formulae accounting for normal derivatives at a curved boundary are extremely awkward so that, for this type of problem, one should consider the finite element method.

5 Numerical Solution Procedures

The exponential dependency of n and p on ψ makes Poisson's equation nonlinear and the generation-recombination mechanisms couple the two current continuity equations and introduce strong nonlinearities. There are basically two different solution strategies adopted for the discretized system namely (a) the decoupled (or Gummel) and (b) the coupled (or simultaneous) procedure. Both require an initial guess of the solution followed by an adjustment of the guess until an acceptable degree of accuracy is obtained.

5.1 The Decoupled Approach

This approach is sometimes referred to as the Gummel iteration [11] by electrical engineers and the Jacobi / Gauss-Seidel iteration by mathematicians

and treats the three equations independently. The iteration proceeds by associating with each g_i the highest-order differential dependent variable (e.g. for the variables of equation (20), taking ψ in the first equation and n and p in the other two respectively). In solving $g_1 = 0$, v and w are treated as fixed to obtain a new solution for u. Let $z_k = (u_k, v_k, w_k)^T$. The iteration can be written as

$$\begin{aligned} g_1(z_k \rightarrow u_{k+1}) &= 0 \\ g_2(z_k, u_{k+1} \rightarrow v_{k+1}) &= 0 \\ g_3(z_k, u_{k+1}, v_{k+1} \rightarrow w_{k+1}) &= 0 \end{aligned} \tag{32}$$

where the variables left of $\rightarrow$ are considered as input variables. This symbolic representation of the iteration allows considerable flexibility in determining the sequence of one-variable equations.

The partitioning of the complete PDE system into a series of three equations that can be solved independently made this procedure very popular particularly in the early years of device modeling. The applicability of the Gauss-Seidel approach however depends critically on the level of current flow inside the device structure. In [4] the procedure has been found to work very well for conditions of low to medium current flow and negligible generation-recombination terms k_2. This corresponds to a weak coupling of the PDE system where the density or quasi-Fermi potentials act as small perturbations to the Poisson equation.

5.2 The Coupled Approach

Using the Gauss-Seidel iteration in physical situations where there is heavy coupling between the variables (typically in high current conditions) will usually prove to be difficult and, in most cases, convergence will not be realised. As a result, reasearchers quickly turned to a more robust procedure based on Newton's method. This requires both the assembly and the approximate solution of the system

$$\frac{\partial g_k}{\partial z_k} x = -g_k \tag{33}$$

where $g_k = g(u_k, v_k, w_k)$ with $g = (g_1, g_2, g_3)^T$. In expanded form this leads to the system

$$\begin{pmatrix} \frac{\partial g_1}{\partial u} & \frac{\partial g_1}{\partial v} & \frac{\partial g_1}{\partial w} \\ \frac{\partial g_2}{\partial u} & \frac{\partial g_2}{\partial v} & \frac{\partial g_2}{\partial w} \\ \frac{\partial g_3}{\partial u} & \frac{\partial g_3}{\partial v} & \frac{\partial g_3}{\partial w} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = - \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} \tag{34}$$

where the Jacobian matrix $\frac{\partial g}{\partial z}$ and the right hand side are evaluated at the k-th iteration (u_k, v_k, w_k). Unlike the decoupled algorithm, the simultaneous or coupled approach has the advantage that it is practically independent of the applied bias conditions and injection levels in the device. If the linear algebra modules are subsequently carefully chosen, it also proves to be sufficiently robust for general purpose codes. However, it suffers somewhat because of the increased requirements of CPU-performance and memory and, in this context, the use of approximate Newton methods are becoming more popular in an attempt to offset these factors.

5.3 Algorithmic Aspects

Assuming we wish to solve the nonlinear system of type (20) or (21) for example. Firstly, the large sparse linearized system

$$M_k x_k = -g(z_k) \equiv -g_k \tag{35}$$

is solved by the sparse direct or iterative methods (to be described) where M_k is an approximation to the exact Jacobian $\frac{\partial g}{\partial z}$. The next iterate is taken as

$$z_{k+1} = z_k + t_k x_k \tag{36}$$

where $t_k \in (0, 1]$ is chosen to satisfy the sufficient-decrease condition

$$1 - \frac{\| g_{k+1} \|}{\| g_k \|} > \epsilon\, t_k \tag{37}$$

and ϵ is the machine epsilon. In other words, the step-length parameter t_k damps the step x_k to insure that $\| g_{k+1} \| < \| g_k \|$ increasing the robustness of the nonlinear equation algorithm [3]. Moreover, t_k is biased towards unity so that a traditionally quadratically convergent method is recovered in the Newton-attraction region.

A popular Newton iterative method is the Newton-Richardson algorithm [4] which assumes that g_k can be written as

$$g_k = M_k - N_k \tag{38}$$

with

$$\| M_k^{-1} N_k \| = \| I - M_k^{-1} g_k \| \leq \rho_0 < 1 \tag{39}$$

for all k. In this case, M_k represents a previously factored Jacobian so that the method clearly attempts to exploit the fact that the time to factor a sparse matrix is much larger than the CPU time to perform a backsolve with a previously factored matrix. The x_k is found by an inner iteration

$$M_k(x_{k,l} - x_{k,l-1}) = -(\frac{\partial g_k}{\partial z_k} x_{k,l-1} + g_k) \tag{40}$$

where $x_{k,0} = 0$ and $x_k = x_{k,l_k}$ for some l_k. This inner iteration can be controlled by monitoring the quantities

$$\alpha_{k,l} = \frac{\| g_k + \frac{\partial g_k}{\partial x_k} x_{k,l} \|}{\| g_k \|} \tag{41}$$

Note that both quantities used to compute $\alpha_{k,l}$ are already available. The inner iteration is terminated when

$$\alpha_{k,l_k} \leq \alpha_0 \left(\frac{\| g_k \|}{\| g_0 \|} \right) \tag{42}$$

where $\alpha_0 \in (0, 1)$ is an experimentally determined parameter. We refer the interested reader to [3] for further details where it is shown that the resulting procedure is quadratically convergent.

5.4 Linear Equation Solution

The set of linear equations which must be solved at each iteration of the Newton procedure (outer iteration) is large and sparse. It may be symmetric or unsymmetric depending on the choice of dependent variables. The efficient solution of systems of linear equations is the key to any successful device simulator where typically the solution of a system of 6000 equations may be required. In terms of CPU and memory, the linear algebra portion of the overall solution process dominates. Iterative techniques are usually preferred since they can fully exploit the sparsity of the equation set. Equally important, they can exploit the fact that the linear equations are part of the outer loop and only need to be solved sufficiently accurately to ensure that the outer loop converges rapidly.

In simulating an arbitrarily shaped structure, the box integration technique is often used with the finite difference method. Using this method, the solution of Poisson's equation and the current continuity equations can be reduced to the solution of linear systems of equations involving symmetric band matrices. The symmetry of the coefficient matrix not only saves memory but also enables the application of the highly popular ICCG (Incomplete Cholesky-decomposition and Conjugate gradient) [13] method which results in a very high speed simulation.

For unsymmetric matrices, both preconditioned ORTHOMIN [2] and preconditioned conjugate gradient squared have been studied with the latter showing a slight advantage. The conjugate gradient iteration (inner iteration) is continued until the norm monitored in the outer loop has decreased and the residuals of the preconditioned equation set has been reduced by a factor of typically 10^3.

To obtain the required convergence whilst exploiting the sparse structure, the choice of preconditioner is critical. The preconditioner employed relies heavily on the method in which the equations are ordered. To ensure that the coupling between the three equations is taken into account, the equations and variables associated with a given node are ordered sequentially. In this way, the coupling between the equations is represented by 3 by 3 submatrices distributed within the matrix. In addition, the nodes are ordered in a way that nodes with strong coupling are neighbours — this is achieved by numbering the nodes sequentially along a path through the mesh that always advances by moving to the closest node which is not yet ordered. Thus strong coupling of nodes is assumed to be correlated with a close spacing of the nodes. With this ordering of the equations, the matrix describing the set of linear equations is dominated by the components in the 3 central diagonals of 3 by 3 submatrices. An exact inverse of this "tridiagonal block" structure can be economically calculated and used to precondition the linear system.

6 Conclusion

We have attempted to provide a simple overview of some of the techniques and issues which are involved in the numerical analysis of semiconductor devices. During the last several years, the modeling of basic semiconductor device structures has grown in importance as integrated circuit complexity has increased. To reduce risk and enable aggressive circuit design, efficient and accurate compact models are required for the simulation of circuits containing hundreds and thousands of devices. These compact models must be fully characterized for the target fabrication process and be valid over a wide range of operating conditions including temperature variations. The predictive capabilities of detailed numerical device and process simulators can be used to enhance device characterization techniques and to form a basis for technology optimization.

Simulators that can perform DC-analysis of specific device types in two space dimensions are today in use at many process development facilities and this level may be considered state of the art. With a reduction in the geometrical dimensions of, for example, the MOSFET, there is an increasing necessity for three-dimensional device simulators and current research is rapidly providing the building blocks. In addition to three-dimensional device analysis, the simulator must also be capable of (a) treating irregularly shaped structures; (b) one- and two-dimensional simulations and (c) device parameter calculations. Two three-dimensional simulators in current use are FIELDAY [5] and TRANAL [18], both of whom first appeared in 1980 and, since then, attention has been focused on reducing the large memory capacity and computing time required for the simulation.

References

[1] Agler W., Markowich P.A., Selberherr S., *A singular perturbation analysis of discretization methods for the parabolic semiconductor equations*, Proceedings of the NASECODE III Conference, J. Miller (Ed.), Boole Press, Dublin 1983.

[2] Axelsson O., *Conjugate gradient type methods for unsymmetric systems of linear equations*, Linear Algebra and its Applics. **29** (1980).

[3] Bank R.E., Rose D.J., *Parameter selection for Newton-like methods applicable for nonlinear partial differential equations*, SIAM J. Numer. Anal. 17 (1980), 806–822.

[4] Bank R.E., Coughran W.M., Fichtner W., Grosse E.H., Rose D.J., Smith R.K., *Transient simulation of silicon devices and circuits*, IEEE Trans Electron Devices, ED-32 (1985), 1092–2007.

[5] Buturla E.M., Cottrell P.E., Grossman B.M., Salsburg K.A., *Finite-element analysis of semiconductor devices: the FIELDAY program* , IBM J. Res. Dev. 25 (1981), 218–231.

[6] Carroll J., *A one dimensional numerical model of the NPN transistor under arbitrary conditions*, Proceedings of the NASECODE IV Conference, J. Miller (Ed.), Boole Press, Dublin 1985.

[7] De Mari, *An accurate numerical steady state one-dimensional solution of the pn-junction*, Solid-State Electron **11** (1968), 35–58.

[8] De Mari, *An accurate numerical one-dimensional solution of the pn-junction under arbitrary transient conditions*, Solid-State Electron **11** (1968), 1021–1053.

[9] Fichtner W., Rose D.J. (Eds), *Special Issue on Numerical Modeling of VLSI Devices* , IEEE Trans Electron Devices, ED-30 (1983) Number 9.

[10] Fichtner W., Rose D.J. (Eds), *Special Issue on Numerical Modeling of VLSI Devices* , IEEE Trans Electron Devices, ED-32 (1985) Number 10.

[11] Gummel H.K., *A self consistent iterative scheme for one-dimensional steady state transistor calculations*, IEEE Trans. ED-11 (1964), 455–465.

[12] Kennedy D.P., O'Brien R.R., *Two dimensional mathematical analysis of a planar type junction field effect transistor*, IBM J. Res. Dev. **13** (1969), 662–674.

[13] Meijerink J.A., van der Vorst H.A., *An iterative solution method for linear systems of which the coefficient is a symmetric M-matrix* , Math. Comput. **31** (1977), 148–162.

[14] Mock M.S., *Analysis of Mathematical Models of Semiconductor Devices*, Boole Press, Dublin 1983.

[15] Sharfetter D.L., Gummel H.K., *Large-signal analysis of a Silicon-Read diode oscillator*, IEEE Trans Electron Devices ED-16 (1969), 64–77.

[16] Shockley W., *Electrons and Holes in Semiconductor Devices*, van Nostrum, Princeton 1950.

[17] Smith R.A., *Semiconductors*, Cambridge University Press, New York 1978.

[18] Yoshii A., Kitazawa H., Tomiyama M., Horiguchi S., Sudo T., *A three-dimensional analysis of semiconductor devices*, IEEE Trans Electron Devices, ED-29 (1982), 184–189.

Department of Mathematics
Dublin City University

MATHEMATICAL EDUCATION

The Dynamic Role Of Mathematics In Business

Dr. Con Power

This article comprises the text of a talk given by Dr. Con Power, Director of Economic Policy, Confederation of Irish Industry, at a Seminar on "Mathematics : Industry–RTC–University Interaction" , organised by the National Sub-Commission for Mathematical Instruction of the Royal Irish Academy and held in the offices in the Royal Irish Academy, 19 Dawson Street, Dublin 2, on Wednesday, 13th September 1989.

Where did it all begin?

Historians claim that mathematics can trace its origins back about 6,000 years to the Middle East. The Babylonians and the Egyptians used elementary mathematical applications in their everyday life. Mathematics was a key element in the early application of science by the more advanced among the prehistoric people who were curious enough to try to ascertain the workings of the world around them. The Egyptians are probably the first recorded people to have used arithematic and geometry perhaps as long ago as 4,000 B.C. and this led to the Egyptians developing expertise in areas such as civil and structural engineering, medicine and astronomy. The Greeks made many advances on the intellectual side of geometry, physics, astronomy and biology.

Mathematics was used in ancient times not only for intellectual and scientific pursuits but also for accounting purposes. The Egyptians considered this

task so important that they entrusted the keeping of accounts to a hereditary class of scribes.

A historian could undoubtedly trace the history and development of mathematics from the early times of Babylon, Egypt, Greece, Rome and Byzantium through the Dark Ages to the new beginning from about the fifteenth century onwards. The new beginning in a commercial sense owes a lot to the Italians who set about the problems of keeping accounts, calculating compound interest and solving insurance problems which were part and parcel of their growing international trade. It is at this stage that we meet the man who is acclaimed as the Father of Accountancy — an Italian Franciscan friar, Luca Pacioli — who is credited with the publication in 1494 of a book on algebra and who is also revered as the Father of Double Entry Book-Keeping.

Economic development is the key

The major motivating force in the development of mathematics from the earliest stages seems to have been a desire by man to understand the workings of the world about him and to harness those workings to improve his own well-being. In modern language, the desire would be expressed as a desire to earn more jobs and higher living standards. Modern man would probably claim that the motivating force for scientific discoveries, including mathematical discoveries, is ultimately that of economic development. This can be brought up-to-date in terms of the Programme for the Completion of the EC Internal Market by the end of 1992. One of the primary objectives of the integration of the economies of the 12 Member States of the EC is to strengthen the EC's capacity in science and technology. This is to meet competition in the international marketplace mainly from the USA, Japan and some of the dynamic emerging nations in the Pacific basin.

Science and Technology is an area in which Ireland can fully participate in the completion of the EC Internal Market and in which we can gain a major advantage because of our highly qualified and well educated workforce. It is also an area in which there can be even greater partnership between education and industry. This partnership needs to be founded to a far greater degree on the reality that it is the *business enterprises* in both the private and the public sectors which *transform scientific and technological knowledge and advancement into economic and social progress.* In other words, it is the private and public sector enterprises which provide the mechanism whereby the work in the schools, colleges and universities is ultimately translated into viable long-term jobs and higher living standards for the entire community.

The challenge to Ireland

The National Development Plan 1989–1993 which was submitted by the Irish Government to the EC Commission on 22nd March 1989 and published on 31st March 1989 acknowledges that the Irish GNP per capita measured in current prices and purchasing power standards amounted to only 58% of the EC average in 1987. The problems in Ireland are accentuated by high unemployment and by a demographic structure which results in a high dependancy ratio. Unemployment is currently more than 17% in constrast to the EC average of less than 11%, and Ireland is second after Spain in the high unemployment league.

Ireland's present situation poses an enormous challenge. An increase in living standards and a fall in unemployment will both depend upon economic development. In order to achieve higher living standards and provide more jobs, it is essential to increase the output and added-value of goods and traded services which are produced by private and public enterprises. There is no escaping the reality that Ireland must and can achieve a far higher level of industrial and commercial output. This places an emphasis on competitiveness in terms of the general environment for enterprise, in terms of specific business input costs, in terms of innovation and in terms of management expertise. The main focus must be upon making optimum use of our natural resources with particular reference to our high quality workforce. In this context, the Confederation of Irish Industry in a number of submissions made to the Government as part of the consultative process for the preparation of the National Development Plan 1989–1993 sought significant up-grading in the following areas of public economic infrastructure — transport, energy, education and training, environmental services and telecommunications. The Confederation of Irish Industry also sought infrastructure investment related to three major areas of natural resource development — the food industry, forestry and tourism.

Skills of people

There is a growing emphasis on developing the skills of Irish people in various internationally traded service areas such as finance and consultancy. These are areas in which Ireland does not necessarily suffer from the infrastructure cost disadvantages which attach to the development of heavy manufacturing industry on a broad basis in a small island nation. An emphasis is being placed on computer services, data processing, health services, educational services, and architectural and engineering consultancy as well as on an extremely wide range of financial services activities. The engineering and applied science emphasis is on high technology and high added-value areas which tend towards the arena of skills and software rather than towards bulk materials utilisation. The greater emphasis on high technology and the utilisation of high level skills is already bearing fruit in manufacturing industry which is reporting strong demand for engineers, computer, marketing and financial staff.

Mathematical awareness

Mathematics has a key role to play in the development of a modern high technology economy. It is essential to create an awareness of the value of mathematics not only as an intellectual discipline in itself but also in terms of its contribution to the generation of higher living standards and more jobs. A discussion document issued by the Curriculum and Examinations Board (now the National Council for Curriculum & Assessment) in November 1986, was entitled "Mathematics Education: Primary & Junior Cycle Post-Primary" and this contained a clear set of desiderata for mathematics teaching and a syllabus content which, if realised, would lead to a wonderful increase in the level of mathematical awareness in society. It is vital, in this context of awareness creation, that mathematics be presented in as attractive and stimulating a way as possible to young people in their early formative years.

Securing the educational base

The Confederation of Irish Industry believes that in order to optimise the contribution which mathematics can make to the development of the economy, it is essential to provide a very secure base in terms of mathematical education at the junior cycle level. The Confederation made a submission to the Syllabus

Committee for Mathematics of the Department of the Education in May, 1983. The recommendations contained in that submission are as valid to-day as when they were first made. A repitition of those recommendations here will give a comprehensive view of the attitude of industrialists towards mathematics education during the formative years of junior cycle post-primary.

1. There should be some emphasis on "traditional" mathematics, but the syllabus should also include logical puzzles which would be solved by discovering the underlying logical pattern.

2. There should be a certain amount of drawing to scale and construction of solids. Geometry is important from the viewpoint of graphical literacy which is the cornerstone of many professions in the field of engineering and architecture. In this context, there is a case for the introduction of a basic technical drawing programme as an adjunct to mathematics.

3. There should be an elementary introduction to matrices based on shopping lists and on the prices of normal household commodities. The examples should include addition and multiplication of matrices.

4. There should be more emphasis upon applied problems, and less emphasis upon theorems. This approach should make the subject more interesting and stimulating for young pupils.

5. The course should include library assignments on the history of mathematics with a particular emphasis on the role which mathematics has played in the development of science.

6. Topics which would relate to everyday experiences could include translating situations into mathematical sentences, identifying problems with too little or too much information, conjunction of sentences, translating to conjunctions, solving conjunction of equations, lever or torque problems, motion problems and problems connected with dynamic situations.

7. It is important to introduce the young mathematics student to computing.

8. Mathematics cannot be adequately taught in isolation from other subjects. Mathematics should be related to science, applied science, business

subjects, and to all other aspects of the work of the school which relate in any way to numeracy.

9. The use of a standard textbook and workbook for mathematics should be considered by teachers and the Department of Education. It is essential that some guidance be given to teachers on the pedagogical approach in addition to guidance on syllabus content.

10. The use of audio-visual aids, including film and video, should be encouraged. This will help to relate the subject, mathematics, to the world outside the classroom.

11. Mathematics should cover at least three important areas in which the student will be involved after school:

    - personal and social
    - vocational and work-related, and
    - leisure.

12. In relation to employment, the course should include the use of mathematics in areas such as:

    - bank accounts
    - wage and salary calculations
    - invoices, discounts and VAT
    - income tax
    - services bills, such as electricity
    - interest rates
    - profit and loss
    - the keeping of all types of statistical records

13. In relation to society, the course should cover such topics as:

    - higher purchase and loans
    - energy utilisation in the home
    - insurance and life assurance
    - social welfare benefits

- expenses in relation to home ownership
- everyday household outgoings
- the expenses of owning and running a car
- all other aspects relating to living in a modern industrial society

14. The course should cover the whole range of basic competency areas such as the ability to add, subtract, multiply and divide, and also the use of decimals, graphs, simple formulae, basic statistics and a knowledge of shape, area, volume and other related matters. The basic competency areas should also be related primarily to the needs of the pupil in the three areas: personal & social, vocational and leisure.

15. Some emphasis should be placed upon concepts such as neatness and layout. The presentation of information is often marred by inadequate layout or by the lack of neatness, and the effective communication of data is thereby lost. The concepts of neatness and layout should be emphasised within the mathematics syllabus from the earliest possible stage.

Those were the observations of the Confederation of Irish Industry about the junior cycle syllabus in May 1983. They are relevant not only to junior cycle but also to the teaching of mathematics at any higher level.

Applications oriented Mathematics

The Confederation of Irish Industry while recognising the need for a *balanced curriculum*, and while recognising the need to prepare young people for entry to higher education, must necessarily place some emphasis on the educational preparation of the majority of post-primary pupils who immediately enter the workplace directly from the post-primary school. In this context, the Confederation of Irish Industry has for a number of years financially supported the **Applications Oriented Mathematics Project** which has been developed by the North Tipperary Vocational Education Committee in co-operation with Thomond College of Education in Limerick. Not only has the Confederation given financial assistance to this project but the Confederation has been represented from the initiation of the project on an Advisory Committee which was established by the North Tipperary Vocational Committee to promote

the development of this prototype mathematics course at senior cycle which is intended to meet the direct immediate needs of young people entering employment. This course is now recognised by the Department of Education and the validation process involves people from education and from industry. The Confederation of Irish Industry endorses the aim of the course which is to equip pupils who will directly thereafter enter employment with a sound educational experience which is mathematically significant and which is appropriate to their needs. The specific objectives of this course are worth mentioning:

- provide pupils with a systematic approach, viz., modelling which is comprehensive enough to fit every situation including work situations
- promote mastery of selected mathematical topics including concepts, techniques, know-how and applications
- promote computational facility and use of electronic calculators and associated skills
- build the pupils' confidence in their ability to understand mathematics and to use mathematics
- provide an appropriate industrial/commercial context for the pupils' use of mathematics through practical applications, case studies and industrial visits.

Calculators and Computers

The Confederation of Irish Industry believes that there is a need for some memorisation of basic results but wishes to stress that, in modern circumstances, there is a wider role for the electronic calculator in taking the tedium out of more involved numerical manipulations and this role extends even to the primary school level. The Confederation believes that the calculator can be used creatively to explore relationships between numbers and to experiment with numerical patterns and procedures. At a more advanced stage, the Confederation believes that there is not sufficient stress on the role of the computer in mathematics education integrated with related applied areas. A great number of schools now have computer facilities and many of these have good graphics capabilities. These are ideal for illustrating mathematical notions and results in the form of graphs, charts and other visual displays and are also useful for the implementation of excellent teaching and assessment programmes. The role of the computer even from the every earliest stages of education is vital as computers are now becoming so cheap and so powerful that they are a part of the total home and societal environment of many children.

Mathematics as a communications tool

The two subjects which are essential for any school leaver who enters business and industry are the language which is used in the workplace and the subject, mathematics. These are the tools with which people in industry communicate — those tools are required in order to learn, to understand, to evaluate, to plan, to record and to measure. In this context, it must be stressed that mathematics is a "language" and its great virtue is that, when properly applied, the communication in terms of mathematics is concise, unambiguous, and readily understood internationally. Mathematical symbols in common usage and mathematical symbols in the highest reaches of science are as readily understood in Paris, Washington and Tokyo as they are in Dublin and London. A message in terms of mathematical symbols can frequently be delivered much more concisely and precisely than a similar message given in terms of words.

Mathematics for the employee

Mathematics in terms of wage and salary calculations and in terms of wage negotiations are important to every person in employment. The new entrant to business and industry will find that the modern payslip contains so many items that it sometimes resembles more of a scientific computer printout and this can lead to misunderstanding and to grievances if the content of the payslip is outside of the mathematical experience of the young person. In order to competently analyse a payslip the young person requires a knowledge of basic arithmetic, a familiarity with percentages, cumulative totals, overtime multipliers, bonus calculations, net and gross figures, and other numerical concepts. The conclusion is that tuition in the intricacies of PAYE, PRSI and wage calculations should not be confined to business subjects alone but should be included in the general mathematics syllabus at junior cycle level.

Mathematics for the young worker

Young people who leave the post-primary school may well start production work in batch or process operations and all the stages of the young person's work will be measured in mathematical terms. This means that times will be recorded, weights and volumes will be recorded, there will be records of units produced, percentage unproductive waiting time, bonus ratings and other quantified information. In addition, the young worker would probably obtain instructions from the supervisor which are sometimes expressed in mathematical terms — degrees, weights, speeds, times, volumes and percentages. If the young worker is unable to correctly interpret these instructions, not only will production suffer but there may be safety risks. This latter point means that the young worker must not only be trained to understand the written instructions in the work schedule but must also be motivated to actually read and to act upon those instructions.

Promotion within Industry

Promotion for the young worker within industry from the shopfloor level to supervisory and first line management posts invariably requires that the young person should undertake some element of clerical work. This work includes recording daily/shift outputs, requisitioning materials from stores and recording other information in relation to work of operatives. The young person who is weak at mathematics may well find that his or her promotional prospects are limited on this account. One of the best ways in which to help a pupil who is still at school appreciate applied mathematics in an industrial sense is to involve pupils in group projects within the school. These projects can be part of subjects such as home economics, woodwork, metalwork and building construction. There are many examples of this type of group project run on a "mini company" basis in the post-primary schools.

EC Developments

If Irish industry is to continue to expand then it must continue to modernise and to introduce improved production techniques. There will be a growing emphasis on R & D and upon applied research. Research and development is one of the important factors in the integration of the EC Internal Market.

Special emphasis is being placed on areas such as biotechnology, information technology, mariculture and high technology areas of engineering.

The largest single EC Programme in the area of research and development is ESPRIT (European Strategic Programme for Research & Development in Information Technology) which will be worth more than IR£1 billion in its second phase, 1988 – 1992, and new applications programme will cover the information technology aspects of medicine, transport and education. Ireland is well placed to share in ESPRIT.

The most recent annual report of a second EC Research & Development Programme, SPRINT (Strategic Programme for Innovation & Technology Transfer) noted that the Irish Robotics Project is progressing smoothly. There is another programme called RACE (Advanced Telecommunications) which has recently added 40 additional projects in its second phase.

A wide range of other programmes were adopted by the Research Ministers of the EC at their meeting on 20th June, 1989, and these include areas such as preventative medicine and agri-research.

The Commission has put forward a 7.7bn ECU budget for the Third Research Framework Programme which covers the period to 1994. The French want Ministers to agree to this ambitious plan by the end of the current year. The proposed budget is likely to be challenged by some Member States. Unanimity is needed before the plan can be adopted. Resources will focus on six specific programmes whose proposed budget (in million ECUs) is as follows:

		ECUs(m)
1.	Information and communications technologies	3,000
2.	Industrial and materials technologies	1,200
3.	Environment	700
4.	Life sciences and technologies	1,000
5.	Energy	1,100
6.	Human capital and mobility	700

Existing and potential Irish beneficiaries should become acquainted with the significant new orientation of the EC's research strategy.

The ECLAIR research programme which relates to agriculture has a budget of 80 million ECUs for the five years 1989 to 1993. The EC Commission is promoting six lines of action designed to develop an EC policy for space research. The EC Commission recently announced details of new research and development programmes in the field of marine science and technology.

Relevant areas of research and development being promoted by the EC Commission include:

- scientific research
- technological development
- diffusion of research results
- strategic economic analysis
- databases
- technical standards
- Higher Education

Higher education has an extremely important role to play in order to ensure that Ireland can participate to the full in the EC programmes. On a wider basis, it is obvious that an increasing number of production and administrative staff will be required to master the use of "microchip" projects which will depend upon digital and numerical information for their operation. The Confederation of Irish Industry is acutely aware that investment in education and in training in Ireland is needed in order to enable Irish industry to offset the disadvantages of being an island nation on the periphery of Europe with a low per capita income.

The Confederation is recommending to the Government in a Pre-Budget Submission 1990 that public expenditure on education should at least be maintained at current levels in real terms. The Confederation is recommending that any funds which would otherwise be saved because of the anticipated phased drop in primary school enrolements by up to 20% should be redeployed to increase the Irish participation rate in third level education to 50% of the relevant age cohort compared with about 25% at present. This increase should focus particularly on qualifications in greatest demand for economic development purposes, such as engineering, computer science and business studies.

The Confederation in ranking the priorities for investment in Ireland with aid from the EC Structural Funds placed special emphasis on investment in higher education which should include post-graduate work and research and development. The Confederation emphasised links between industry and higher education not only in terms of the direct educational input but also in areas such as industrial parks, innovation centres, research parks and science parks. The Confederation placed a focus not only on technological education but also on business education and on all faculties in the higher education institutions on the grounds that all faculties contribute to intellectual development.

Three Major Points

There are three major points which need to be emphasised from a business viewpoint in relation to mathematics.

1. It is essential to inculcate a positive attitude towards mathematics in all school leavers and, in order to do this, it is essential that parents, teachers and employers must have a positive attitude towards mathematics. Points which need to be emphasised in this context include:

    - basic mathematics is essential as a skill which, just like riding a bicycle, can only be developed through individual effort and through practice.
    - Mathematics is not just a subject for the "boys"; if girls are to avail of the increased range of careers now opening up to them, it is essential that they be able to compete on an equal footing with boys in the area of mathematics. This implies that there must be a new approach particularly to the availability of higher level mathematics at senior cycle in "traditional" single sex girls schools.
    - Mathematics is just as important in the private life of the individual as it is in the workplace. Calculations in relation to the monthly mortgage payment or in relation to the economics of using a motor car are acquiring a new importance in the life of the individual.

2. The presentation of mathematics in the schools should, in so far as possible, be related to the applied subjects which are studied by students. Problems based upon examples from home economics, woodwork, metalwork, chemistry, physics, business organisation, accountancy and other applied subjects should be introduced as an integral part into the mathematics syllabus. An increased emphasis should be placed upon graphical presentation of information in recognition of the fact that there is a growing emphasis upon chart and diagram format of presentation of information. This is seen in the print and television news media where

there is a growing use of flow charts, histograms, and pie-charts which are intended to enable the reader or the audience to rapidly assimilate statistical data. Mathematical applications are also seen in the growing sophistication of computer games and toys.

3. Teachers should be encouraged to maintain and to strengthen links with business and industry. This type of contact is important not only for the guidance counsellor but should extend for the benefit of all teachers, including teachers of mathematics. In this context, the Confederation of Irish Industry established an education trust in December 1983, to financially assist research and development work undertaken by teachers' subject associations. Financial assistance has, to-date, been given for projects in chemistry, physics, junior science, biology, mathematics, geography, history, computer studies, art and design, business subjects and modern languages.

The business community fully recognises the dynamic role which education plays in the development of a modern industrial economy and recognises the centrality of mathematics for this purpose within the education process. The important role which mathematics plays in the curriculum needs to be continuously reinforced by keeping syllabuses up-to-date, by researching "state of the art" practices and procedures and by the implementation of a planned programme of in-service training for teachers with appropriate mechanisms to link the teaching of mathematics to its application in business and in the wider general community. Progress does not stand still in the workplace or in economic development; it cannot be allowed to stand still in the world of the mathematics and in the classroom of the mathematics teacher and of the mathematics student.

Confederation of Irish Industry
Confederation House
Kildare Street, Dublin 2.

NOTES

Venn Diagrams: A Combinatorial Comment

T. B. M. McMaster

When, in tender years, we all first learned how to draw Venn diagrams, those in charge of our education insisted that these be depicted as in Fig. 1; and if, through inquisitiveness, amnesia or sheer cussedness, we produced a deviant hieroglyph such as those in Fig. 2, they generally informed us that
(i) We were silly, and
(ii) even though some examples could be devised which fitted into our 'wrong' diagram, the vast majority of instances could only be accomodated on the 'general case' picture which we had been told to emulate.

Now a detailed analysis of proposition (i) may not perhaps be appropriate at this juncture, but assertions such as (ii) have a habit of surfacing in the mind after lying dormant for years. So it has come to pass that several members of our Department have recently been exploring some of the combinatorial/probabilistic questions which are raised by subjecting it to scrutiny and generalization. This brief note presents a report on one of these investigations.

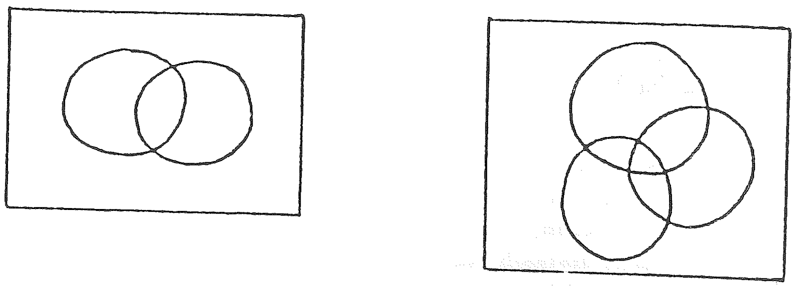


Figure 1: 'right' Venn diagrams

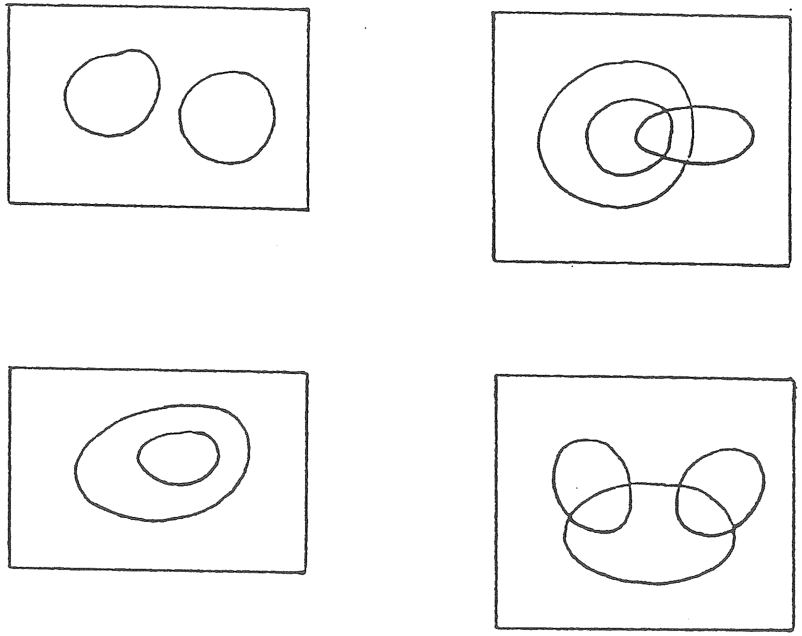


Figure 2: 'wrong' Venn diagrams

The problem to which we here address ourselves is this: given a finite set X (with n elements say), a (small) positive integer s, and a random selection of s distinct subsets of X, how likely is it that the 'general case' Venn diagram is the only correct one to describe their relationships? As the number of such random selections is easily obtained, namely

$$T(s) = 2_n(2_n - 1)(2_n - 2)\ldots(2_n - s + 1)/s!$$

the equivalent combinatorial problem is: how many selections (of s subsets) does the general case diagram alone depict?

Since the feature which distinguishes the general case here from the various degenerate ones is non-emptiness of the disjoint 'regions' on the diagram (except perhaps for the 'outer zone'), it is convenient to begin with a formula for the number of ways of choosing a specified number of pairwise disjoint non-empty subsets of X (not necessarily covering the whole of X).

Lemma *Consider a positive integer $k \leq n$. The number $\Omega(k)$ of ordered choices of k pairwise disjoint non-empty subsets of X is given by*

$$\Omega(k) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (k+1-j)^n ,$$

and the number of unordered choices is $\Omega(k)/k!$

Proof The formula claimed for $\Omega(1)$ gives $2^n - 1$, which is evidently correct. Assuming now its validity for integers from 1 to $k-1$, observe that there are $(k+1)^n$ ways of distributing the elements of X across k boxes and one wastepaper basket, and that we shall determine $\Omega(k)$ by subtracting from this total the number of distributions in which one, two, three, ...,k of the boxes remain empty. This gives

$$\Omega(k) = (k+1)^n - \binom{k}{1}\Omega(k-1) - \binom{k}{2}\Omega(k-2) - \ldots \binom{k}{k-1}\Omega(1) - 1 ,$$

and when we substitute in the assumed formulae for $\Omega(k-1)$, $\Omega(k-2)$, ...,$\Omega(1)$, the coefficient of $(k+1-j)^n$ in the resulting expansion is

$$(-1)^j \left\{ \binom{k}{1}\binom{k-1}{j-1} - \binom{k}{2}\binom{k-2}{j-2} + \ldots + (-1)^{j+1}\binom{k}{j}\binom{k-j}{0} \right\}$$

which is easily evaluated as $(-1)^j \binom{k}{j}$. Induction completes the demonstration.

Let us now return to the simplest case ($s = 2$) of the original problem. There are $T(2) = 2^{n-1}(2^n - 1)$ ways of selecting an (unordered) pair A, B of subsets of X, and this selection will be non-degenerate in the Venn diagram sense if and only if none of the three sets $A \cap B$, $A \cap B'$, $A' \cap B$ is empty. Now there are $\Omega(3)/3!$ ways of selecting three non-empty disjoint subsets (call them K, L, M) of X, but each such selection resolves itself into three distinct choices of $\{A, B\}$ when we try to identify K, L and M with $A \cap B$, $A \cap B'$ and $A' \cap B$; for A and B could be

either	$K \cup L$	and	$K \cup M$,
or	$K \cup L$	and	$L \cup M$,
either	$K \cup M$	and	$L \cup M$;

hence we see that:

Proposition 1 *The number of non-degenerate (in the present sense) choices of two subsets of X is*

$$\frac{\Omega(3)}{3!} = \frac{1}{2}(4^n - 3.3^n + 3.2^n - 1) .$$

In the same way, there are $T(3)$ ways of choosing three subsets A, B, C, of X, and the choice is non-degenerate precisely when none of the seven sets $A \cap B \cap C$, $A' \cap B \cap C$, $A \cap B' \cap C$, $A \cap B \cap C'$, $A \cap B' \cap C'$, $A' \cap B \cap C'$, $A' \cap B' \cap C$ is empty. There are $\Omega(7)/7!$ choices of seven disjoint non-empty subsets (call them K, L, M, N, O, P, Q) of X, each resolving itself into several distinct choices of $\{A, B, C\}$. To be precise, for each of the 7! permutations of $K, L, \ldots, Q$ we could identify those sets in order with $A \cap B \cap C$, $A' \cap B \cap C$, $\ldots A' \cap B' \cap C$, thus constructing an ordered triple (A, B, C); it will be necessary to divide by 3! to disregard the order and so $\{K, L, \ldots, Q\}$ actually resolves itself into 7!/3! unordered combinations of $\{A, B, C\}$. Thus we have shown that:

Proposition 2 *The number of non-degenerate choices of three subsets of X is*

$$\frac{\Omega(7)}{3!} = \frac{1}{6}(8^n - 7.7^n + 21.6^n - 35.5^n + 35.4^n - 21.3^n + 7.2^n - 1)$$

Although Venn diagrams themselves cease to be of much use for $s > 3$, the above analysis requires no significant change to cope with larger values. Thus one reaches the following conclusion:

Theorem *Let n and s be positive integers, and let X be a set with n elements. The number of ways of choosing s distinct subsets $A_1, A_2, \ldots, A_s$ of X (irrespective of order), subject to the condition that every one of the sets*

$$C_1 \cap C_2 \cap \ldots \cap C_s$$

(where for $1 \leq i \leq s$, C_i is either A_i or A_i', but $C_i = A_i$ for at least one value of i) is non-empty, is

$$\frac{1}{s!} \sum_{j=0}^{2^s-1} (-1)^j \binom{2^s-1}{j} (2^s - j)^n$$

The following table records, for $1 \leq n \leq 10$, the calculated values of $T(2)$ and $T(3)$, of $\Omega(3)/2!$ and $\Omega(7)/3!$, and of the probabilities p_2 and p_3 that a randomly chosen pair or trio of sets is non-degenerate, recorded to four decimal places. It shows as expected that the probabilities, though small for small values of n, rise as n does. Better information on their behaviour for large n is easy to extract from the above formulae, which yield that

$$\begin{aligned} p_2 &= 1 - 3(3/4)^n(1 + o(1)) \\ p_3 &= 1 - 7(7/8)^n(1 + o(1)) \end{aligned}$$

and in general, where p_s is defined as $\Omega(2^s - 1)/s!T(s)$, that

$$p_s = 1 - (2^s - 1)(1 - 2^{-s})^n(1 + o(1))$$

as $n \to \infty$. Since these probabilities tend to 1, we are obliged to concede that they told us the truth all those years ago. Rather a pity, really.

n	$T(2)$	$\Omega(3)/2!$	p_2	$T(3)$	$\Omega(7)/3!$	p_3
1	1	0	0	0	0	0
2	6	0	0	4	0	0
3	28	3	.1071	56	0	0
4	120	30	.2500	560	0	0
5	496	195	.3931	4,960	0	0
6	2,016	1,050	.5208	41,664	0	0
7	8,128	5,103	.6278	341,376	840	.0025
8	32,640	23,310	.7142	2,763,520	30,240	.0109
9	130,816	102,315	.7821	22,238,720	630,000	.0283
10	523,776	437,250	.8348	178,433,024	9,979,200	.0559

Department of Pure Mathematics
The Queen's University of Belfast
Belfast BT7 1NN, Northern Ireland

HISTORY OF MATHEMATICS

Giovanni Frattini 1852–1925

Maurizio Emaldi

(Communicated by M.L. Newell)

A little over one hundred years ago, between 1885 and 1886 three papers by the Roman mathematician, Giovanni Frattini *"On the generators of groups of operations"* appeared in the proceedings of the Royal Academy of Lincei. In the first of these the author introduced the subgroup Φ of a finite group of operations consisting of the set of all operations which "cannot effectively contribute to the generators" of the group. This can be characterized as the intersection of all proper maximal subgroups. He demonstrated that the group in question is nilpotent and in doing so used a most elegant argument which today is called "the Frattini argument". The results contained in these three papers, the full scope of which were not fully grasped at the time of their publication, are amongst the most significant contributions of Italian mathematicians to the theory of groups in the latter half of last century. The definition of the subgroup Φ of a finite group given by Frattini has been extended to groups in general and today is generally called "the Frattini-subgroup". (As far as we can determine, this name appeared explicitly for the first time in a paper by G. Zacher: "Construction of finite groups with trivial Frattini-subgroup." Rend. Sem. Mat. Padova, vol. 21, 1952). In group theory the Frattini-subgroup and more generally the analogous notion in algebraic structures play a central role in many questions. Thus it seems opportune to give a brief biography of the author and document his mathematical interests. While our investigations have led to a complete list of his publications, we shall give but a selection here. We have used the writings of R. Marcolongo "Bollettino di Matematica (1926)", of P. Teofilato "Memorie della Pontificia Accademia dei Nuovi Lincei (1926)," G. Zappa "Supplemento ai Rendiconti del Circolo Matematico di Palermo (1985)" and research on Frattini undertaken by O. Vanuzzo in 1982/3. Furthermore, we have availed of the help of other people, notably A.C. Garibaldi of Genoa.

Biographical Notes

Giovanni Frattini was born in Rome on the 8th of January, 1852 and in that city he did all of his studies. He was admitted in November 1869 to the first year course in Mathematics and in June 1870, passed his Bachelor's examination with honours. His tutors at University were Battaglini, Beltrami and Cremona and in July 1875, he was awarded his degree in Mathematics. In 1876, he was in charge of Mathematics in the Liceo di Caltanisetta and from there, in November 1878, was transferred to the Technical Institute of Viterbo. In 1879 he became principal teacher of Mathematics and Descriptive Geometry. In February 1881 he obtained a transfer to the Technical Institute of Rome. He remained in Rome until the end of April 1916, having opted for the position of lecturer in the Military College where he taught right from its foundation in 1884. On the 1st of August 1921, he retired.

He was one of the most senior members of the "Mathesis", an association founded in Rome in 1895 by secondary school teachers of Mathematics aimed at improving the schools and raising teaching standards. Between 1900 and 1902, he held the office of President and although his term was extended for a further two years, he did not wish to accept in order to secure the re-election of the association's first President R. Bettazzi. In 1914, he was awarded lectureship qualification in Algebra at the University of Rome, but never took up an appointment. He was a member of the Mathematical Society of Palermo. In 1917, he became an associate member of the Pontificia Romana Accademia dei Nouvi Lincei and an ordinary member in 1918.

He held office as city counsellor and in this guise, helped in the erection of a monument to his favourite poet G.G. Belli, whose sonnets he was wont to recite to his pupils in perfect Roman dialect. This practice of his led to a ministerial enquiry which ended happily. He declared himself willing to replace these sonnets by the poetry of a highest undersecretary of Education. Without further ado, he was allowed continue his recitation of Belli's poetry. A gifted teacher, he wrote several books for the elementary schools in Rome, for the two year course at the Technical Institute and for the Military College. These concise books, bereft of superflous abstraction and rich in elegant brilliance makes one aware of Frattini's art of teaching. The spirit of the master is evident from the following verse which heads a collection of exercises in Frattini's book "Elementi di Calcolo Letterale" (Paravia, 1908).

— Gino mio, l'ingegno umano
partorì cose stupende,
quando l'uomo ebbe tra mano
meno libri e piu faccende — (Giusti)

implying that greater things are achieved by doing more and reading less.

As a researcher and scientist, Frattini cultivated all the mathematical disciplines and studied with particular benefit differential geometry, the theory of groups and the analysis of second degree indeterminates. On this latter topic, his noteworthy contributions to the Periodico di Matematica 1891/92 simplify significantly the classical methods of Euler, Lagrange and Gauss. Two large tracts in 1883 and 1884 on groups of transitive substitutions which were the fruit of his study of the classical work of Jordan and some brilliant works of Capelli, made him eligible for appointment to the Chair of Complementary Algebra in Naples in 1886. Amongst the contestants in this truly famous competition were Capelli, Cesàro and Besso.

Subsequently, however, for family reasons, and because he did not wish to leave Rome, he did not compete for other University posts and remained in secondary teaching, making a magnificent contribution "turning the advances of science to the benefit of the school". This motto of Frattini's was imprinted on the journal "*Il Bollettino di Matematica*" founded by A. Conti at Bologna in 1902.

The last years of his life were not happy ones. He was deeply troubled by the war and further by the illness of one of his sons, a war invalid, the loss of his wife and the difficult economic conditions which forced him to work even though he had left teaching for some years. All of this left him dejected but he still managed to retain his old habit of frequenting his favourite Roman cafe where as he often pronounced he "held his chair". He died on the 21st July, 1925, while acting as supervisor of examinations for the School Leaving Certificate of the Liceo Scientifico di Roma.

Apart from 21 papers in the "Periodico di Matematica", 6 articles in "Il Bollettino di Matematica", 3 contributions to "La Matematica Elementare" and over 13 general publications on mathematical topics, some written in witty poetic form – his main works are the following:

Memorie R. Accademia dei Lincei

1883 I gruppi transitivi di sostituzioni dell'istesso ordine e grado. Serie III, Vol. XIV, pagg. 143–172.

1884 Intorno ad alcune proposizioni della teoria delle sostituzioni. Serie III, Vol. XVIII, pagg. 487–513.

Rendiconti R. Academia dei Lincei

1883 I. Gruppi a K dimensioni. Serie III, Vol. VIII, pagg. 260–264.

1885 Intorno a un teorema di Langrange. Serie IV, Vol. pagg. 136–142.

1885 Un teorema relativo al gruppo della transformazione modulare di grado p. Nota I, Nota II, Serie IV, Vol. I, pagg. 142–147, 166–168.

1885 Intorno alla generazione dei gruppi di operazioni. Serie IV, Vol. I, pagg. 281–285.

1885 Intorno alla generazione dei gruppi di operazioni. Nota II. Serie IV, Vol. I, pagg. 455–456.

1886 Intorno alla generazione dei gruppi d'operazioni e ad un teorema d'Aritmetica. Serie IV, Vol. II, pagg. 16–19.

1886 Extensione ed inversione d'un teorema d'Aritmetica. Serie IV, Vol. II, pagg. 132–135.

1892 Due proposizioni della teoria dei numeri e loro interpretazione geometrica. Serie IV, Vol. I, pagg. 51–57.

1892 A complemento di alcuni teoremi del sig. TCHEBICHEFF. Serie V, Vol. I, pagg. 85–91.

1893 Di un doppio isomorfismo nella teoria generale delle sostituzioni. Serie V, Vol. II, pagg. 253–259.

1903 Di un gruppo continuo di trasformazioni decomponibili finitamente. Serie V, Vol. XII, pagg. 74–82.

Atti Della Pontificia Accademia Romana dei Nouvi Lincei

1917 Di una dualità reciproca tra coppie di quadrilateri inscritti nel medesimo cerchio. Sessione VII, pagg. 136–139.

1918 Intorno a una questione di minimo relativa alle equazioni. Sessione VI, pagg. 210–219.

1920 Alcune considerazioni poligono vettorali. Sessione I, pag. 25.

1922 Una formula di approssimazione relativa al gruppo della radice quadrata. Sessione IV, pagg. 99–104.

1923 Un problema di ampliamento per isomorfismo cui dà luogo la teoria della relatività. Sessione III, pagg. 94–99, Sessione V, pagg. 146–156.

1924 La relatività e le frazioni continue. Sessione II, pagg. 80–81.

1924 Intorno alla proprietà caratteristica dei numeri primi. Sessione III, pagg. 103–105.

1924 Intorno a una proprietà caratteristica delle funzioni intere d'una variabile. Sessione VII, pagg. 174–178.

Dipartimento di Matematica Pura e Applicata
Università di Padova

BOOK REVIEWS

STATISTICAL INFERENCE FOR SPATIAL PROCESSES
by Brian D.Ripley. Cambridge University Press,1988,viii + 148pp. $34.50

In their quest to isolate signal from noise, statisticians know that a good statistical procedure depends not only on the data but also on assumptions they make to account for information they have concerning the mechanism generating the data. Depending on these assumptions, they then adopt a purely data- analytic angle, or attack the problem from a classical, Bayesian, robust, non-parametric or other inferential approach. The strength of the results obtained will depend on the validity of the approach used and on the criterion of performance adopted. Because of special peculiarities associated with spatial processes, the above standard scientific process leads to severe difficulties in the analysis of spatial data. This scholarly essay by Ripley is an excellent attempt to describe and accomodate these difficulties.

Chapter 1 explains a number of problems that arise in statistical inference for spatial processes, including the major difficulties caused by edge effects and long-range dependence. The inappropriateness of time series models for spatial processes is highlighted, along with certain inadequacies of procedures based on the likelihood function of the data. These latter inadequacies are explored in more detail in Chapter 2 where the author explores an autoregressive Gaussian process as model.

Moving away from Gaussian processes, Chapter 3 examines spatial point processes employing a geometric approach. The author examines the behaviour of certain statistics that are based on distances between points and which accomodate edge effects. The asymptotic results presented are useful for detecting departures of the model from the simple binomial or Poisson process, but their complexity and corresponding incompleteness force a comparison of various estimates only in terms of their first two moments. Continuing in a similar vein, Chapter 4 deals with Gibbsian models, which, as the author notes, have their origin in statistical physics and account for interaction between points. Here estimation procedures are approximate and an

interesting feature is the incorporation of Monte-Carlo methods for portions of the estimation.

Chapters 5 and 6 involve the highly important study of digital images. Chapter 5 accomodates background knowledge of the signal in the form of a prior distribution, and hence invoves Bayesian inference procedures.On the other hand, Chapter 6 belongs more to the area of (exploratory) data analysis, the methods discussed being useful from a descriptive viewpoint, in addition to permitting one to suggest a suitable model and allow appraisal of the procedures of the previous chapter.

Because of the inadequacy of time series models and the limitations of standard likelihood methods, the reader may feel despondent upon reading the first two chapters. This however should not be the case, because the author clearly demonstrates that the problems are not insurmountable. True, the theory is somewhat inconclusive, but this reflects only on the difficulties involved, and there is no doubt that the author has achieved his objectives of describing the state-of-the-art and offering much food for thought. It is this reviewer's opinion that while asymptotic theories have in recent years been developed by probabilists for dependence structures as severe as those implied in 'strong mixing sequences', the state of probability theory is not yet at a level appropriate for deriving a totally definitive theory for spatial processes. For this and other reasons, the author and others in his area deserve high praise for their efforts. From a purely statistical viewpoint, I note that while asymptotic procedures are at the core of much of statistical inference theory, spatial processes beg the question: 'asymptotic in what?' As pointed out by the author, asymptotic results differ according to whether we fix the region E (within which observations will be taken) and let the sample size increase or, as essentially done by Mardia and Marshall(1984), let E expand.

This timely essay is a must for specialists in the area of spatial processes who desire to keep abreast of recent developments in the area. Theoretical statisticians too will welcome this well-written addition to the literature, for it abounds in open research problems. For example, there is plenty of room for the development of procedures that are robust against, e.g., misspecification of the functional form of certain models and against forms of dependence among the observations other than those forms accomodated in this essay. (Needless to say, the mathematics will be enormously involved in such projects.) A valuable graduate course for students could be based on this book along with Ripley(1981). One caveat to the reader: the writing style is lucid but, as befits an essay, highly terse, and the author wastes no time in reviewing background probability or statistics. For such a preparation, it is not necessary that the reader be familiar with theory and techniques at the level of Billingsley(1968) and Serfling(1980), but be aware that the author liberally sprinkles measure-theoretic concepts and non-elementary limiting techniques throughout the essay! An understanding of the ideas in Ripley(1981) is also highly desirable for appreciating the many elegant ideas in this outstanding essay.

References

[1] Billingsley, P. (1968) *Convergence of Probability Measures.* New York: John Wiley.

[2] Mardia, K.V.and Marshall, R.J.(1984). *Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression.* Biometrika 71,135-46.

[3] Ripley, B.D.(1981). *Spatial Statistics.* New York: John Wiley.

[4] Serfling, R.J.(1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley.

Jerome N.Sheahan
Department of Statistics
University of Alberta

STATISTICAL MECHANICS: RIGOROUS RESULTS
by David Ruelle, Addison-Wesley (1969,1989)

Addison-Wesley have reissued Ruelle's famous book as part of a new series *Advanced Book Classics.* This book is a landmark in modern statistical mechanics. The basic concept of the book is the use of functional analysis as a foundation for statistical mechanics, and this idea is behind much study in the past three decades. Not only have the techniques of functional analysis provided insight upon physical problems, but standard methods of statistical

mechanics form the underlying basis for the *theory of large deviations*, which is of considerable current interest to researchers in probability theory.

During the twenty years since this book was first issued, the problems it approaches have become clarified, and some have been solved, but most remain as an open challenge. Mathematicians with an interest in functional analysis may wish to have a go. This book is a good place to start, but they will not find the going easy: Ruelle packs a lot into 200 pages.

The most interesting aspects of statistical physics involve phase transitions. For the standard models, phase transition does not occur with finite systems, so one must start with infinite systems which are the limits of finite approximations. One such is a Newtonian system of infinitely many point particles, but the simpler model of an infinite lattice where each lattice point must be in one of a finite number (usually 2) of states is also studied. The continuous and lattice systems can be considered as classical or quantum. Thus there are several stages of increasing difficulty, from the classical lattice to the quantum continuous systems.

The first step is to deal with the limit of finite systems in such a way that for energy considerations the *boundary* of the finite approximation can be ignored. The assumptions on the strength of the interactions are those needed to make the limiting process work. It has since been discovered the a slightly more restricted family of interactions has much nicer properties concerning phase transitions. For continuous systems a rather special class of interactions is considered. The important case of the Coulomb interaction is not treated in this book.

One area in which reasonably satisfactory results obtain is that in which the interactions of the system are sufficiently weak. In this case one can prove that the behavior is quite close to that of non-interacting systems. For slightly stronger interactions, even in the classical lattice model, one has the presence of several phases in the sense that the infinite limit with different boundary conditions yields different states.

The case of stronger interactions is more interesting and more difficult. Limited progress has been made in this case. The book also deals with various probabilistic, group theoretic, algebraic and functional algebraic methods of treating statistical mechanical systems.

Wayne Sullivan
Department of Mathematics
University College Dublin.

AN INTRODUCTION TO ALGEBRAIC TOPOLOGY
by Joseph J.Rotman, Graduate Texts in Mathematics 119, Springer-Verlag, 1988, 433pp., ISBN 0-387-96678-1

This is a well written, often chatty, introduction to algebraic topology which "goes beyond the definition of the Klein bottle, and yet is not a personal communication to J.H.C. Whitehead." Having read this book, a student would be well able to use J.F.Adam's "Algebraic Topology:A Student's Guide" to find direction for further study. The book begins with a sketch proof of the Brouwer fixed point theorem: if $f : D^n \to D^n$ is continuous, then there is an $x \in D^n$ such that $f(x) = x$. Functorial properties of homology groups imply that the sphere S^n is not a retract of the disc D^{n+1}, and then a simple argument by contradiction shows that f must have a fixed point. This illustrates the basic idea of studying topological spaces by assigning algebraic entities to them in a functorial way. There follows a rigorous account of the singular homology of a space which assumes only a modest knowledge of point-set topology and a familiarity with groups and rings. The account includes the Hurewicz map from the fundamental group to the first homology group, and ends with a proof of the Mayer-Vietoris sequence. By page 110 a complete proof of Brower's theorem has been given.

Singular homology is good for obtaining theoretical results, but not so good for computations. So simplicial homology is introduced in Chapter 7, and used to compute the homology groups of some simple spaces such as the torus and the real projective plane. A proof of the Seifert-Van Kampen theorem for polyhedra is given at the end of the chapter. Continuing the search for effective means of computing homology groups, Chapter 8 introduces CW complexes and their cellular homology. Chapter 9 begins with a statement (without proof) of the axiomatic characterisation of homology theories due to Eilenberg and Steenrod, and then introduces enough homological algebra to prove the Eilenberg-Zilber theorem and Künneth formula for the homology of a product of spaces. Chapter 10 deals with covering spaces. The higher homotopy groups are studied in Chapter 11 using the suspension and loop functors. Results obtained include the exact homotopy sequence of a fibration, and its application to the fibration $S^3 \to S^2$ to show that the group $\pi_3(S^2)$ is non-trivial. The isomorphism $\pi_3(S^2) \equiv \mathbb{Z}$ is beyond the scope of the book. In the final chapter a short discussion on de Rahm cohomology is used to motivate the study of the cohomology ring of a space.

The book is nicely structured, with explanations of where the theory is heading given at frequent intervals. Important definitions are often accompanied by a discussion on their origins. Many exercises are given at the end of sections. Proofs are usually given in full detail. Even though probably every result in the book (and many more besides) can be found in E.H. Spanier's classic text "Algebraic Topology", J.J.Rotman's style of exposition makes the book a useful reference. However a lecture course based on this book may turn out to be a bit slow and dry. (Unfortunately the book corresponds to the syllabus of a one year course given at the University of Illinois, Urbana.) For example the homology of a space is defined on page 66 but we have to wait until page 157 until the homology of the torus is calculated, and until page 226 for the homology of a lens space. The fundamental group is introduced on page 44 but is not calculated for a wedge of two circles until page 171. Maybe too much rigour and generality in a first course on any topic is not a good thing!

Graham Ellis
Department of Mathematics
University College Galway.

PROBLEM PAGE

Editor: Phil Rippon

Just two new problems this time. The first is an inequality which is elementary, in the sense that it can be solved easily by rearrangement to give a perfect square. However, there is a much more interesting solution using an approach which makes the inequality 'obvious'!

23.1 *Find a context within which the inequality*

$$\frac{(a+b)(c+d)}{a+b+c+d} \geq \frac{ac}{a+c} + \frac{bd}{b+d}, \quad a,b,c,d > 0,$$

is intuitively obvious.

I heard the next problem from several tutors at an O.U. Summer School; apparently, it has became a popular 'investigation' at teacher training colleges. Roughly speaking, the problem is to find the number of triangles which have integer sides and perimeter n.

23.2 *Let $s(n)$ denote the number of triples (a,b,c), where a,b,c are positive integers with*

$$a \leq b \leq c \text{ and } a+b > c.$$

Determine a simple formula for $s(n)$.

Now here are the solutions to two earlier problems, which I described as geometric doodling.

21.1 *What is the minimum number of (strictly) acute angled triangles into which a square can be partitioned?*

The solution is given in Fig. 1, which contains three construction lines (one straight line and two semi-circles).

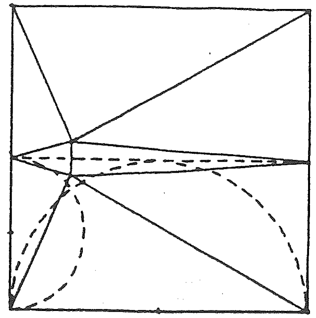
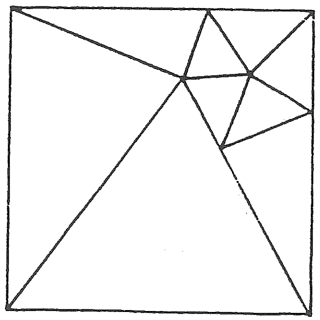


Fig. 1

This problem appears in H.S.M. Coxeter's book 'Introduction to Geometry' and I think it was discussed by Martin Gardiner in Scientific American in the sixties. It must surely be impossible to find a solution with fewer than 8 triangles, but I don't know a proof. It is possible to solve the problem with any $n \geq 8$, however, as Fig. 2 shows (it is of course easy to go from n to $n+3$).



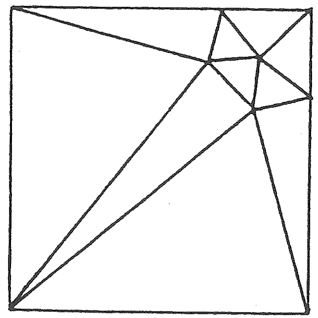


Fig. 2

21.2 *Find a configuration of finitely many points in the plane such that the perpendicular bisector of the line segment joining each pair of points passes through at least two of the points.*

The solution is to take the points to be the vertices of four equilateral triangles built on the sides of a square. It takes only a little work to check that this configuration has the required properties.

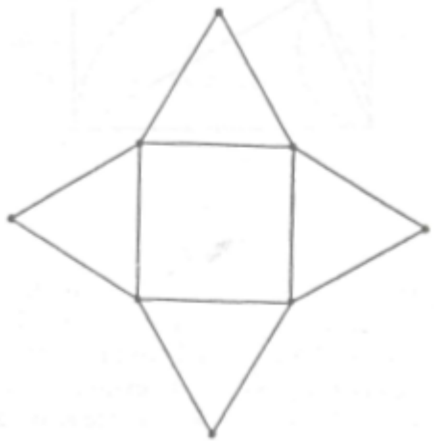


Fig. 3

Thanks to H.S.M. Coxeter and W. Moser, I have been in touch with L.M. Kelly of Michigan State University who considered this problem back in 1964 when he was visiting the University of Cambridge. He believes that the problem may have originated with Paul Erdős and thinks that it is still unknown whether this is the unique such configuration. The nearest I have come to finding another solution was to replace the four equilateral triangles in Fig. 3 by their reflections in the corresponding sides of the square. I'll let you discover why my momentary excitement was quickly followed by disappointment!

Phil Rippon
Faculty of Mathematics
The Open University
Milton Keynes MK7 6AA, UK

INSTRUCTIONS TO AUTHORS

Authors may submit articles to the Bulletin either as TeX input files, or as typewritten manuscripts. Handwritten manuscripts are not acceptable.

Manuscripts prepared with TeX

The Bulletin is typeset with TeX, and authors who have access to TeX are encouraged to submit articles in the form of TeX input files. Plain TeX, AMS-TeX and LaTeX are equally acceptable. The TeX file should be accompanied by any non-standard style files which have been used.

The input files can be transmitted to the Editor either on an IBM or Macintosh diskette, or by electronic mail to the following Bitnet (EARN) address:

MATRYAN@CS8700.UCG.IE

Two printed copies of the article should also be sent to the Editor.

Authors who prepare their articles with word processors other than TeX can expedite the typesetting of their articles by submitting the input file in the same way, along with the printed copies of the article. The file should be sent as an ASCII file.

Typed Manuscripts

Typed manuscripts should be double-spaced, with wide margins, on numbered pages. Commencement of paragraphs should be clearly indicated. Handwritten symbols should be clear and unambiguous. Illustrations should be carefully prepared on separate sheets in black ink. Two copies of each illustration should be submitted: one with lettering added, and the other without lettering. Two copies of the manuscript should be sent to the Editor.