Computing as a Sport

A life doing science with computers

Roy Williams

This is the best of me;
for the rest, I ate, and drank, and slept, loved and hated, like another ...
this, if anything of mine, is worth your memory
– Ruskin

for Jane, Jim, Jess, Jane, & Charles

Contents

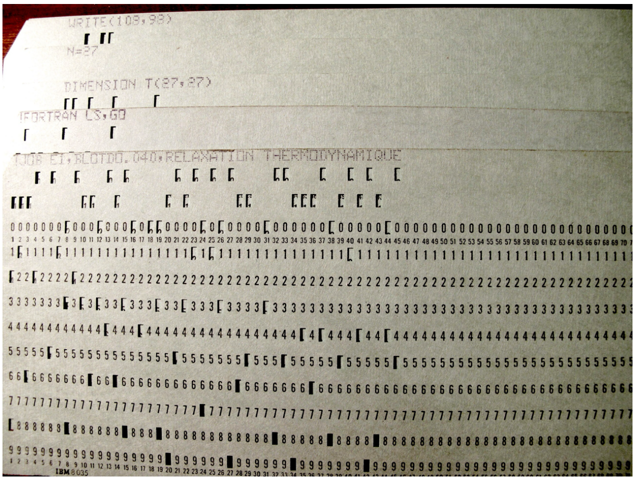
1. Fortran

Computing as a Sport

It was a nasty afternoon in South London in 1972, cold and drizzly, with darkness coming early in the midwinter. A soccer game on the so-called "slopy pitch", where it is best to play the ball downhill, for obvious reasons. As luck would have it, my team was supposed to play the ball uphill, but I was just standing there wondering if maybe we could go in early if the rain came harder. Wondering about how a star can make X-rays (answer: black hole). I was too late to take evasive action, and a wet, muddy soccer ball bashed the side of my head and a cry of ***"PASS!! PASS THE BALL!!"***. I also remember afternoons of cross-country running, slogging and jogging through cold, slippery mud. Some of my school-fellows seemed happy with the exercise, as a horse would be, but not me. Either this or marching with the cadet force, but all, except swimming, out in the rain. These were the Tuesday and Thursday Games Afternoons at my school.

One day a couple of friends showed me their "computer program", a few lines of FORTRAN written with a pencil. And this does this, and that does that, that's what this program does. And just about this time word came down that a new option would be available on Games Afternoons, that would involve Computing. Well obviously I was all in, anything to get out of the mud. It turned out that volunteers were needed to carry boxes of computer cards to Imperial College in central London, which seemed to me infinitely preferable to either soccer or cross-country running. Changing out of school uniform, taking the bus and the Underground, walking the tunnel at South Kensington, mixing the undergraduates as if I were one myself.

Another aspect of computing at Alleyn's School was the acquisition of an IBM 29 card punch, looking like something from Star Trek, with its blue, grey and black trim, switches above the keyboard, and chutes to hold, move, and save the cards being typed. See Figure 1-1. Every time the user types a character on the keyboard, several holes are drilled in the card, powerful but controlled strokes of unseen machinery. The idea is that you type in your program, in Fortran, onto the cards, ending up with a half-inch to two inches thick of punched cards. Each card had special columns with special meanings: columns 1 to 6 were for a numeric label for a given card, column 7 was for continuation of a previous statement that did not fit on one card, then most of the columns for the Fortran code, and columns 72-80 reserved for numbering the cards in the deck. This numbering was in case you dropped the deck on the floor and the cards when everywhere, and once they are all picked up and dusted off, can be assembled into the right order using the numbers. Of course none of us in the Computing Nerd Club typed in numbers, because every time you got them all perfect, it would be time to insert a bit of code, some extra cards, and the numbering would be no good.



1-1 Some computer cards, each with 80 columns and 10 rows that can be either a hole or not, thus encoding a maximum of 800 bits = 100 bytes of data. Usually most of the card was empty. Each card weighs 2.5 grams, so about 0.1 grams per byte. In 2023, a 2 terabyte flash card weighs 6 grams, or 0.000000000003 grams per byte

One of the most enjoyable activities with the IBM-29 was the auto-duplicate function. The idea was to copy an existing card deck by putting a blank card after each already-punched card, then the machine could read and remember the existing card, and punch that into the new one. In these days when computers operate silently, it is difficult to emphasise enough how much noise was produced by the IBM 29 in auto-duplicate mode. A continuous wall of sound, like a hundred hammers, so loud the floor would vibrate. And it was nice that if any figure of authority came to the door to complain, they could be sent away because “we are doing valuable computing work”. So delicious for the schoolboy to be able to talk back to his teachers!

Next, you might ask, what happens to the *chads*? The tiny bits of card that are punched from the card. That word made famous in the 2000 US election with the counting of votes on punched cards, and does it still count if the chad is still hanging on to the place it was pushed out, rather than being completely free. Anyway, the IBM 29 was a prodigious producer of chads, and they could – obviously – be used to throw at other schoolchildren. The chads were a continual sort of fog-of-war in the room that housed the card punch, always in the corners and on the floor. Rather like glitter in kindergarten rooms, impossible to completely remove.

Punched cards are really heavy! Each card has 80 columns and 10 rows, so it can hold a maximum of 800 bits of data, or 100 bytes. In fact, each card would just have part of its space used up: the statement `SIGMA=SIGMA-10`, for example, would take up a whole card, but is only 14 characters (bytes). Going up to Imperial College, those of us in the computer club would carry a few boxes of 2000 cards each, a heavy load that represented much less than a megabyte of storage. Now we have a terabyte disk in a laptop – a million times that heavy load. A 2 Tbyte thumb drive (2 grams) is more than a hundred billion times lighter per byte than those boxes of cards that I carried on the bus and the train. Later in this book we will think about deep time and deep space – a billion years and a billion light-years – but the story of my life is also one of big data, from thousands of bytes of trillions of bytes.

The boxes of cards would be deposited with the operator of the IBM 7094, which was already obsolete in the early 1970s, and that was why we schoolchildren were allowed to use it. The boxes of cards would be fed into a card reader, and over many hours the programs were read in, the programs run, and the output printed on a “lineprinter” with 14-inch-wide paper with holes in the sides. Huge amounts of paper would be produced, as each program would be printed entirely before it even started executing. Great sheaves of paper to be carried back wearily through the London rush hour and taken to school the next day so everyone could see how it went with their program. Usually, of course, there would be some minor error, perhaps writing `SIGM=SIGMA-10` in place of `SIGMA=SIGMA-10`, meaning that the offending card would be re-punched, inserted into the deck, then wait until the next courier carried all those cards to the city again. An amazingly tedious process, but one that instilled a fastidiousness and attention to detail that I had never achieved before.

The IBM 7094 was famous in a way, and featured in the movie *Hidden Figures*. The plot is about Black female scientists being the unsung heros of the early age of space exploration. While the traditional image of that time is a man, wearing a white shirt and with a pocket protector, there were a lot of smart women who are only now getting recognition. One of the was Dorothy Vaughan, a supervisor of a group of computers in the 1950s, which in those days meant a group of humans with calculating machines, keying in numbers and turning the crank to add and multiply. But NASA needed the latest technology for the complex calculations needed for the early *Mercury* and *Gemini* missions, and the new machine was ordered, the first one based on transistors rather than valves, and with the new FORTRAN programming language so that programs could say things that resemble what a human would write (`SIGMA=SIGMA-10`) instead of the old assembler code that would look like this:

```
CLA 00500
SUB 00501
STO 00500
```

It is assumed that we already know that `SIGMA` is stored in location 500, and the first instruction loads that into the “accumulator”, then we already know that 10 has been put into location 501, and it is subtracted, then the result put back into location 500.

There were a very few highly trained people who could program in assembler, but Dorothy Vaughan started looking at the new way, the FORTRAN language, and realised how it made programming infinitely easier. She set about learning, well before the new machine appeared, so that when it was installed and running, her team of Black females were astonishingly capable with their understanding of FORTRAN and how to use it for solving complex problems.

The Computing Club

I am still in contact with some of those from the computer club, and one of them has written his own account of those days. The following is from Roger Shepherd, who spent his life, like me, with a career in computers:

I started doing computing as an activity at thirteen. The school was part of "The Schools Computing Project" run by Imperial College which made use of a then ageing IBM 7094 computer which had been retired from general service. Although our programs could run for a maximum of 30 seconds, it was a day from putting the punched cards into a pigeon-hole in Imperial College, to the proceeded cards and corresponding printer output, appearing in another pigeon-hole at Imperial. For a programmer at school, the turnround was much longer. We relied on someone transporting punched card from school to Imperial College and bringing back the output from the previous batch, this would be done twice a week, on sports afternoons, when someone keen on missing out on the delights of football and cold showers, would act as courier.

In addition to the card punch, the school also won a year's access to a time-sharing system (remote access computer terminal). This came with a Type 33 Teletype with paper tape writer and reader. Even though the connection was only 300 baud dial-up, it could outrun anyone's typing speed, and to keep phone bills down, and to ensure accuracy, programs could be prepared off-line and then submitted by running the tape through the reader.

What programs did people write? There were a huge variety.

There were the simple "Hello World" class of programs. Not a great deal of programming involved, but you had to be prepare the "job". There would be a $JOB card which contained information like the maximum amount of time the program could run for (I think 10s was the default, and you could go up to 30s). Then there would be the FORTRAN program (ending with an END statement). Finally, there was an $EOF (end of file card). So much had to be correct in order the get the simplest of programs to run. And certain errors, such as omitting the $EOF card, would result in the following job not running, and the person who submitted that job having to wait another week.

Then there were some very ambitious programs. Someone had a (large) program to write a Fugue. There were programs (in FORTRAN!) that attempted to parse English text. On my part I remember writing a program, inspired by Alfred Watkins' "The Old Straight Track" which processed location information for churches, megaliths, etc. extracted from OS Maps (by hand) to try and identify Ley Lines.

Compared to what you could do today, the opportunities for "hacking" (as in cracking hacking) were limited, but they did exist. The batch computer system to which we had access ran an "operating system" and Fortran Compiler called PUFFT (the Purdue University Fast Fortran Translator). With hindsight, it was pretty neat, much lighter weight and higher throughput for small, short (run-time) programs than IBM's standard IBSYS system. One of the ways it managed to squeeze operating system, compiler and the FORTRAN libraries into core was for the compiler and operating system to use the FORTRAN run-time (I/O system etc), rather than have its own. What made PUFFT fast was that it was core resident and didn't reload between programs. Of course, there was a danger that a program might corrupt PUFFT, preventing subsequent programs from running correctly.

Even before someone managed to acquire a source-listing of PUFFT we'd managed to explore a lot of how it worked. As part of the FORTRAN library there were a set of list processing primitives, called MINI-SLIP (based on SLIP). This library included routines to read from and write to arbitrary memory locations. Using this it was possible to read and dump the 7094's memory which consisted of 32,768 36-bit words, each storing 6 6-bit characters - so 196,608 characters in all. Various things could easily be identified in a dump - in particular the compiler's error messages. Once located these could be modified and PUFFT would subsequently use the modified versions. So, you could easily arrange for the "JOB TERMINATED" message to read "BRAIN EXPLODED" instead. The challenge was to overcome the integrity check. It turned out, this check was implemented by performing a one's-complement sum of the memory and checking it against a stored value. So all that was needed was make modifications, compute the new sum and overwrite the old sum.

For more on Ley Lines and their significance, see Chapter 9.

Haiku is a finite system

Somebody who influenced me greatly is John Jaworski, a teacher at my school who started up this whole idea of teaching the pupils about computers. He smoked a pipe, so of course I had a pipe to smoke by my early 20s. Teaching logical games, then the next step, what would happen if you played forever? Unanswerable. He was teaching the advanced classes about how to use computers to analyse text, to using its detailed statistics of its style to identify the author, even to make synthetic poems. This was the age of *"Godel, Escher Bach – An Eternal Golden Braid"* that everyone was reading in 1972, about logic and knowledge, about the limits of logic, and about infinity of all kinds.

Jaworski would mutter to himself about "all possible Haikus"; the idea that a short poem style like the Japanese Haiku is a finite set of possibilities. Each poem has five syllables the first line, seven the second, and five on the third, a total of 17 syllables, for example:

Cha'os reigns with'in.
Re'flect, re'pent, and re'boot
Or'der will re'turn.

(Where I have added word-breaks to show the syllables). It is not so easy to estimate the number of syllables that the English language uses, but it could be 15,000; in contrast Japanese is said have less than 200 syllables. But let us call this N, the number of syllables. Therefore there are at most N^{17} possible Haikus. A very large number indeed, but not infinite. If we think of the Haiku as an item of beauty, we now have two conceptions of its nature: the *creation* of beauty, and the *discovery* of beauty. It is this latter idea that the Haiku shows: because the style is so minimal, there are a finite number of them. They are all pre-ordained, so it's just a matter of selection. Thus we see that from *selection* can come beauty. The landscape was there all along, it is the eye of the artist that makes it a beautiful picture.

Revenge of the nerds

Being an old machine, the IBM 7094 was easy to attack. When it was built, nobody was thinking of inky-fingered British schoolboys tinkering with the innards. There was no virtual memory or memory protection, so we just printed out the entire contents of memory. There was the program doing the printing, there is the Fortran compiler. Part of the compiler is tables of information, something particularly easy to recognize while poring over a character dump of all 32 kilobytes of memory. Tables that recognise the ‘+’ sign in an arithmetic expression, and knows to call up the assembler instruction that adds numbers. Another such table converted the characters 0,1,2,3,4,5,6,7,8,9 to equivalent numbers. They are different things, the string “5” and the number 5. The string can be used as a label (“come upstairs to Flat 5”). The number can be used to count, 4+1=5. Different things. Anyway, part of the Fortran compiler was the conversion form string “5” to number 5, and the other digits. But we not only read the whole contents of memory, but then 2 weeks later were modifying the contents of memory. The same compiler would be used for all sorts of other jobs – users from Imperial running their research codes. And if some schoolboys messed up the compiler, it messes up all the subsequent jobs. So we changed the character “5” to mean 6, and “6” to mean 5. When a program contains “VALUE = 55”, then variable VALUE would actually be set to 66. Ha ha ha! Oh dear, we got into big trouble over than one. Jaworski had to use his best diplomatic skills to repair the relationship with Imperial College.

The magic of computers

But you might ask what were we doing with these machines and why? It wasn’t just escaping the mud, it wasn’t just the freedom of roaming London unsupervised, it wasn’t just the machine-gun noise of the Card Punch auto-duplicating a deck of cards. It was the idea of precise action taken in response to precise orders, the magic of asking for something difficult to be done, and *presto*, it’s done. There was obviously some hidden potential in these machines.

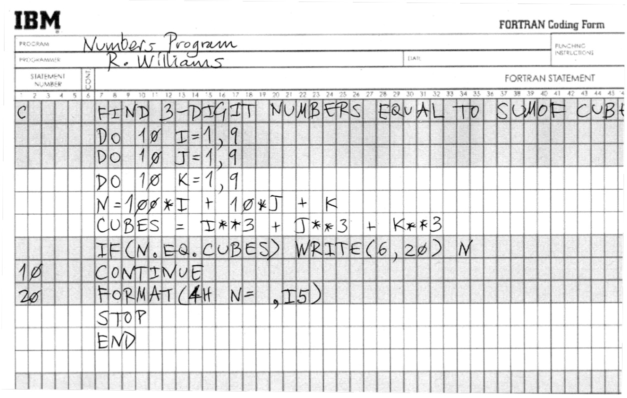
At the most basic level, computers are made from binary “gates”. The AND gate for example has two inputs and an output, the output being “on” only when both inputs are “on”. It’s a bit of electronics, and can be built with two resistors and a transistor, where “on” means working voltage (perhaps+5 volts), and “off” means zero volts. The transistor “normalises” the output, so that even it is +4 or +7 volts going in, it is exactly 5 volts going out, so properly “on”. The role of a transistor in a digital circuit is to amplify and dampen, so that high numbers become exactly the working voltage, and low numbers become exactly zero volts.

This is the magic connection between the real and the abstract world. In the real world, there is voltage, and transistors to normalise it, and in the abstract world there are the binary digits 0 and 1. Alan Turing realised this in 1936, even before there were any actual computers, and defined an “abstract finite-state machine” that manipulates binary bits on an abstract storage tape: despite the simplicity of the abstract machine, it can implement any computer algorithm. We can compute anything! We can make codes for letters and strings, store not just numbers, but information and programs to manipulate that information. It’s not just numbers, we can store images and sounds and movies! The genie is out of the bottle.

In Figure 1-2 is a coding form, from which a card deck can be punched, a deck that might be sent up to Imperial, just 13 cards. In addition it would need “job control” cards at the top and bottom. Usually these would have a different colour, so when all the different user’s card decks are all put together, you can tell where one person’s ends and the next starts. This job control says to use the “PUFFT” compiler on the following cards.

This program objective is to find all the 3-digit numbers which are equal to the sum of the cubes of the digits. It was one of my first programs, I liked the idea of all the difficult calculations that would have to be done without mistake so that the result would be right. There are three nested `DO` loops, each controlling a single digit of the number, the code inside is executed until card `10  CONTINUE`, and it starts again with the counter incremented. The number is made from its digits with 100s, 10s, and units, the sum of the cubes of the digits computed, then the comparison. If success, then we `WRITE` on unit 6 (the lineprinter output), using format card 20.

Fifty years later, the fossils of the cards are still there, at least for astronomers, through the FITS file (Flexible Image Transport System). The file is very much in use in as of the 2020s, and consists of a sequence of 80-character card images, where the header “cards” use variable names up to 8-characters, all of which must be capital letters. So very reminiscent of the card-punch back at school, and writing Fortran with variable names of 6 characters, all capital letters.



IBM — FORTRAN Coding Form

PROGRAM: Numbers Program

PROGRAMMER: R. Williams

```
C     FIND 3-DIGIT NUMBERS EQUAL TO SUMOF CUB
      DO 1Ø I=1,9
      DO 1Ø J=1,9
      DO 1Ø K=1,9
      N=1ØØ*I + 1Ø*J + K
      CUBES = I**3 + J**3 + K**3
      IF(N.EQ.CUBES) WRITE(6,2Ø) N
1Ø    CONTINUE
2Ø    FORMAT(4H N= ,I5)
      STOP
      END
```

1-2 A coding form, used to write out computer programs before being punched into the cardpunch to create a deck of cards.

Computers are very good at computing, that we had found out, and could print out a lot of numbers on that 13-inch wide lineprinter output. But not so good at anything more visual.

The Magic of Pi

As with all maths nerds, I was enchanted by Pi, that magical number that appears in unexpected places, is entwined in higher maths, and yet is so easy to understand. I read about the calculations of its value through the ages. In the Bible (1 Kings 7:23), Pi is approximated as 3: *"And he made a molten sea, ten cubits from the one brim to the other … and a line of thirty cubits did compass it round about."*. By the year 480, the Chinese astronomer Zu Chongzhi knew about the approximations 22/7 and 355/113, and by 1615 the Dutch mathematician Ludolph van Ceulen had spent 25 years computing Pi to 32 digits (figure 1-3).

By the time I started on the quest to compute Pi in 1973, it was already known to a million digits, and by the 2020's there are 10^{14} digits known. Anyway, I had found out about Machin's formula from 1706:

$$\frac{\pi}{4} = 4\arctan\frac{1}{5} - \arctan\frac{1}{239}$$ and

$$\arctan x = x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \cdots$$

The power series for arctan converges rapidly, since each term is $5^2 = 25$ times smaller than the previous term. Here are the first few terms and the partial sums for arctan(1/5):

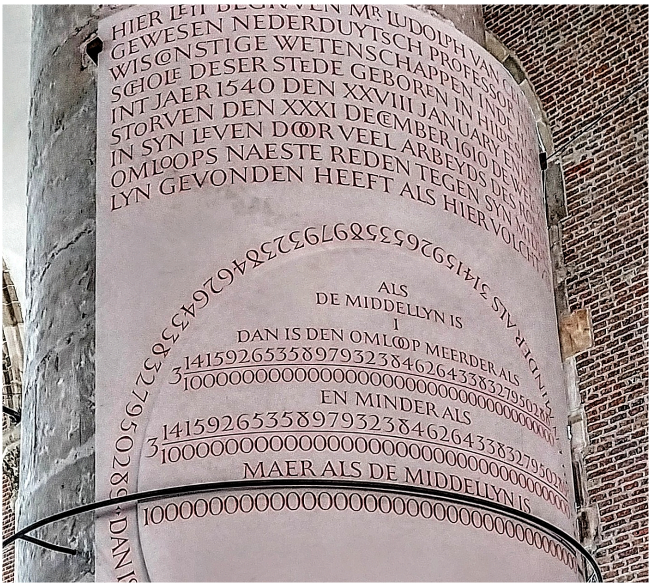
```
N'th term.       Sum
 0.2000000000     0.2000000000
-0.0026666667     0.1973333333
 0.0000640000     0.1973973333
-0.0000018286     0.1973955048
 0.0000000569     0.1973955617
-0.0000000019     0.1973955598
 0.0000000001     0.1973955599
-0.0000000000     0.1973955598
```

This was the calculation for ten digits. But the computer doesn't use 5,000 digits of accuracy, so I had to store the digits not as a single floating-point umber, but as an array of digits. At first, I started with an array of a 5,000 digits, so 3.141 would be an array if integers 3, 1, 4, 1. I needed to be able to add numbers represented this way, and divide them by ordinary numbers. I coded up the algorithms from primary school – 4 + 8 is 2 carry 1, that sort of thing. And it actually worked! I had a value of Pi to 32 digits, that which had taken Ludolph van Ceulen 25 years! Then in maths class we covered "bases", doing arithmetic in base 2 or base 16 in addition to the traditional base 10. The teacher remarked that it is really easy to convert base 16 into binary, because 16 is a power of 2: for example the base 16 number 7F3 becomes 0111 1111 0011 because 7 is 0111, F is 1111, and 3 is 011. Anyway, I realised that my code would be much more efficient if I worked in base-100000, so that Pi would be an array like 3, 14159, 26535, … and a 5,000-digit number could be stored with an array of size 1,000. I finally got the program working – it took a long time given there were only two attempts per week, when somebody went up to London with the cards. I got special permission to run the ancient room-sized IBM 7094 for three hours, and it printed out Pi to 5,000 digits. I still have that printout!

Just to finish the story, all this is very easy now. You can compute Pi to 5,000 decimal places with this Python program:

```
from mpmath import mp
mp.dps = 5000
pi = 4*(4*mp.atan(1/mp.mpf(5)))
pi -= 4*mp.atan(1/mp.mpf(239))
print(pi)
```

and it runs in 0.02 seconds on my Mac laptop, half a million times faster. And of course I don't need to count up how much memory I am using either, squeezing the calculation into 100 kilobytes, since my laptop has 32 gigabytes, just about half a million times as much.



1-3 Memorial to Ludulph van Ceulen in the Pieterskerk in Leiden, showing the 35-digit approximation to pi that he spent 25 years calculating. Today, my phone can do this in a few milliseconds.

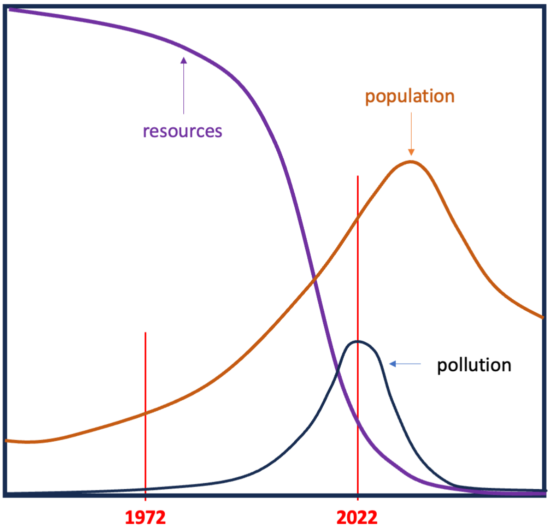
Hidden lines

My heart was taken when I saw the pen-plotter, on a trip we nerds took to the University of London Computer Centre. Yes of course we were all sniggering about getting somebody else to “push that switch” to see if the whole operation would come to a halt. But that is what schoolboys do, they snigger. What really impressed me was the pen plotter, a flat bed the size of a pool table, with a pen magically dancing across the surface dipping up and down like a bird. And the most precise and lovely drawings being made; the precision that I knew the computer could do, but made into drawing accuracy. I set myself a problem, to find out how to draw three-dimensional objects using a pen plotter. Was not as easy as I thought it would be. I thought the main task would be locating exactly where the endpoints of lines should be, the perspective transformation, and so on. But in fact that was straightforward. I was drawing shapes derived from the icosahedron (see Stained Glass, Chapter 13), so the coordinates were just the golden ratio. The real problem, it turned out, was removing *hidden lines*.

It is the difference between a wire-frame cube and a solid cube: in the former all 12 edges are visible, and for the solid cube, between 4 and 9 edges are visible. When the solid is complex, it is a very difficult computation to decide which of the wire-frame lines are hidden, or partially hidden. Not difficult for the machine, but difficult for the programmer: all the special cases where lines are parallel or collinear, lines that look like other lines, that the code (at least when I did it) became a real thicket. It never worked properly, that code to remove hidden lines from solid polyhedra, but it taught me to be a bit clearer in writing code, put in explanations, start the documentation. Because the person who eventually tries to disentangle the code is probably not somebody else, but rather oneself in 6 months time!

Limits to Growth

By 1973 I was 15, and had just started programming at school and I was browsing at the bookshop in Dulwich, where I lived, and came across the newly published “*Limits to Growth*”. The authors (Meadows, Meadows, Randers, and Behrens) are reporting on a larger collaboration called the “Club of Rome”. See Figure 1-4. The book showed results from a computer simulation for the world economy, and I was quite impressed that such a complex thing could be modelled, so much more complicated than the simulations I was trying to do – Newtonian gravitation that controlled planetary motion. So I bought the book and was fascinated. I remember thinking at that time, when I was 15, how their predictions went into the future 50 or 100 years, and for the first time thought about what the world would be when I was much older – as I am now.



1-4 The "standard" world model run from "Limits to Growth" assumes no major change. All variables follow historical values from 1900 to 1970, with exponential growth until the diminishing resource base forces a slowdown. Both population and pollution continue to increase for some time after the peak of industrialization.

The book is subtitled "… project on the predicament of mankind", and now, 50 years later, I am amazed at the authors' prescience. It starts with compound interest and exponential growth, and that many measures of human society are exponential in nature: population, energy use, economic growth, etc, and that in real-world systems there are feedback loops: the rabbits grow exponentially but then the wolves increase because of the food supply, thus reducing the rabbits. Natural resources such as metals are different if the supply is exhausted, rather than recycled; however some types of pollution are the opposite – growing but never being reduced. Much of the book is defining and justifying their modelling of feedback loops.

The Club of Rome saw the exponential increase of pollution back then, and its potential for danger, although I don't think they knew that CO_2 would be that world-smashing pollutant. They thought it would be replaced: "If man's energy needs are someday supplied by nuclear power instead of fossil fuels, this increase in atmospheric $C0_2$ will eventually cease, one hopes before it has had any measurable ecological or climatological effect.". However, they lump it in with other pollutants: CO_2, mercury, DDT, and all the others. Their famous Figure 35 is adapted here, with two dates marked by me in red: publication (1972) marked, and the 50-year anniversary (2022).

The astonishing thing about the study is its accuracy. All the major curves have been vindicated in the last 50 years, a chilling thought that has run through my entire life since that day in the bookshop in Dulwich.

Two Orphans

My whole life I have lived with not very much family. Never met any of the grandparents, and the one aunt and one cousin were thousands of miles away. So just mother and father and brother, the four of us all alone. Certainly, it made it easier to move from UK to US back and forth a few times. But I was always a bit envious of most people, who had grandparents.

My mother, Jane, and her twin, Marianne, were born in Illinois, USA, a surprise to their aged parents, Both girls did very well academically. By 1951, Jane was a professional career woman long before that was an accepted thing to do. Her work was testing hearing and fitting children with amplifiers. She owned her own car and used it to drive across the Great American West. In the summer of 1954, Jane took a five-week coach tour of Europe, and the tour guide was an engaging British man named Jim Williams, who handled all the arrangements with aplomb, speaking enough of all the relevant languages to smooth the way.

My father, Jim, was the son of a man who made leather upholstery for carriages, successful enough that he bought a house in the suburb of Kilburn in 1896, but died in 1919 from the Spanish Flu, leaving his young wife and daughter. My father spent his youth in that house with his mother and grandmother. There were summers in Suffolk, where his grandmother had grown up. Uncle Charles kept ferrets that were used for rabbit hunting, sending the ferret down a rabbit hole and setting snares where the frightened animals would emerge. There were picnics and boating, and a huge number of cousins of all ages. By 1946 my father had a house but no family, alone in the world except for distant relatives. The political feeling at the time was that communism was the best way to feed people and avoid the horror of war, and my father was caught up in this, starting a long belief in the power of communism. After graduation from the London School of Economics, he spent summers as a tour guide. He would meet parties of American tourists in London, and set off, with a coach-driver, through France, Germany, and Italy, for a tour of four or five weeks, working with hotels and restaurants, pointing out the sights, and solving problems. He also directed the tourists to gift shops, at which a commission was promised for the tour guide! Long after, while my family was travelling in Europe, he would point out tight corners of medieval streets where coach navigation was particularly difficult, or a place where a passenger had asked a fatuous question.

In the summer of 1954, Jane Irvine was living in Chicago and had a good income and a car of her own. Her mother had recently died, leaving her emotionally bereft, but with a small inheritance. Jane took the coach tour of Europe with a friend, and my father was the tour guide: by the end of the tour, Jim and Jane were very close, enjoying the romantic atmosphere of the European capitals, far from the ordinary world. The practically minded couple found themselves drawn to each other. By the time they reached Paris, they were engaged. Jane went back to Chicago, then three months later took the ship to England, and they married and moved into Jim's house. When Jane became pregnant, she somehow already knew what it took the rest of the world another 20 years to discover: that a pregnant woman should not smoke. She used her cigarette money to buy a great delicacy: fresh grapes flown in from South Africa. Jane had brought the money she had saved in Illinois, and together with Jim's house and a lot of hard work, they were able to provide for the family and also start a property development business. Jim put his energy there, as well as a career in academia, as a Senior Lecturer at the City of London Polytechnic.

My father was quite political when I was young in 1960, having seen the ravages of fascism in the 1930s and 1940s. He had inherited the house in northwest London that his mother and grandmother had rented out for forty years. Remaining in the top floor were the 96-year-old twins who had been there all that time, I remember being given cake by them, these two who were young women in the age of steam and Sherlock Holmes. Other flats were converted to the new arrangement of “bed-sitting room” or bedsit, for a single person. My father was a believer in socialism, he wanted to help the African countries freeing themselves from the British Empire, and assumed independence would be the start of democracy and the end of authoritarianism. They were rented out to a progression of students from newly liberated Ghana. All I can remember from that time is the huge smile and deep laugh, and that the palms of their hands were not the same colour as the backs of the hands.

Childhood

My brother, Fred, and I made it a family of four, living in the ground floor of the ancestral home, with the upper floors rented to students, who also served as occasional baby sitters. Through the early 60's, Jim’s friend George Marlow was buying property in south London, and encouraging Jim to do the same, converting large Victorian houses to the more modern model of self-contained flats, with money provided by what Jane had brought from Chicago, as well as by the rents. The relationship with Marlow eventually turned sour, but was replaced by a 30-year partnership with a lawyer, Roy Brien.

In 1959 our family of four moved to a house in Colindale, further out in the suburbs, with a school that my brother and I could walk to. My parents kept the Kilburn house as a rental business. At Colindale School, I remember a girl explaining to me that even very complicated sums are actually made up of lots of little sums, a principle that has remained with me my entire life. We sat together at a desk made for two, and she was left-handed, while I was right-handed. It was quite awkward when I sat on the left and she on the right, as our elbows bumped while learning writing. What a revelation it was when the teacher suggested we change places, each with our own space, no more collision!

By 1965 Jim had a day job as a Lecturer at the City of London College, teaching undergraduates in Business Studies and Economics, while spending the weekends driving to south London, and learning enough of the building trades to supervise the conversion work. The family would travel to “the Continent” in the summer, driving the familiar routes. Jim had enough languages to be able to negotiate prices at hotels at the end of the drive. He became interested in wine, with a small cellar at home, and enjoyed dinner at those restaurants where the waiters behaved with proper decorum. Jim was fascinated by history, especially economic and industrial history, not the history of the aristocracy. He enjoyed visiting ancient castles, and trying to discover the remnants of sanitary arrangements. He believed that social and political change occurs through technology and industry. Meanwhile Jane pursued art and painting, getting her unwilling family to join her in touring the museums and galleries.

The Gang

When I was eight, there was a group of kids from up and down the road. We ranged all around with a freedom not afforded to children now. We would put our fishing nets in the little stream and catch sticklebacks, take them home in a bucket of water and try to keep them alive, but they didn’t last long. My friend Graham told me that God is writing it all down in a big book and if there are too many fish deaths then I wouldn’t be allowed into Heaven. Graham and I explored the empty lands on the other side of the stream, and one day came across an isolated, deserted hen house. Being naughty boys, we had matches, and decided to keep our selves warm with a fire. Not a good idea in a wooden structure, as I found out that day. Running away, I took a look back at the blazing shed, and for weeks afterward had nightmares being arrested by the police.

In 1966, my family moved to Dulwich, a salubrious area of south London famous of its village-like atmosphere. Schools for the boys were within walking distance, as well as the growing property empire nearby. Owned with Roy Brien, the peak size was eight houses split into some 25 flats by the beginning of the 1980s. Each Saturday, there would be work to do at one place or another, the maintenance alternating with major renovation. Jim, often with one or more of his sons, would set off with complex collections of keys, specifications, notes, and tools, laying out a schedule of work for several people throughout the day, all organized with military efficiency. Weekdays, Jim would take the train to London, and try to be back for a cup of tea at five o'clock.

The house in Dulwich was perfectly located with a walkable triangle of schools, train to London, and pretty village with local shops. The school was partly fee-paying, partly supported by the government, and I especially liked the dedicated science rooms: the chemistry lab with fume hoods and etched glass bottles labelled “Conc. Sulphuric Acid”, the physics labs with oscilloscopes and electrical components. There was a display model of a periodic table, each element having a glass tube containing a sample of that element. I told my friends that I know for a fact that all the samples of elements that are colourless gases are actually just air, because who would know?

Youth Hostels

During summers of my teen years, it would be a week or two hiking or biking. A group of us would ride our bikes 50 miles a day or hike 10 miles a day, taking a train to get to the start point, and staying at Youth Hostels. I am a bit surprised at the freedom we had, managing our lives like this at 14 and 15 years old.

1-5 Youth Hostels Association membership card. Each hostel would stamp the card, so there would be a permanent record. These stamps are from the last few days of the Pennine Way, a 250-mile hike from Derbyshire to Scotland.

Each village would have a local shop, so we could get lunch or dinner and not have to carry it too far. Lunch would often be just a loaf of sliced white and a jar of jam, each of us consuming a great deal and yet staying slim and muscled because of the massive exertion.

I loved Youth Hostels and the culture, I am sad they are disappearing, and I think they are just the right thing for today's youth as they were for me. I liked the communal arrangement, being encouraged to spend time in the common room rather than in separate bedrooms, the casual meeting and friendliness. On the Pennine Way, a long distance walk, my friend and I would see the same people each night, as there was a convenient string of hostels. There would be a drying room, with a familiar smell of wet wool and socks. You couldn't be there during the day, and you were not welcome unless you arrived by foot or by bike. Everyone would be assigned a chore in the morning, like sweeping or deadheading roses. Figure 1-5 is my YHA card from back then, a rubber stamp from each visited hostel.

Trips to Europe

There were many trips to Europe where my father thought he was the tour guide, but perhaps they were really organised around the major art collections that my mother wanted to see. There were lots of old churches to visit on the tourist circuit, and I particularly remember how European churches would have a gruesome statue of Jesus on the cross, with three streams of blood and the crown of thorns, whereas the protestant Church of England would just have the geometric antiseptic crucifix so you were not aware of the details.

One of the European holidays, we met up with my father’s friend Arnold and his family, and I particularly remember a dinner the two families shared at a restaurant. At the end, the two men were arguing, and I assumed in my schoolboy way that each was trying to get the other to pay, and reducing their own commitment. But then I realised it was the *opposite*: each was demanding the right to pay the bill! How civilised, I realised, how classy, to want to be seen by your friends as the generous one, rather than minimising the cost.

Astronomy

I was born 2 weeks before the first artificial satellite – Sputnik – and have always been intrigued by the space program. Big rockets full of explosive fuel, the anxiety of launch. The stirring words of John F. Kennedy promising a trip to the moon. Even as a young child I would implore my parents to let me stay up for the broadcast from the Gemini missions, that the Americans showed at their prime-time evening TV slot. Spacewalks, docking, re-entry, the technologies needed for moon landing.

By 1967, I was 10 and reading science fiction about colonising the planets, when I started to understand the “Grand Tour” concept: an alignment of Jupiter, Saturn, Uranus, and Neptune in the 1980s that would enable a single spacecraft to visit all of those outer planets. It relied on the idea of gravity assists, where a spacecraft could be accelerated “for free” by swinging around a planet, without the need to carry extra fuel. Ten years later, Voyager 1 and Voyager 2 were launched, and in 1989 I was at Caltech with some of the world’s most famous planetary scientists, watching the Neptune flyby in real time. Some highlights: volcanoes on Jupiter’s moon Io, braids and spokes of Saturn’s rings, rings of Neptune, and so much more.

The Voyager spacecraft are still out there, still functional after nearly 50 years in space and 15 billion miles from Earth. They are astonishing engineering achievements. Now out of the solar system, they rely on technology from my youth. There is a digital 8-Track Tape Recorder to store data before transmitting it to Earth. Now hopelessly outdated, its mechanical durability means it's still running after 50 years in vacuum. The computer has only 70 kilobytes of memory, a millionth of today's smartphone. With many original engineers retired or deceased, modern teams rely on handwritten notes and drawings, and knowledge passed down through generations. Commands take 22 hours to reach Voyager 1, and responses take another 22 hours to return. Reminds me of waiting for the next trip to Imperial College for a run on the old IBM 7094 – Tuesdays and Thursdays only.

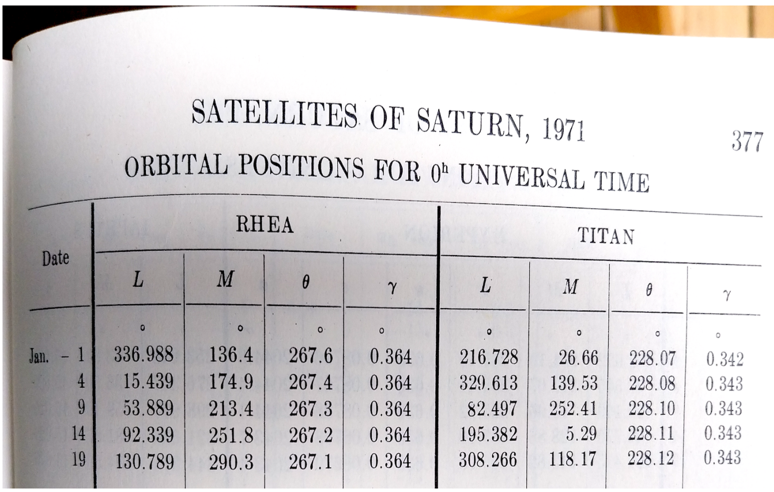
The moon landings were of course amazing, I was watching everything I could, speaking to my nerdy friends at school with the acronyms so nobody else could understand. It was a time when anything seemed possible, where mankind could do anything, the future was up and up. The film "*2001: A Space Odyssey*" really inspired me – and so many other people too. In spite of the horrors of Vietnam, there was another, heroic side to the USA. In 1972 I had convinced my parents to buy me an astronomical telescope. I had done all the reading, know that a steady mount is just as important as the optics. I sunk a piece of 4-inch iron pipe into concrete in the back garden, to which the 6" aperture Newtonian telescope was mounted. The moon was the most spectacular, the edges shivering from the Earth's atmosphere. But the rest of the heavens, blah. Jupiter was a tiny disk and I could barely see the Red Spot. Mars was a featureless red disk. The best time for observing was the dark winter nights, and of course it was bloody cold out there at midnight. By this time the first images were coming in from Pioneer 10, absolutely spectacular detail of Jupiter and its moons. Ever since then, I've been fascinated by astronomy, but never again as an observer, rather using computers.

There was an astronomy club at my school, with various nerdy people meeting to talk, I can't remember anything we achieved. All I remember is the "Banda" spirit duplicator that was owned by the school, and we had permission to use to duplicate our agenda and minutes. First the master sheet was typed on to special paper impregnated with wax, which would then be inserted into the machine, then an absorbent pad soaked in "duplication spirit" would dissolve the ink and put it only where the wax had been removed from the master. The operator would turn the crank, and gradually get rather lightheaded from the fumes. It was methyl and ethyl alcohol, with a bit of sweet-smelling trichlorofluoromethane – the active ingredient of the whole intoxicating process.

Celestial Mechanics

While the other boys at my school (no girls) knew all about football players, my hero was Isaac Newton. People have this facetious idea of him sitting under an apple tree in 1666, then the apple falls on his head, then he says "Aha! Gravity!" and claims to have invented it. But the real story is that he wondered why the apple falls toward the centre of the Earth, and wondered if the attraction of gravity extended much further, even to the distance of the moon. Newton conceived that gravitational force is proportional to the inverse square of the distance between masses. By 1687, Newton had applied the tools of newly-discovered calculus to compute orbits of two masses – such as Earth and Moon, or Jupiter and Sun. He recovered Kepler's laws of planetary motion, that had been published 80 years earlier: that planets travel in elliptical orbits, going slower further from the Sun, and that the cube of the orbital radius is the square of the orbital period. The old word planet means wanderer, because to the ancients the way they wandered the sky was unpredictable. But after Newton, their motions were absolutely predictable, a triumph of 17th century science.

Newton's theory, extended and elaborated by Lagrange, Laplace, and Tisserand was able to predict the motion of the solar system with exquisite precision. The boys of the astronomy club were punching out cards with FORTRAN programs to compute these orbits, reading up on how to predict eclipses, and other things that most 15-year-olds are not doing.


SATELLITES OF SATURN, 1971 377

ORBITAL POSITIONS FOR 0ʰ UNIVERSAL TIME

Date	RHEA				TITAN			
	L	M	θ	γ	L	M	θ	γ
	°	°	°	°	°	°	°	°
Jan. − 1	336.988	136.4	267.6	0.364	216.728	26.66	228.07	0.342
4	15.439	174.9	267.4	0.364	329.613	139.53	228.08	0.343
9	53.889	213.4	267.3	0.364	82.497	252.41	228.10	0.343
14	92.339	251.8	267.2	0.364	195.382	5.29	228.11	0.343
19	130.789	290.3	267.1	0.364	308.266	118.17	228.12	0.343

1-6 Page from The Astronomical Ephemeris 1971. It was obtained at Her Majesty's Stationary Office in High Holborn, London. (Astronomical Ephemeris, USNO, 1971)

We were also rather interested in data: the predicted positions of celestial bodies, known as *ephemerides* (Figure 1-6). Not that we could do anything with the data, but like a stamp collector we just wanted to *own* the data. So there were trips to London to buy the books with these long tables of numbers. They were published by the British government, since 1786, because in the past, astronomical ephemerides were an aid to navigation of ships, thus contributing the creation and maintenance of the British Empire.

Finding Neptune

The greatest triumph of celestial mechanics has of course been the navigation of spaceships throughout the solar system since the 1960s. But well before that was another triumph: the discovery of Neptune.

Since ancient times, there have been seven planets: Mercury, Venus, Earth, Moon, Mars, Jupiter, and Saturn. In 1781, William Herschel was hunting for comets from his back garden in Bath, England, and thought he saw a comet, but the nearly circular orbit convinced him it was a planet. For the next eighty years, Uranus completed an orbit around the Sun, and its position was measured by astronomers to accurately compute its orbit and future ephemeris. But there was a problem: for the first time, Newton’s theory of gravity was not sufficient, as the planet was not exactly where it should have been.

Two mathematicians, independently, in 1845, had the idea that the irregularities of the orbit of Uranus could be caused by yet another planet, as yet unknown, and even further from the Sun. They were John Adams, a brilliant mathematician who had just graduated from Cambridge, and Urbain Leverrier of Paris, who had spent a lifetime in astronomy, and knew everyone. Both used the irregularities to compute the mass, position, and orbit of the inferred planet. Adams communicated his work to James Challis, the director of the Cambridge observatory, in September 1845, but somehow it was lost. A later letter, with more accurate prediction elicited the remark from Challis that “while the labour was certain, success appeared to be so uncertain”. By June 1846, Leverrier had published a similar conclusion, and the English tried to find the mystery planet. In September 1846, Leverrier communicated to his colleague in Berlin with a position accurate to one degree, and Neptune was discovered the same night. There was a great controversy between France and England about the true discoverer, with criticism of the English establishment of failing to support young talent. However Adams was always humble in ascribing the discovery to Leverrier. The discovery showed also that Newton's law of gravitation prevailed even at the limits of the Solar System, at 4 billion kilometres, unifying the attraction of the apple to the Earth and the attraction of Neptune to the Sun, just as Newton had theorised.

Another discrepancy of measurement from Newton's gravitation was discovered by Leverrier: the precession of the orbit of Mercury. After the success of discovering Neptune, a new planet Vulcan was hypothesised, with an orbit between Mercury and the Sun. But the explanation was not so simple, and had to wait for the brilliance of Einstein, see chapter 9.

2. Birmingham

Non-Ferrous Metals

The end of school was a bit lonely. All my friends had finished exams in summer 1975 and headed off to University or apprenticeship, and I stayed on for an extra term at school to swot for the Cambridge Entrance Exam. My tutor was Dave Wallis, the new maths teacher. I had been very impressed when he arrived and after a bit of a chat we talked about the idea of a space with “infinity minus 4” dimensions, which seemed intriguing. I recently met him again, many decades after 1975, and he revealed that he was just as afraid as I was of the questions on the exam, and worried about his ability to teach. In other words we were both a bit anxious.

After the exam, I worked as an intern for the now defunct Imperial Metal Industries in a down-at-heel area of Birmingham, living at the YMCA and taking the bus every day to work. I left my parents’ house on a frosty January day, my brother driving me up there. I was deposited in a little room with my music cassette player for company. But within a couple of weeks, I struck lucky. On the noticeboard at the YMCA was an ad for the North Birmingham Youth Hostels group, so I went over to meet with them. Turns out they were exactly what I needed: Friday evening we would crowd into a couple of cars and speed off to a Youth Hostel, Saturday with hiking and then to a pub in the evening. A real sense of community that made that time in Birmingham quite acceptable.

At the YMCA, breakfast was the same every day: bacon and eggs and fried slice (of bread). There were a few “custom characters” let’s call them, but one morning I sat at breakfast with Martin Atkin. We gingerly chatted, he was also just out of school, doing an internship at another non-ferrous metal company. How funny that we were both with companies making semi-finished non-ferrous metal parts. But then we talked about where are you going next, and well yes its University. Being surrounded by regular people, it was easy to be branded a snob or a Tory because of going to University. Then – and this was very sheepish – actually yes its Cambridge. And actually we were both headed to the same Cambridge college! Non-ferrous metals and the same Cambridge college, what a coincidence! Martin of course came along on the Youth

Hostel weekends, and was just as pleased as I for having this little community of friends.

IMI knew I was on my way to University, and the idea was that I would come back for a career there. The work was interesting, a bit of statistics of which products sold at which rates, some programming in Fortran. I made a statistical model of sales and stock at the Coventry branch, the conclusion being that the stock levels could be seriously reduced, freeing up capital, but without impacting the customer experience. After writing it all up, and presenting it to the manager of the Coventry branch, he seemed a bit stunned, then spat out "I won't have that little twerp telling me how to run my business". So I learned that being right is only a little part of the story, that reaching out to people is much more important!

The programming at IMI was a strange experience. They used the same card punch that I had used at school, the IBM 29, but there was a typing pool of women who operated them. I would write out the code on pre-printed "coding forms", give it to the typing pool manager, who would give it to one of the typists to be punched on cards. Of course it was full of mistakes! The typists were able to turn out a perfect result when typing English, indeed they would correct mistakes in the original. But Fortran was a foreign language to them, and they could not see what was right or wrong. And a language with strange constructions, like "`FUN(A(I+1))`" and "`FORMAT(I10, 10H TOO LARGE)`". The computer notices differences that humans might not – lowercase `l` and capital `I` and `1` are all different.

A couple of times while in Birmingham, I visited Martin's parents' house in Preston, hitchhiking up the motorway. I recall a gracious place, where attention was paid to how things look, it wasn't just functional like in my parents' home. There was a big Irish Setter dog, beautiful but a bit stupid. Whereas my family always would worry about dog hair on the sofa, Martin's family simply enjoyed having the big dog around. There was one room I was particularly taken by, quite small and with swathes of Arabian textile draped around the walls and furnished with big soft pillows. Quite a non-functional room, but just delicious to be in. The sound system made it an all-enveloping experience. I couldn't imagine my stolid parents spending money and effort on such a pointless, but wonderful place. In other words, Martin's parents' house made me realise there is more to life than good value for money and plain speaking.

Strange Numbers

There was a beautiful book published in 1979: *Gödel, Escher, Bach: An Eternal Golden Braid*, by Douglas Hofstadter. It was a popular account of all the deep thinking that had happened in the 1930s about the nature of mathematics and computing. Naively, we think of them as separate empires. First as babies we learn the natural numbers, 1,2,3,…, then zero and negative numbers, rational numbers like 3/7, and on to decimals and infinite decimals like 1/3 = 0.333333… recurring. We get a tiny peek under the curtain when the maths teacher talks about pi, and its March 14 (pi day in the USA), and we consider the decimal expansion of pi that goes on and on for ever. What does that mean, goes on for ever? Who will do the counting to ensure it goes on for ever? A peek under the curtain from the monochrome, ordinary world to the bright colours and high peaks of advanced maths. What does it mean for this decimal expansion to go on "for ever"? Suddenly we have an infinite series.

Consider the infinite decimal x=0.99999… recurring. Two things are obvious: that any finite version of x (say 0.999) is definitely different from 1.0, but also that the longer the decimal expansion of x, the closer it is to 1.0. It is now a leap of faith to assert that actually x=1.0, that this one number has two names: 1.0 and 0.99999… recurring. Back to pi. Each finite decimal expansion (3, 3.1, 3.14, 3.141, 3.1415, etc) is actually a rational number (3, 31/10, 314/100, 3141/1000, 31415/10000), and yet the limit of this sequence is not a rational number. In fact pi is not only irrational (proved by Lambert in 1768), but is "transcendental", meaning that it is not the solution of any polynomial equation with rational coefficients. The existence of these transcendental numbers can be proved by counting arguments: since rational numbers are countably infinite, and therefore the numbers that are roots of polynomial equations are also countable; and yet there are an uncountable number of real numbers. Therefore there are a lot more transcendentals than the numbers that we know about. Even though it is not difficult to prove that there is an uncountable infinity of transcendental numbers, it is actually very difficult to find one and prove it is so! It was not until 1882 that Pi was shown to be transcendental.

Back to *Gödel, Escher, Bach.* Turns out there is a lot more magic that can wrung out of the number system. The book is lively with the artist (Escher) and the musician (Bach), but the real heroes are Gödel, Church, and Turing. For years, mathematicians had been trying to put their subject into a strong framework of axioms: start with some basic truths (the axioms), and then be able to prove everything that is true and disprove everything that is false. The search was for an automated way to prove theorems that would prove everything true about numbers.

For example, take this simple "theorem": ***There is a whole number which when multiplied by 2 equals 6***. I think we can all see that the theorem is true because that number is 3. A slightly more complex theorem is: ***There are no whole numbers a, b, c so that*** $a^3 + b^3 = c^3$. This latter is difficult to prove, and requires several pages of mathematics. Such statements and be encoded into a formal language, with symbols encoding concepts like add, multiply, "for all" and "there exists". This idea of encoding statements into numbers is what happens when computers execute a program. A proof of a given theorem can be written as changing the encoding of the statement step by step, in ways allowed by the axioms, until a true statement is reached (perhaps 0=0 or another axiom). Once theorems can be encoded as symbols, each symbol can have a numerical equivalent, and now a theorem is encoded as a number. The axioms of logic provide a way to convert one of these numbers to another, and eventually we get to the encoding of "0=0", which we know is a true theorem. We therefore have a set of numbers that represent the theorems that are provable, let is call these "Gödel numbers".

Now comes the heist, the place where we are swept off our feet. Every theorem about numbers is itself a number, so we can encode some paradoxical statement into such a theorem, something like ***"This statement is unprovable"***. This statement is undecidable – neither true nor untrue – and yet is generated by the formal axiomatic system that is supposed to generate all true mathematical theorems and no untrue theorems.

From there the book describes the "Turing Machine", an abstract model of a digital computer that is simple enough to analyse in detail, where programs become Gödel numbers, and yet complex enough to describe almost all the computers we use today. Turing invented the model in 1936, and its legacy still lives on. All these ultrafast logic machines that are so much a part of our lives are at root Turing machines, (except perhaps the new-fangled quantum computers). Once the Gödel encoding is available for computer programs, Turing proved that it is impossible to decide algorithmically if a given program will halt or continue in an infinite loop. And that there must be numbers that are *not computable*. While Pi hides among the infinite number of transcendental numbers, now we find out that there is a huge host of other numbers that we cannot even compute!

The book is one of those with so much rich material that you can't just read it from start to end, you have to keep going back ("hang on what was that again?") and checking the thread. I was absorbed for weeks.

3. Cambridge

Living the Dream

A year in dingy Birmingham made me very grateful to be in graceful ancient Cambridge. Ever since then, I have striven to live in a place of beautiful architecture.

Cambridge offered something alluring, something I have searched for ever since. A sense of history and tradition, yes, but also a sense of things being done a certain way because they have always worked well. The college Hall, basically the same plan that Henry VIII or the Anglo-Saxons would recognise – long tables scarred from use, with gloomy pictures of the ancestors all around. Everyone eats together – they Break Bread – and that encourages collaboration and a sense of common purpose. The people in charge sit at the most impressive table and give the word to allow the eating to start. At my college there is a musician's gallery at the other end of the hall; during a ceremonial dinner (a “feast”) the sounds of Tudor instruments drift down from high above. In the great kitchen beside the Hall are huge turtle shells on the walls, with painted calligraphy "Feast of the Benefactors 1885". From a time, obviously, when turtle soup was a celebration of the power of the British Empire rather than an assault against Nature.

Accommodation was in a “set”, meaning a 2-bedroom flat with two students sharing. Martin and I were in “New Court”, built in 1825, with lovely views over “The Backs” – the gracious lawns surrounding the river. Friends would come by for tea parties like in *Alice in Wonderland*, toasting crumpets at the gas fire with a toasting fork, and eating Chelsea buns from the famous Fitzbillies bakery. Buns doubly slathered in syrup, oozing it everywhere. With tea in cups with saucers and holding the pinky finger up while lifting the cup, to look elegant and regal. It is said that the custom arose in the royal court, where a lady would indicate interest in a man by raising her little finger this way. The music would be dreamy house music, perhaps Keith Jarrett in the *Köln Concert*, orgasmically sighing and groaning to the piano.

Walking to lectures every day I went through ancient lanes that Isaac Newton would have recognised. I would take a detour on the way to be in the magnificent Kings College Chapel, how simply being in that space, often all alone, would uplift my spirit and get me ready for the algebra and differential equations waiting for me in the Maths lecture. Coming back to college on a winter evening, the fog rolling in from the Fens, the ancient architecture lit by lights haloed in the fog. Once again uplifted by beauty.

These days, Cambridge seems a lot more locked down than it was, with hordes of tourists in the public streets and the college porters working hard to gather payment so they can walk in. Trinity college has a large and beautiful Great Court with an architectural fountain in the middle – which was often in use for student's birthdays, as their friends would dunk them in the water at midnight. Another custom of the college was the “Great Court Run”, where the best runners would wait at a corner of the Great Court for the first stroke of midnight from the clocktower, the race around the circumference (370 metres) before the last stroke (44 seconds). Even Olympic athletes cannot finish in time! The Great Court Run was memorably part of the film *Chariots of Fire* (1981), even though it was filmed at Caius college, not Trinity.

Punting

One of the first days after arrival a few of us took a punt for the afternoon (Figure 3-1). It’s a heavy, flat-bottomed boat, originally used to move around the marshy land on which Cambridge was founded. Gathering reeds for thatching, and harvesting the eel traps, that sort of thing. The person driving the punt stands at the back, holding a pole that is about 6 meters long. They hold it vertically, then allow it to fall through their hands until it is stopped by the river bed, then push it to propel the punt forward, and when the pole is trailing behind, push to the left or right, like a rudder, to adjust the forward direction. Once it's into a rhythm, everything works well: fall-push-rudder-lift. But it takes a bit of practice to get the knack, but it was my first day out, and there was a bit of a problem at Silver Street bridge.

It was a warm day and a hundred people were sitting in the sunshine outside The Anchor pub, as I pushed my punt forward. But I misjudged the bridge, and lifted the pole for a last push to carry it underneath. Unfortunately, the top end of the pole didn’t quite clear the bottom of bridge, and of course the bottom end was stuck into the river mud, with me holding on to the middle. I thought the punt would slow to a stop, but there were five people in it, plenty of inertia, and it kept going under the bridge. With me holding on to this vertical pole, suspended over the water. The pole gradually started tipping over and then of course Splosh! In the river! As I came up for air there was a round of applause from the pub patrons. Luckily punts also come with a short oar, and my loyal friends were already on the way back to pick up my sorry self.

3-1 Punting at Silver Street Bridge in Cambridge. When your punt passes under the bridge, it is important not to have one end of the pole in the river bottom, and the other end on the arch, and the punt moving away under you. (Nigel Graver, Shutterstock)

I had a lovely time with mathematics. Such a joy to simultaneously discover that I could do the work, and to have a girlfriend for the first time in my life – also a mathematician, so a lot of snuggling over the abstract algebra and axioms of analysis. Now Sarah Rees is a professor of mathematics at a prestigious university! Before coming to Cambridge, I had seen a book from 1937, *Night Climbers of Cambridge*, and wondered about having a try. Of course a tremendous courage and confidence is needed to be high enough that a fall might be fatal, yet relying on dimly-seen projections as hand-holds. I did not have such courage and confidence.

In the old days, the Cambridge colleges were worlds unto themselves, and closed up like a fortress every night. To get in past curfew, you needed to knock at the ancient door and the porter would take your name, and there might be some kind of scolding or fine or something. Anyway, Sarah and I were keen to give it a try, and one late evening, when the college was closed, we scaled Nevile's Gate, barely avoiding a disembowelling from the pointed railings (figure 3-2).

One of the less appealing features of the ancient College was the snobbishness. Many students got their place through hard work (ahem), but it seemed that "the other half" got there because their parents had paid to put them through an expensive private school. At the most expensive of these, the headteacher knew the Oxford and Cambridge admissions people personally, and could make a phone call to get a place for a particularly stupid, but very rich student. Many of these would move in herds, making noise and mess, throwing sherry parties, and so on. The brighter ones would go into broadcasting or politics, while the less able studied "land economy", so they would know what to do with the vast acreage they were due to inherit – with a nice family trust of course to avoid inheritance tax. There was a film made about these people, *The Riot Club* (2014), subtitled "Filthy. Rich. Spoilt. Rotten."

At Cambridge, I was lucky to meet some people that remained friends for many decades, including David Womersley. He studied English, and he explained that it isn't just summarising the plots of novels and plays, which is how I had always tackled the subject at school. Rather there was an rigorous analysis of metaphor, allegory, foreshadowing, and so on. David was quite good at this game, and went on to be a professor at Oxford. In 1988, I would become godfather to his daughter Kate, who now lives in Edinburgh, near me. She is a medical doctor, a researcher, a writer, and a mother, and is very busy and quite charming.

3-2 Nevile's Gate, Trinity College, Cambridge. A possible climbing route is shown by the red stars at left. Do not attempt this! (photo by the wub under CCAS3.0 via Wikimedia commons)

A Summer in the USA

I applied for a scheme to spend summer vacation in the USA, and there were all sorts of requirements – the Americans didn't want any hippy types sponging off their welfare system – so there had to be a sponsor writing a letter saying I could stay with them, as well as the British University saying I was OK to come back in the Autumn. There were lots of hours of standing in line at the US embassy in London. Finally I showed my papers to the clerk, who refused me on the grounds that the sponsor letter didn't have the envelope it came in – which I had binned. After a bit more waiting in line, I got some time with an actual consular official. She wanted to know more about the sponsor, and her ears perked up when I said it was my aunt and uncle, and my mother is a US citizen. She suddenly smiled broadly “Well if your mother is a US citizen, we’ll just give you a passport!”. So that solved the problem of the visa, and informed me that I am a US citizen. I spent a summer working night-shift in a Styrofoam cup factory near O’Hare airport in Chicago, and got to meet some real Americans. I learned not to tell my supervisor that my work was all done, because that would mean sweeping the floor.

Driving my uncle’s car to the factory late one night, I did not see the place where the road became divided by a concrete kerb, and the front wheel hit the kerb. The tyre deflated and the wheel was bent. While I was pulling out the spare and working out how to change the wheel, a police car pulled up, red lights flashing, and a man with a gun got out. This is where I found out the difference between the friendly British bobby and the American police. He spotted a sign that had fallen down, a sign warning drivers that the road was divided. If the sign had been in place, then I wouldn’t have hit the kerb! Anyway, he was all ready to arrest me for colliding with the sign and knocking it down with the car. Writing up papers and so on. But my brain kicked in and I went to the front of the car, then pointed out that there was no damage, no dent, no scratch on the front of the car, so how could I have knocked down the sign? Luckily he backed down and admitted I was right. Then helped me with the spare wheel, so that turned out well. The next day a man from the factory took me to his home, with piles of automotive parts, and found me a brand-new wheel. So that also turned out well.

I took a 3-week Greyhound Bus tour of the USA, from Chicago to Los Angeles, and back to Washington and New York. I somehow got invited to a party in the Hollywood Hills, and I remember being on a balcony with the sweep of the lights out to infinity, stars twinkling above in the warm velvet night. The Americans all full of optimism and promise, not like the British, who were falling into a recession and complaining about the weather. I had not known I was an US citizen. That passport, plus the memory of that party, decided where I would be two years later. Anyway, there was a certain rhythm of the Greyhound Bus, the long hours of driving and the rest stops. A later section of this chapter is a little story I wrote after that long bus trip.

Infinity and Limits

At Cambridge I learned about different kinds of infinity. We all realise early on that there is no largest number, because you can always add one to it, and yet we talk about that largest number as if it really exists: “infinity + infinity = infinity ha ha”.

The first key concept is countable versus uncountable infinity. Quite instructive is “Hilbert’s Hotel”, that has an infinite number of rooms numbered 1, 2, 3 etc. Even when full, it can always accommodate another guest, by moving the guest in room 1 to room 2, the guest in room 2 to room 3, and so on. The hotel can also accommodate an infinite number of new guests: move existing guest in room 1 to room 2, the guest in room 2 to room 4, so that the even-numbered rooms are full and the odd-numbered rooms all empty, ready for the infinite number of new guests.

Notice that in the Hilbert Hotel, the rooms have numbers 1,2,3, and the new guests also have numbers 1,2,3: they are *countably* infinite, meaning there is a one-to-one correspondence with natural numbers 1,2,3, … But there are other kinds of infinite sets that are not countable, in particular the real numbers, meaning numbers with an infinite string of digits, like Pi. There is a famous proof by Georg Cantor that the real numbers are not countable. It is a proof by contradiction: if we assume the real numbers are countable, then we can find a number that is not in the sequence, and so the initial premise must be false

The second key concept is a rigorous approach to limits, invented by Augustin-Louis Cauchy in 1821, and is known as the *epsilon-delta* or *epsilon-N* definition. Take for example the sequence of fractions: s_1=1/2, s_2=3/4, s_3=7/8, s_4=15/16, s_5=31/32, etc. You just look at it and blithely state that at infinity the value is 1; because you can see it getting closer and closer. But what exactly does this mean? Cauchy’s idea was to set up a challenge: in the epsilon-N formulation the challenger provides a very small positive number epsilon, let’s say 0.0000000001. Then the other person must find N so that the sequence is – and stays – within epsilon of the limit value 1. For the sequence above, s_N is within epsilon of the limit as soon as N is greater than 34, and stays within that bound as N increases. And therefore, by Cauchy’s definition, we can say that the limit is 1. Notice that you don’t need to use the word infinity.

The epsilon-N and similar epsilon-delta ideas add rigor and structure to much of the mathematics of the last 300 years. While Newton computed the motions of the planets with his version of calculus, it was all hand-waving until Cauchy came along. One example is the idea of a continuous curve. Heuristically, it a curve is one that can be drawn without taking the pen off the paper: a circle or a parabola for example. The rigorous version of the idea of continuity doesn't involve pencils or paper: if a function f(x) is continuous, then for every epsilon there is a delta such that changing x by less than epsilon causes changes in f(x) that are less than delta.

Greyhound Bus Romance (1977)

This is an account I wrote in 1977, describing a bus trip across the USA.

Finally he got off that damn bus. I saw him ask an official, almost as beef-faced as himself, where was the bus for Little Rock, Arkansas. That was something that annoyed me about him: every other gape-mouthed man he met was told in no uncertain terms about how he was going to see his nephew in Little Rock, Arkansas. I have more grievances. There was some kind of mauve stuff under his fingernails, and he had something like old onion exuding from both the dark patches on the armpits of his shirt. For the last eight hours, through the top of Texas and Oklahoma, his body repeatedly slumped into my half of the seat, grunting. And each hair on his arm was surrounded by a tiny pink patch, I dunno why that upset me. You can imagine why I readily talked to another human being, when I had washed and shaved myself in the washroom in Tulsa Greyhound Station.

I was reading an account of the downfall of the Russian Czar. All through Oklahoma and Missouri the Empire fell. The political situation became extremely tight, and I'm sure a few more pages would have seen the entire royal family shot.

The rest stops are good for peeing and leg-stretching, but I have this feeling when I travel that when I'm not traveling, rather I'm wasting my time. A pint of chocolate milk made up my supper: I had a sort of jackdaw urge when I was in the States not to spend my hard-earned dollars, and besides, not spending money is just like emulating Bob Dylan and Woodie Guthrie, and after all, they are folk-heroes.

Those Greyhound buses are always hell to get on to after a walk in the fresh air. I'd been down to the St Louis Centennial Arch, and laid down with my eye, the top of the arch, and a star all in a line; just to see if it really swayed in the wind like they say it does. In the bus there are candy wrappers, and brown spills of Coke, and small children sticky with having spilled on themselves, and grubby flustered people who haven't had the sense to go for a walk. And Marines drinking beer and jostling people like me. Perhaps it's un-American to jostle old ladies or cripples. The bus was pretty empty so I was hopeful of a double seat to myself. But I started talking with a woman in the rest-stop, her name was Deborah, and we sat together. In spite of wanting to sleep to pass the time.

But Deborah and I talked. And talked and talked.

"… and this guy didn't convince you about … about the supernatural?"

"No my scheme of the Universe …"

"You don't ever feel that an aura might exist from some people - don't tell me you've never felt unreasoned like or dislike of a person without ever meeting them?"

"Oh well –"

"Yeah, and you were so busy telling me about all your math and stuff, that I didn't even have a chance!"

We decided to get a bottle of something at the next stop and loosen up. But Effingham, Illinois was so crazy Puritan there were no liquor stores open. Not at 1 am anyway. Deborah taught ballet in Trenton, New Jersey, and we sat on the bus and watched the sun rise over the low-lying mist. We thought we'd try and sleep. There was nothing for it if I wasn't to be torn up with tiredness later. The half second between making that decision, for half second it must be, and actually kissing her was like an age. My head lay on her shoulder for a full minute before my neck became stiff, so we talked more. Mmm ... Bloody Mary ... Together People ... Snug.

Indianapolis Bus Station is not to be recommended after a hangover. It's fluorescent light revealed our faces and we both revised our visual impression of each other, on the basis of what we had been saying. Her nose jutted clearly and forcefully; let that be all of description. I have vague memories of deserted streets in downtown Indianapolis, and fire-hydrants like little soldiers, and a very American looking policeman carrying big guns. There was a bar with some weasel-like people in it. Deborah strides in impressive style asking for drinks and Marlboros; Roy pretending to impressive style and tagging behind, eyes slipping involuntarily to the notice about Indiana state liquor regulations and being over 21. I really felt seventeen again. No drinks for weary travellers though, even after a harangue which at one stage seemed to be going the right way. She swept out like a queen without a crown. I had the feeling I catch from being near pretty women, that why is she talking to me, I'm just a small vole. The feeling soon left me: I've always been a hedonist, so if I'm enjoying myself I let sleeping, well, dogs lie.

We'd discussed the word ‘nice’. It's one I've been taught to hate from an early age, and I use it therefore scornfully. One of my teachers from primary school explained that it is a negative word, expressing little. I've always tried to be economical with words, so it is offensive to me to substitute just ‘nice’ for anything from ‘fine’ to ‘jolly’ to ‘splendid’. Deborah said I was a sarcastic Englishman, and my use of ‘nice’ she quoted as an example; that I use it cynically to refer to what a person who uses ‘nice’ often appreciates. Thatched Cottages. The 1812 Overture. I wasn't meaning to be nasty, saying the State Capitol Building was nice, but Deborah had liked it, so I guess I was being nasty. She even taught me a little of her craft underneath a memorial which was dedicated to something. We splashed in the fountain. Hers was the first ear-lobe I ever hurt. We had a glorious snog like wet liver sloshing in a brandy butt.

Ohio struck me as a pretty state, though I'd been told otherwise. It seemed too like England after the hot dryness of the Southwest. It was green and sweet, with cows and saplings. And pink haze in front of the eyes helps. I explained that I thought many friendships – or should I say 'friendships' – are based on a subdominant personality making way for a dominant one. Rather like a dog and master. We both seemed to

be holding a leash to some others, we agreed, and Deborah seemed to have become more of a friend now, in so far as my heart wasn't thumping so much. She told me my psychological theorems were crazy. Oh well there went another part of my Scheme of the Universe. Perhaps a good thing too, it always was a cumbersome thing to carry around. There was a deep white mist outside, the cows wading through it, and a barely visible sun trying to rise. We were due into Pittsburgh at noon, at which point Deborah was to take the New York bus and I the Washington one. Surely a place like Pittsburgh would have a liquor stores. She changed her clothes in the Ladies Room there and made a few sneaky comments on how often Roy changed his shirt. As a matter of fact I hadn't changed it for some time. Deborah was wearing her sexy (as she called them) jeans and kept smoothing them over her thighs, Like I imagine Faye Dunaway doing offstage. Two double Bloody Marys, two 'businessmans' lunches in a fast food hole, a peck on the cheek and bye.
"Next time you're in Trenton.."
"Next time you're in Cambridge.."
And that was that. A twenty-four hour love affair.

The Magician of Mathematics

I can’t leave Cambridge behind without mentioning the magician – Ramanujan.

Ramanujan was the first Indian to be a Fellow of Trinity College, and one of the youngest Fellows of the Royal Society. His notebooks continue to be analysed and studied as a source of new mathematical ideas, where comments about “simple properties” and “simple results” lead modern mathematicians to profound new areas of research.

Ramanujan suffered from ill-health much of his life, now thought to be amoebic dysentery contracted in Madras. He spent time in hospital in Cambridge, and went back to India, where he died in 1920, aged 32. His life has been popularised in a film “*The Man Who Knew Infinity*” in 2014.

When I arrived at Cambridge, Hardy, Littlewood, and Ramanujan were my heroes, and I was so proud to be at the same place they were. I ran into Littlewood one summer morning, walking in his bathrobe, aged

90, across the architectural splendour of Nevile's Court. I stopped and said I knew who he was and who Hardy was, and he gave a little smile before heading into the bathroom. I felt touched by glory.

In 1913 a mathematician named G. H. Hardy, Trinity, Cambridge, received papers with a cover letter that began: "*Dear Sir, I beg to introduce myself to you as a clerk in the Accounts Department of the Port Trust Office at Madras on a salary of only £20 per annum. I am now about 23 years of age….*" and went on to talk about divergent series and the distribution of prime numbers. The cover letter ended: "*Being poor, if you are convinced that there is anything of value I would like to have my theorems published…. Being inexperienced I would very highly value any advice you give me. Requesting to be excused for the trouble I give you. I remain, Dear Sir, Yours truly, S. Ramanujan*".

Hardy and his colleague J. E. Littlewood were fascinated by the papers, containing some results that were unknown and far from their experience, yet strangely beguiling. And other results that were well-known. Ramanujan, it turned out, was completely self-taught, so derived his results without knowing if other mathematicians had already done so. There were no proofs in the papers, yet Hardy and Littlewood could see the truth in the exotic results. Later Hardy said the results "*must be true because, if they were not true, no one would have the imagination to invent them*." Ramanujan was a great calculator, and much of his mathematics was experimental – where the result was determined by calculating, then it and other results derived more rigorously with algebra and calculus. His astonishing mind produced many bizarre formulae, all correct, and yet he hated the conventional mathematical proof. For example this is true to 18 decimal places:

$$e^{\frac{\pi\sqrt{190}}{12}} \cong (2\sqrt{2}+10)(3+\sqrt{10})$$

and if that doesn't make you slack-jawed with wonder I don't know what would. Hardy and Ramanujan exchanged letters, and finally Ramanujan was invited to Cambridge, and arrived in April 1914. While the three mathematicians made great strides in analytic number theory, modular forms, and other lovely things, Ramanujan suffered: from the cold damp weather, from the intolerance of his vegetarian diet, from the racism all around him. Nobody wanted to believe that a brown man is a genius.

Some beautiful mathematics comes from "modular forms": a combination of symmetry and the complex plane. We ask what functions exist such that a large number of symmetry transformations leaves the function essentially unchanged. For functions on the real line, we can ask which functions $f(x)$ are unchanged if we demand $f(x) = f(x+1)$, and the answer is every periodic function, everything that repeats itself infinitely like $\sin(2\pi x)$ and $\cos(2\pi x)$. Its richer in the complex plane, and there is a particularly interesting symmetry group – SL(2,Z) – that generates the formulae quoted above, and was Ramanujan's insight after the first calculation. In Figure 3-3 is an image inspired by the paper of David Lowry-Duda[1], and my picture of SL(2,Z) transformations. An interactive version is available[2].

It has been nearly fifty years now that I have been intrigued by this number, that I first saw in a paper of Ramanujan from 1914:

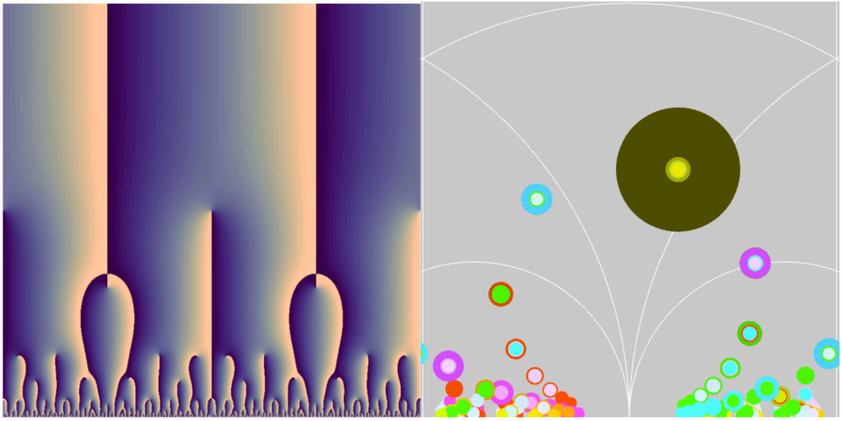
$$e^{\pi\sqrt{163}} = 262537412640768743\,.\,9999999999992$$

It is easy to see that it is close to a whole number, in fact there are 12 nines after the decimal point. During lockdown in 2020, I made a serious effort with the mathematics underlying this curious coincidence. I got quite a way with the modular forms part, but there's another part that I didn't understand, concerning *imaginary quadratic forms* and Heegner numbers. So it remains a mystery to me.

---

[1] https://arxiv.org/abs/2002.05234

[2] https://roywilliams.github.io/play/js/sl2z/

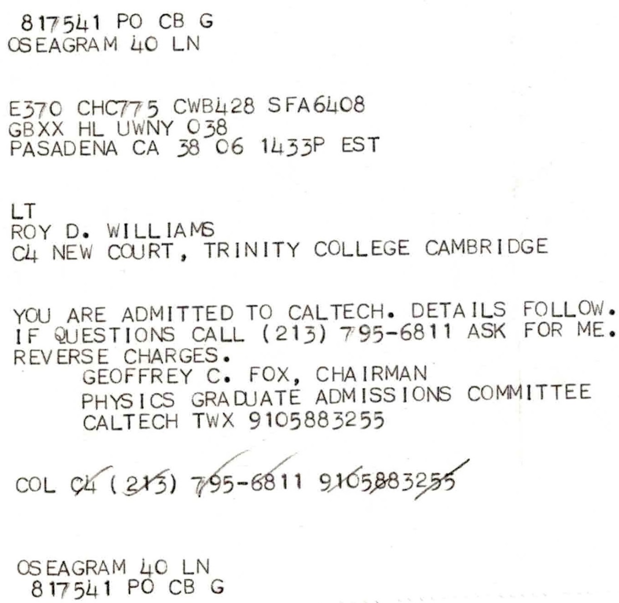


3-3 Left: Phase plot of a modular form function. The symmetry group that generates it is evident from the self-similarity that can be seen at the bottom of the image. Right: Repeated application of SL(2,Z) transformations starting with the largest disk create the smaller disks; values of the function at the coloured points are related.

4. California

Golden State

As my three years at Cambridge approached its end, it was time to go somewhere else, preferably a place with science and computers. There was a letter in the university careers office from one Geoffrey Fox, representing the California Institute of Technology, inviting applications. At that time, I was reading Hunter S. Thompson's *Fear and Loathing in Las Vegas*, with stories of drug-crazed binges and driving through Nevada at high speed while enjoying a blow job. Or something like that, but that was my impression at the time. I applied. Took an exam (GRE) in London with other hopefuls. Several months later, there was a knock at the door of my room at Cambridge, and an excited college employee brought a thin envelope with a blue border, exclaiming "It's a telegram! A telegram from America!". This was 1976, when these things were unusual. I ripped it open and there it was "You are admitted to Caltech" (figure 4-1).


817541 PO CB G
OSEAGRAM 40 LN

E370 CHC775 CWB428 SFA6408
GBXX HL UWNY 038
PASADENA CA 38 06 1433P EST

LT
ROY D. WILLIAMS
C4 NEW COURT, TRINITY COLLEGE CAMBRIDGE

YOU ARE ADMITTED TO CALTECH. DETAILS FOLLOW.
IF QUESTIONS CALL (213) 795-6811 ASK FOR ME.
REVERSE CHARGES.
GEOFFREY C. FOX, CHAIRMAN
PHYSICS GRADUATE ADMISSIONS COMMITTEE
CALTECH TWX 9105883255

COL C4 (213) 795-6811 9105883255

OSEAGRAM 40 LN
817541 PO CB G

4-1 The telegram brought in dramatically by a college employee, handed over breathlessly.

Graduation at Cambridge was lovely but also sad. I realised that I wouldn't see many of my college friends again, and that we were all leaving the protected harbour out into the storms of the real world. What I had not understood back then is the importance of these relationships when making progress in that real world. My life was destined to be "permanently" moving across the Atlantic not once, but four times, each time giving up valuable friends and work connections. Anyway, there were strawberries and cream in the majestic Neviles's Court at Trinity, for once being allowed to walk on the grass. One of the Fellows told me there's an airport bus from Los Angeles airport all the way to Caltech.

I was due to fly to California in just a week, but there had been a huge crash of a DC10 airplane, and the whole fleet was grounded, so places were short. I spent 24 hours standing in line for a "standby" ticket, to save money. Finally I was on the plane for the 11-hour flight, with a huge bag of vinyl records under the seat in front of me, so even less legroom. I took the airport bus to Pasadena, and through luck got talking with a woman who told me the bus doesn't go to Caltech but somewhere else. Her boyfriend picked her up in a car and they were kind enough to take me to the campus and drop me off. It was a beautiful June evening, the sun setting in the west, I had been awake for about 36 hours, and I knew nobody and had no place to stay. But I saw some people – it was a dorm and they were watching TV – so I just walked in, ready to collapse. Somebody was kind to me, booked me into a room at the Caltech Athenaeum (the faculty club), saved me. That somebody was Stephen Trentalange, who has been my good friend for many decades since.

The next day I woke early, jetlagged, and walked out past the beautifully kept lawns of Caltech's Olive Walk, when suddenly there was an explosion! But then I realised it was the automatic sprinklers – which I had never seen. I kept going and ended up at a "coffee shop", where I ordered steak and eggs, because I remembered that's what the astronauts would have before heading to the Moon. From the time that plate was in front of me, things looked up.

I was confused by the selection of milk in the California supermarket. Low fat, half-and-half, soy milk. All they had in Britain at that time was just one thing: milk. And the churches! In Britain it was

Protestant, Catholic, or Jewish (I had not seen a mosque). But in Pasadena there was the Four-Square Church, the Baptist, the Science of Spirituality, the Vision Christian Fellowship, the Scientologists, and so on, as if every group of people decided to have their own church, instead of following Henry VIII, the Pope, or Moses, as did all the religious people I knew.

While there was far too much choice in some ways, there were a few food problems: beer, cheese, bread, and coffee. While Britain had moved away from mass-market beer in the 1970s, pushed by the Campaign for Real Ale, the craft brewing movement had not yet taken off in the USA, and there was a lot of Bud and Miller and Coors. While a few places in the USA could supply European style cheese, it is still considered an extra, not a main course – and always vacuum-sealed in plastic, so there is no delicious aroma. I think Americans are a generally a bit grossed out by the idea of adding mould to improve flavour. The bread tended to be just sliced square loaves in plastic bags, no jolly baker pulling out crusty baguettes from a hot oven. Coffee was made in a drip machine, perfectly OK in its way as long as it was strong enough and not burnt. Of course all these problems disappeared over the next few years, and now in the USA you can get the most wonderful beer, cheese, bread, and coffee.

Southern California

The climate and terrain of Southern California was marvellous. Always sunny but not too hot; the sun shone even in winter. In May and June, mornings were foggy, with the sun appearing in the afternoon. When it rained, it really rained, then it suddenly stopped; no more drizzly English days with a sky like pewter. A bit boring, the endless blue sky, but not bad! Instead of twisty English roads, there were fast highways that could quickly get you to the beach, to the mountains, to the desert. You can surf at the beach and ski in the mountains on the same day! The Caltech Athenaeum is a splendid and graceful place, like a Cambridge college: restaurant with oil paintings of the famous scientists, lodging rooms, lovely gardens, a palatial room with armchairs and newspapers. If you want to impress a visitor, put them up at the Athenaeum. Many of its members were from the tony suburb of San Marino next to Pasadena, with enormous homes like royal palaces, surrounded by vast gardens.

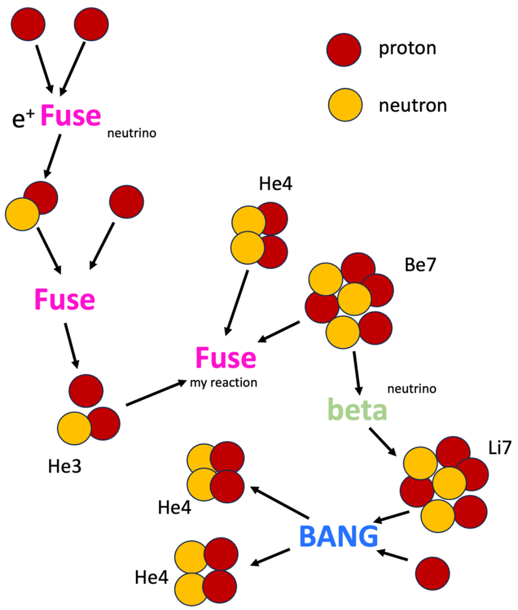
The only people you would see, walking around San Marino, were the immigrant gardeners working tirelessly.

I had spent years camping and hiking in Britain, but it was so much bigger and brighter in California! Instead of the smell of woollen socks that have been wet for days, there is a smell of pine resin and gentle zephyrs of desert air. Instead of the mountain summit being smoothed off by glaciers, in California the plates are still active: great sharp mountains with pocket glaciers. In Britain the hike (or walk as it's called) would go through land owned by somebody, with public right of way, and there would be a pub for lunch. But in the US it was all much more gung-ho: heading into the wilderness with a tent and dried food. No herds of cows and sheep, rather distant deer and taking precautions against bear attack when camping.

The deserts were astonishing also to a boy from London: great wide vistas and purple sunsets as the glistening coach of night draws in. Slot canyons, where the rainfall is a combination of almost never and extremely fierce; you feel like you're in a narrow passage but full of light, and you imagine the force of water that has carved the canyon. Great sand dunes, where you can run and fall and roll like swimming in warm sand. Often the desert landscape is full of previous human efforts: a century-old mine works, lime kilns, or even a ghost town. I have always loved Death Valley, the huge dry lake that is below sea-level, and known for extreme temperatures, the sound of wild burros, the brilliance of the light, the extreme adaptation of species, the silence. I'll take Death Valley over Las Vegas any day!

Learning How the Sun Works

I landed in the Kellogg Lab at Caltech, a place founded by the cornflake king in the 1930's to investigate the healthful properties of radioactivity and radiation treatment. But Kellogg quickly became a world-class research lab of nuclear physics, and was made particularly famous by its association with discovery of how elements are created in stars, the new field of "nuclear astrophysics". Willie Fowler was part of Kellogg, and together with Fred Hoyle and the Burbidges – a married couple – elucidated how most of the chemical elements could be synthesised in stars like our Sun. The work continued 20 years after the seminal paper of 1957, usually called the B2FH paper: understanding details and checking the predictions of the model. There are three “chains” that drive the solar furnace, but I worked on the PP2 branch shown in Figure 4-2. Reaction rates can be measured in the lab, with an accelerator such as Kellogg had in its basement: high vacuum plus powerful magnets accelerate protons of helium nuclei to high energy and collide the beam with a target. Once the impact energy is high enough (maybe 500 keV), then the rate of the reaction is high enough that it can be measured.



4-2 Nuclear processes that drive the Sun. Converting raw Hydrogen (protons, top) into heavier elements, with the release of energy and neutrinos. The reaction I studied is in the centre of the picture: ^{3}He + ^{4}He → ^{7}Be + energy. It has a lot of protons coming together, so needs a high temperature.

But inside the Sun, at a temperature of (only) 15 million degrees, the nuclear reaction rates are very low. While we think of the centre of the Sun as an enormously fast conversion of fuel to fire, it is actually a rather unproductive reaction because of the low temperature: less like exploding gasoline and more like the heat generation in the middle of a compost pile! The rates of the reactions are so low that we can’t measure them on Earth. And yet the enormous size of the Sun means it has an huge energy production rate. Furthermore, the centre of the Sun is very well insulated by a very thick blanket, so even a low rate of energy production builds up to great heat.

The centre of the Sun is difficult to access experimentally. You can't go there. But some of the reactions in the fusion chain emit neutrinos, the ghost particles that pass through everything with almost no reactions. Anti-neutrinos had been detected in 1956 from a nuclear reactor, but their opposites, neutrinos, had not been detected. But the conventional wisdom of the solar fusion reactions predicted a rate, and a physicist named Ray Davis and John Bahcall decided to find out. In the late 1960's an experiment was built to detect these solar neutrinos; it was an enormous tank of cleaning fluid – carbon tetrachloride – because very rarely a neutrino will interact with a chlorine atom to make an argon atom, and a detectable flash of light. But after several years of data gathering, the flux of neutrinos was only a third of what the B2FH theory predicted: this was called the "Solar Neutrino Problem", and many of us at Kellogg were investigating. Hence the flurry of experiments and calculations to get all the reaction rates and the temperatures and pressures right, to check that the prediction was correct.

One of those details was about the rate of a particular nuclear reaction, when helium-3 and helium-4 fuse to make beryllium-7. My first scientific paper was a calculation of the rate of this reaction at the temperature of the centre of the Sun, and this was the first paper I wrote with my advisor, Steve Koonin: "*Direct capture cross sections at low energy*". Perhaps the extrapolation down in energy of this particular reaction (Figure 4-2), was incorrect? But in the end, it turned out that there was no error in the reaction rates of the chains, but rather new physics explained the discrepancy. It was already known that there are three types of neutrino: electron, muon, and tau neutrinos, and it was also known that the solar nuclear reactions emitted electron neutrinos, and that the chlorine in the tank of cleaning fluid reacted with those electron neutrinos. But in 1998, it was discovered that neutrinos oscillate between the three types on their way from the Sun to the Earth, which neatly explains the factor three reduction of the measured rate.

Supernova Shock Breakout

There was another puzzle that was doing the rounds in the early 1980's: how to simulate on a computer the supernova explosion of a massive star, something directly predicted by the nuclear reactions that take place in our Sun. The evolution path of a star is influenced strongly by the mass, because that controls the pressure and temperature achieved in the centre, which is where the nuclear reactions start. A low mass star, less than 10% of the solar mass, does not get hot enough to fuse hydrogen nuclei, but may be able to fuse deuterium (an isotope of hydrogen) or lithium, should there be any in the initial mix. Between this limit and about 2 solar masses, a long-lived star will be produced, happily fusing hydrogen to helium over a period of many billions of years and eventually peter out to a white dwarf. But for large stars, more than 20 solar masses, the temperatures and pressures at the core burn hydrogen at a high rate, followed by more exotic reactions producing heavier and heavier nuclei, until there is an onion-like structure: with an iron core, surrounded by a silicon layer, then oxygen, neon, carbon, helium, and the outer layer of hydrogen. Iron is the limit of the nuclear cycle, because making a heavier element *costs* energy, rather than releasing it. So once there is a core of iron vapour, the reactions can go no further, the energy source stops, and the immense gravity of the outer layers is no longer held back. Suddenly the millions of years of nuclear reactions ends in a fraction of a second, and there is a violent collapse. All that matter piles up in the centre of the star, then bounces back, throwing out matter and energy in a titanic explosion that can be seen on the other side of the Universe: a core-collapse supernova.

People started using computers to model supernovae in the 1970's. At first they used a one-dimensional model, where conditions were stored in the computer for every shell of the onion-like structure; at this radius from the centre the temperature and pressure, as well as the abundances of many different elements and isotopes. The knowledge of nuclear reaction rates would be in the program, so that the entire stellar evolution could be predicted. The model would predict the millions of years of evolution quite well, but then at the very end things would fail: the program did not predict the bounce-back of the core and its outgoing shock front. Models improved as computers improved – and of course the nuclear physics improved, but still no spectacular supernova explosion, just a sort of big nuclear bonfire. One part of the puzzle is the nature of "nuclear pasta" – see below for more about this.

It turned out that the one-dimensional modelling was simply insufficient: a full three-dimensional multiphysics simulation is required to capture the fluid instabilities and turbulence that drive the power if the shock breakout. It was Moore's Law that modelled the doubling of computer speed and memory every two years, and this, in the end has allowed the proper modelling of supernovae.

Hoyle Resonance as Proof of God

Creation of all the elements of the periodic table relies on the creation of carbon from three helium nuclei. The route to heavier elements is through carbon production. And this is a narrow pathway requiring the simultaneous presence of three helium nuclei. Imagine how difficult it would be for humans if making a baby required not just two people loving each other, but three! How unlikely for three to be together in the same place, and for that triple coincidence to stay together long enough. It is the Hoyle resonance that provides this magical, narrow pathway to complexity. The religious might say the fortuitous existence of the Hoyle resonance proves the intervention of God, who wanted to build a complex Universe suitable for Life instead of a dull Universe.

The chemical elements are the building blocks of everything in the Universe. There are 92 naturally occurring elements, the Lego blocks from which out world is made. Each nucleus of an element consists of a specific number of protons and neutrons. The basis of life on Earth is carbon, with six protons and six neutrons, or we can forget the distinction and call it 12 nucleons. Nitrogen is 14 nucleons, Oxygen 16, Iron has 56, and Gold has 197 nucleons. Glossing over some subtleties, we can say that each element is characterised by a specific number of nucleons in its nucleus.

Most of the Universe is hydrogen, which is a single nucleon, and helium has four. Like a party with a lot of introverts, nucelons stay apart until something brings them together and only then do they find they like being in a group. That something (stellar nucleosynthesis) was elucidated in the 1950's: the process where light nuclei are combined to make much of the variety of elements that are prevalent on Earth. Nucleons want to stick together – the process of nuclear fusion – and all that is needed is to bring them close enough with some heat and pressure. In this way a star is a self-sustaining reactor, with heat being generated by fusion because of the pressure in its centre.

There is a problem on the way, because while helium-4 is easy to make from hydrogen, the elements with five to 11 nucleons are not energetically favoured, so the step-by-step process of adding nucleons might easily stop there, meaning a Universe with no carbon, or anything heavier, a place with stars but without the rich menu of elements that provide building blocks for planets, for life, for civilisation. Fortunately, there is a magic gateway! The so-called Hoyle resonance is a semi-stable, excited configuration of three helium nuclei that can last long enough for a few of these triples to drop down to form a carbon nucleus. This is the magic gateway.

Heavy Elements

Bigger stars can support higher temperatures and pressures, and heavier elements can be built, all the way up to Iron with 56 nucleons. Like water falling downhill, it is reasonable that these elements will form, given the energetic advantage, the propensity of nucleons to stick together. However, the process of accreting more nucleons is disfavoured energetically for nuclei heavier than Iron. The curve of binding energy has turned over, like expecting water to flow uphill.

But there is another nuclear process that can make heavy elements, the so-called r-process, which needs Iron and heavier nuclei to be bombarded by a massive storm of neutrons, that stick on and convert to protons very rapidly, even against the energetics. Such a massive storm occurs in an explosion of a massive star called a supernova, and the storm is powerful enough to create elements with 100 or more nucleons. However, it is thought that even stronger neutron flux is needed to create the upper end of the periodic table – Gold, Platinum, Bismuth, Uranium etc, with over 200 nucleons – and the only known way to get such a high flux is the collision of two ultracompact neutron stars: the dead cores of massive stars after they have exploded as supernovae.

These very exotic events are extremely rare and occur maybe once in 1000 years in any given galaxy; see Chapter 9.

The formation of the elements is from multiple mechanisms, just as geology has many processes. Geologically, there is erosion, flow, and deposition by water, by ice, and by air, and there are barriers to these processes. Astrophysics has the condensation of hydrogen clouds, with temperature and pressure bringing nucleons together into energetically favourable nuclei, and there is the blockage solved by the Hoyle resonance. There are violent processes that push uphill even against the curve of binding energy; like the slot canyons of desert regions, where the landscape is crucially changed and moulded by very rare, very violent storms.

A Cosmic Chronometer

While on the subject of nuclear astrophysics, I did some work on a specific isotope Rhenium-187. It’s an extraordinary nucleus, hovering between stable and radioactive: its half-life is 43 billion years, three times the life of the Universe since the Big Bang, decaying to Osmium. Just as carbon dating can be used to date organic materials, so Re-Os dating can be used for the universe as a whole. But one of the difficulties is measuring the half-life itself, since the rate is so low. Part of the reasons the rate is so low is that the emitted electron has very low energy – so not much to drive the reaction. If the nucleus is already surrounded by electrons – an atom of normal matter – then there is no place for the new electron, so it must jump all the way out. But if its inside a star, all the electrons stripped, there may be an easier place for it to get to, and thus a shorter half-life. So my contribution was to think about the differences in half-life of this *cosmic chronometer,* depending on whether it is inside a star or not.

Crashing Out

The first weeks in California, I met Dave McNary. Dave and some friends had a big house divided between various single tenants, and I became one of them. We called it Hill House, and there were parties and japes and weed, as I struggled over DeShalit and Feshbach’s magisterial *Theoretical Nuclear Physics*. I also made friends with Jeremy Bentham and Rob Pike, and through them I met some girls! You have to know that since age 11 I had attended a boy’s grammar school, and a Cambridge college that was boys only, so it always took an extra effort to meet women. Jamie Primm looked after a group of Barn Owls at Caltech, that were used in studies of binocular vision, and she had her own pet owl, named Quasar. One weekend, the owl was staying with me at Hill House, and my fellow lodger, Rick, was quietly smoking in the living room, when Quasar swooped down the stairs, circled the living room and flew back up the stairs. All completely silent, as owls are, leaving behind nothing but disturbed air. Rick looked up and said “Was that real or was it the drugs?”

Jeremy and I both got motorbikes, so we could explore California. Such a feeling of freedom, roaring along the road, wind in my hair. This was the realisation of my dream based on *Fear and Loathing in Las Vegas*. All went well until one early evening, driving the bike back to Hill House after doing poorly on an exam. I had the green light going straight, and a Cadillac was waiting in the left-turn lane. But he didn’t see the bike, made the turn, and I flew over the top. I spent a month in the hospital. It was the first time I had an inkling that family is important: it was great to see Dave and Jeremy, but it took a week or so for my parents to fly over and comfort the boy with the leg in traction and the broken jaw. There was a question of whether the leg would recover properly. Would I be able to walk? You can imagine my joy eight months later, after hiking and camping, and arriving at the summit of Mt Whitney at 14,500 feet. There is no joy like the condemned person given their freedom! Figure 4-3 shows me in triumph, having completing a gruelling mountain backpacking expedition just eight months after being told I may never walk without a limp.

Feynman and Others

I took a class from Richard Feynman on applied quantum mechanics in 1980; his lectures were so clear, and made the physics so obvious, that we all felt like masters of the universe. And yet when it came to the problem sets, it was not quite so easy, and even with several students together we had a difficult time. I remember the first time I used a distant computer to solve a problem: something about the wave-functions of the electrons of a hydrogen molecule. After several pages of error-prone crunching formulae on paper, we turned to a symbolic programming system called MACSYMA that was running on a machine at

MIT, 3,000 miles away. Symbolic computing deals with algebra rather than numbers, so you can say

`expand((a*x + b)^2),` and it will respond with

$a^2x^2 + 2abx + b^2$

4-3 On the top of Mount Whitney (14,500 ft), just 8 months after being told I may never walk without a limp.

The connection was made by telephone crammed into an "acoustic coupler" that allowed our terminal to communicate with the remote computer. You had to know the phone number, listen until it connected, then jam the handset into a black foam cradle with two big holes, one for the speaking end of the phone and the other for the listening end. Then you log in and off you go. There would be a mechanical typewriter jumping up and down, typing lines of text from the remote machine, from which you type replies at your end. Anyone who has used the "terminal" utility, or a "command-line interface" will recognise this interaction. By the mid-1980s there were cathode ray displays, also known as a visual display terminal, so no more mechanical typewriter, and no more of the distinctive smell of all that paper and hot ink. How could we imagine, back then, that we would use remote computers “in the cloud”, not even knowing where the server is physically located? And how could we know that children would be using distant computers from a device the size of a hand?

A few times I was at the same lunch table as Feynman, and I remember some things he said. One was “Whenever there is a factor 100 change, then there is a qualitative change”. When you have 100 times the cars or the population density or the number of chickens, then you need new ways of doing things. When distance increases from a mile to a hundred miles, it’s not just the journey time that changes, but also the transportation method. Emails and parked cars and self-driving cars all come to mind. But especially I am reminded of this when a climate denier shows the history of global temperature and says it has all happened before; but the variations were 100 or 10,000 times slower back then! The qualitative difference in time scale means the ecology cannot adapt, it means a huge and quantitative difference in the effect of the warming.

Another thing Feynman stressed to us was the importance of play in the life of a productive scientist, that you just try this and try that according to how much it appeals to you, and maybe some new thought appears, some new solution to an old problem. He also said it was important to *put it down* for a while, to give it time, to walk or play the bongos or socialise. This way the answer appears like a flash while in the shower.

I like the analogy Feynman drew when explaining physics: the underlying principle may be simple, but calculating quantitative results may have a lot of tricks and shortcuts. Understanding the principles is actually a lot easier than calculating quantitative results. The example he used is addition and subtraction of numbers. In principle, all you need is the ability to count. Suppose we start with a jar of beans: to add 15 and 27, you would count out 15 beans, and then 27 beans, join the two piles, then count how many beans total. Yet the actual process of addition is much more complicated: the tabular layout of hundreds, tens, and units; carry the one; memorise your addition table (5+7 is 2 carry 1) then the next column, (don't forget the carry digit). Make sure you do the columns from right to left. Feynman recounted this before talking about path integrals in quantum mechanics. The principle is understandable: that all paths count, not just the classical trajectory, the "optimal" path. But calculating these path integrals, represented as Feynman diagrams, is very difficult. See the section below on Free Will.

Stephen Wolfram

Stephen Wolfram was another contemporary at Caltech, who has become a great expounder of mathematical and scientific ideas, and built the famous *Mathematica* software. I remember his apartment near the campus: a large comfortable chair in the middle of the room, surrounded by a litter of literature and pizza boxes, like planets orbiting the Sun. He was an *enfant terrible* of Eton, a very exclusive British school, publishing papers about nuclear physics at 16 as I recall. He was interested in cellular automata: a line of binary bits with a rule to evolve the string to a different state. A famous example in two dimensions is John Conway's "*Game of Life*", with intricate patterns, and pattern-making patterns. That fascination with cellular automata eventually became a giant book purporting to explain life, the universe, and everything. Wolfram had made a computer system called SMP – Symbolic Manipulation System. It was like MACSYMA, which eventually became Mathematica. But I remember the original name for SMP comes from the "School Mathematics Project" which both he and I suffered. SMP was a sort of convulsion in the teaching of mathematics in 1960's Britain: it had its own imaginary computer language (Simpol), that ran on an imaginary computer called Simon. It was all that stuff about emphasising investigation over rote-learning. So Wolfram's SMP was a tongue-in-cheek reference to that nonconformist way of teaching mathematics. Wolfram was the nearest I have known to Sheldon in the TV show *Big Bang Theory*: completely convinced of his own brilliance and having difficulty working in a team. But unlike many people who just think they are very clever, Wolfram actually is very clever.

Unix: A Proper Operating System

Finally, I will mention Rob Pike, a housemate in Pasadena in 1980, who went on to become a principal at Google. Rob showed me a new and much more efficient way to use a computer, so much better than the old card decks and their singular emphasis on batch-mode computing: present the code then the computer runs it then look at the result. It was the Unix operating system, something that has become a mainstay of my life for many decades. In addition to batch computation, Unix has a command-line style, a sort of conversation where the human and the computer alternate. Once you know a few commands, and how to combine them, it is very powerful. For example, the `cat` command outputs the contents of a file that has a list of names:

```
cat list_of_names
```

And if you only want those people named "John", you can use `grep` to filter the output. The pipe symbol "|" means to take the output of what is before as the input of what is after:

```
cat list_of_names | grep John
```

Then you can sort the result by piping again to sort:

```
cat list_of_names | grep John | sort
```

So from an initial list of

```
Sue Barker
John Jaworski
Roy Smith
John Aardvark
John Salmon
Diane Heron
```

You would get the sorted version of those whose name has "John":

```
John Aardvark
John Jaworski
John Salmon
```

In addition to the pipe symbol, there are symbols to redirect the output of the command into a file or take the input from a file. There are command-line words that modify the behaviour of subsequent words, for example `time` means to execute the rest of the line, but provide timings at the end:

```
time cat list_of_names | grep John | sort
```

Another innovation of Unix is the hierarchical system of data storage, with directories that can contain files or other directories, together with sophisticated access permissions. Unix is also a *portable* operating system, meaning it is implemented not directly on the idiosyncrasies of a specific computer, but with a portable language called C. Thus Unix can be run on a new machine as soon as the C language can be run – a much easier proposition. Unix is a multi-processing system, and a multi-user system, so that it can do many things at once, with multiple processes who do not interfere with each other. Of course it is not really doing many things at once, but rather switching from one process to another so quickly and seamlessly that nobody notices.

Unix was built in the 1970s at Bell Labs – the research arm of the US telephone monopoly – and extended and grown in numerous ways; in fact the machines we use 50 years later are based on Unix: the Android, the Apple OSX and iOS and Linux are all based on that early work. The philosophy that has worked so well is to make simple, well-defined tools, together with sophisticated ways of combining them. So while I was carrying heavy boxes of cards up to London in 1973, to be run on a machine built in 1960, Ken Thomson and Dennis Ritchie were inventing the future.

Also in the early 1970's, email was invented by Ray Tomlinson. Already it was possible to send messages between users on the same machine, but Tomlinson introduces the famous @ usage to separate a user from the machine to which the message should go. That symbol had been used for a hundred years in commerce, where you would write "7 widgets @ $5 each" for example. It was Rob Pike who showed me email, and how it could send messages from one end of the building to the other. It did not seem impressive: why not, I thought, just write it on a piece of paper and walk down the corridor, or even put it in the post? But of course the usage of email grew and grew, and messaging has completely displaced the envelopes and stamps that I grew up with.

Nuclear Pasta

There are many collective phenomena in physics. Begin with an understanding of a few individual entities or particles, but when the numbers are large, entirely new behaviours appear. Road traffic, for example looks like individual vehicles on a road when the density is low, but a crowded freeway is a different matter altogether. We are driving fast but at low density, then in front we see a wall of red brake lights, and suddenly the traffic bunches up to high density and the speed goes down. After a while, the density gradually decreases and the speed increases until we are a high speed again. This sudden deceleration and gradual acceleration also occurs with compressible fluid flow, characterised by shock fronts – see chapter 6.

So this is where the nuclear pasta comes in; modelling the physics of the high-density matter during the few milliseconds between collapse and bounce. Atomic nuclei are sometimes modelled as droplets of nuclear matter; like a water droplet a nucleus is roughly spherical, and there is a surface tension that keeps it together. But the nuclear matter also has a strong electric charge from its protons, that pushes it apart. The spherical nuclei tend to elongate to ellipsoids, and at higher density to long cylindrical structures (nuclear spaghetti). As density increases, the matter configures itself to flat sheets (nuclear lasagne), then to pure nuclear matter but with long cylindrical bubbles (negative nuclear spaghetti), and squeezing harder still, to pure matter containing spherical bubbles. As pressure increases further, there is no more space, no more bubbles, and the density stops increasing. Just like water mixed with air: the gaseous component is compressible, but the liquid is not.

We can think of tourists in a popular place like Venice: early in the morning the density is low, and they move as individuals or family groups. Nuclear matter is the same: protons and neutrons individually or as small nuclei like helium or carbon. But when the density increases, with larger groups of tourists, they move in a long line like a strand of spaghetti. Eventually the density is so high that the tourists are packed together like sardines, with a rare "bubble" of no tourists.

Nuclear pasta is important for understanding supernovae. In the last microseconds of the core-collapse and bounce, the theory provides the equation of state – the relationship between pressure, density, and temperature. Nuclear pasta was one of my first papers in 1984, the energetics of short-range surface tension forces, in conflict with long-range electrostatic forces.

Free Will and Determinism

My PhD thesis at Caltech was about "semiclassical" methods in quantum mechanics. If you imagine a marble in a bowl, the classical and quantum views are quite different, and the semiclassical approach is an attempt to reconcile these. Classically, the marble rolls around and around, and if there is no friction, it never stops. It can move in a periodic orbit, repeating the same trajectory over and over, like a planet around the Sun. Or it can move in a chaotic way, with no repeats. And of course it could be at the lowest point, completely still. The quantum picture, rather, is one of energy levels, each with a very specific energy level and a probability distribution of where the marble might be found if observed. In quantum mechanics, the ball can be at energy level E_0 or E_1 or E_2, etc, but not anything in between. There is no lowest point, rather a lowest energy level, which like the other states, has a probability distribution. Classically, in contrast, the ball can have any energy, and it can be perfectly still at the lowest point of the bowl.

So what can we learn about the quantum energy levels and probability distributions from a detailed knowledge of the classical dynamics? It turned out that periodic orbits – where the same motion is repeated – play a crucial role in predicting quantum behaviour. Because quantum mechanics is predicted by considering *all possible paths*, not just those that the marble would actually follow, but all of them. Let me try to give an illustration of this with the refraction of light.

Refraction as Optimisation

Light travels more slowly in a medium – let us imagine that medium is glass, because we will be talking of refraction of light. The "refractive index", written with the symbol "n", defines that slower speed. If the speed of light outside the medium is c, then in the glass it might be c/n. If n is 1.5, this means the light travels at only 67% of c. The reason we use glass in microscopes, telescopes, cameras, and spectacles is that the interface between air and glass changes the direction of the light-ray, so that it can be focused.

Now imagine that you are walking between two points as shown in Figure 4-4 top-left, one on a smooth paved surface, where you walk at fast speed, but the other point on a difficult surface like long grass or gravel, a slow walk at 67% of your normal walking speed. What is the fastest route between these points? The answer is not the straight line, but rather to shorten the section of slow walk in exchange for a longer section with the fast walk. When analysed in detail, the optimum path satisfies

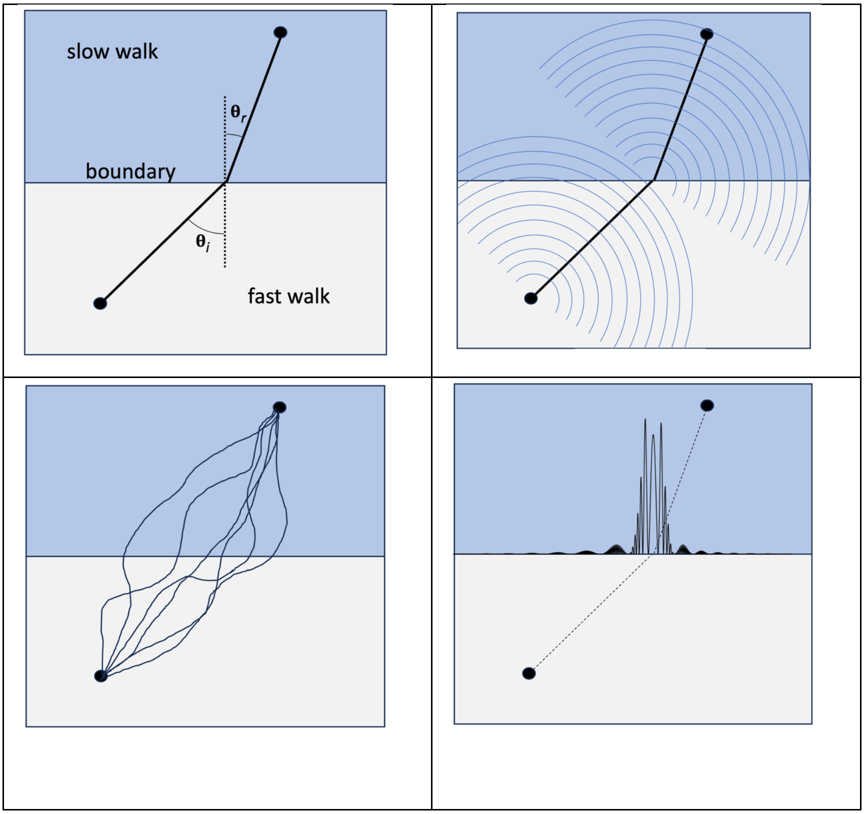
$$\sin\theta_i = n \sin\theta_r$$

Strangely enough, this is precisely the same equation (Snell's Law) that controls how a light beam is refracted through glass. The optimal path across smooth and rough terrain is the same as the path of a light ray refracted at the boundary. This minimisation was stated by Fermat in 1600 as the “principle of least time”.

What is strange about this is philosophical: we think of the light starting at its source, maybe a lightbulb, and moving – at the speed of light – towards the glass, and then the direction gets changed, then it continues on. But the other way to see it is an optimisation problem, like choosing the fastest path between two points. It brings up deep questions of free will versus determinism. Did the light-ray walk through its life, responding to circumstances, or was the whole thing planned in advance, birth to death? We have a dichotomy between free-will and determinism.

There is a story about this paradox *Story of Your Life* by Ted Chiang, made into a movie called *Arrival*, starring Amy Adams as the brilliant linguist. The alien civilisation, the heptapods, communicates with all the thoughts at the same time, with free word order and a two-dimensional pictorial system. This is in contrast to humans, with their linear sentences that have words and glyphs in a definite order. The heptapods experience all events at once instead of in time order. Fermat’s principle of least time enters the picture, and the linguist starts seeing the future and the present all together, including seeing her not-yet-born daughter.

In 1678, the wave nature of light had been recognised, and the *Huygens-Fresnel principle* (Figure 4-4 top-right), saw every point reached by the light is its own source of light. We see the light waves striking the boundary which itself is a source of light. This explains both the model of outgoing rays of light, as well as the diffraction effects caused by its wave nature.



4-4 Top-left: The fastest path from a point on a smooth, fast surface to a difficult, slow surface is not a straight line.

Top-right: The Huygens principle models the question so that each point of the medium re-radiates the light.

Lower-left: A quantum-mechanics approach sees photon particles that move on every possible path, carrying a phase, and these phases add to produce the amplitude (probability) of the event.

Lower-right: Here we see a partial sum of the phases of the multiple paths, considering only paths that are straight lines on each side of the boundary. The minimal path is picked out because others cancel out through interference effects.

The Huygens-Fresnel theory reigned supreme until the advent of quantum mechanics in the 20th century. The path-integral formulation of Feynman and sees optics as propagation of photons – wave-particle duality (Figure 4-4 lower-left). While classical dynamics and the propagation of light rays sees the photon moving only on one path – i.e. the path of least time described above – the path integral method considers all the paths the photon could possibly take, and adds them up. In many ways similar to the Huygens-Fresnel theory, it has no actual waves in the description, only the multitude of classical paths.

But each path carries a phase with it, and the trajectories can add in way that can be constructive or destructive. As shown in Figure 4-4 lower-right, the bundle of paths that constructively interfere, and are therefore the most important, are the paths associated with minimum time. The optimisation of the path is not from some pre-ordained plan, but because all paths have been tried, and those that contribute most are those of shortest time.

5. Oxford

Neutron Scattering

Several times in my life, I have moved “permanently” between the UK and US. I was born in London with a UK passport, and at the age of 19 discovered that my mother’s US birthplace entitled me also to a US passport. So I went to the California Institute of Technology (Caltech) after my undergraduate education, to become a graduate student and teaching assistant.

It was 1984, I had my PhD and a year of postdoc mucking about with quantum field theories – can't really remember – and was looking for where to go next. There was something with Schlumberger, an oil service company, that involved lots of science, lots of computing, building conclusions from data. But even back then I knew there was something unsavoury about fossil fuels, and did not want to spend my career with it. There was another job though: being a fellow at an Oxford college. I remembered the days of being an undergraduate at Cambridge, the gracious buildings and deep history and traditions, it just seemed to the young me to be so gracious and classy. I was offered the fellowship, hooray, so I felt mighty pleased with myself. And my mother and father too, they were happy to have their boy back in the UK again. I married Barbara Lamprecht in June 1984 and we moved “permanently” to the UK. The position was very much two parts in two locations, so let’s describe those in order.

St Catherine’s College, Oxford

I was a Fellow at St Catherine's College in Oxford, a modernist college built in 1960, so not the ancient architecture, but certainly a place upholding all the Oxford traditions. I remembered eating in a communal hall with undergraduates at long tables, then a step up to the “high table”, where the fellows ate, and sitting in high backed chairs presumably having sophisticated conversations. We would each wear a long, black academic gown, and enter in some sort of pecking order (can't remember the details), then a big gong is sounded for silence, followed by a Latin grace, after which dinner could commence. The grace before dinner was shortened to "*Benedictus Benedicat*" (May the Blessed One give a blessing), and after dinner would be a dismissal, also in Latin

"*Benedicto Benedicatur*" (Let a Blessing be given by the Blessed One). One special thing about the architecture of St Catherine's was that it wasn't just the buildings designed by the famous architect Arne Jacobsen, but he also designed the furniture to be put in the buildings, and even the cutlery to be used at the tables. In particular, the soup spoon had a bowl offset from the stem. Giving rise to the astonishing possibility of a *left-handed soup spoon!* See Figure 5-1. Whenever there was a guest, I would prompt them to ask for left-handed spoons, no matter their actual handedness.

Barbara and I were found lodging in the "punt house", a two-story garage. A punt is a long, heavy boat propelled by standing at the back with a pole 5 metres long, dropping it to the river mud, and thus pushing the boat forward. The lower floor of the punt house was punt storage, and the upper floor our living space. I remember a couple of times undergraduates walking into our bedroom looking for somebody who could unlock the punt storage. Rather embarrassing on both sides. Back then it was a modernist cube surrounded by meadows and the the ducks plashing in the river. But going back 40 years later, I found the punt house surrounded by new buildings, looking a bit forlorn.

Rutherford-Appleton Laboratory

The other part of the job was at the Rutherford Appleton Laboratory (RAL). It's a government-issue campus of cubes and cuboids, completely devoid of grace. I was doing research on materials science, with the help of a new piece of equipment for neutron scattering – the Spallation Neutron Source in Grenoble, France. There was some great maths and physics: liquid crystals, polymers, micelles and lots of other things, but I found out in those two years that I was more of a maker than a scientist. On science, there seemed to be a lot of ideas from the big-man professor who ran the department, and the duty of all the researchers was to show which were true and which were not. As far as I was concerned, it was a lot of fun writing code, hacking maths, and making visualisations of results. One part I particularly enjoyed was the daily drive between Oxford and RAL, carpooling with two excellent scientists, Mark Warner and Mike Gunn.

5-1 Cutlery designed by the architect Arne Jacobsen. Note the asymmetric soup spoon, which comes with right and left-handed versions. (Wikimedia, David Hawgood / CC BY-SA 2.0)

Barbara was in a bit of a culture shock, coming from sunny California to gloomy Oxford; the food shops closing at 5pm and the eternal overcast. But she gradually integrated, joining the nuclear protests at Greenham Common, getting a job, making friends. She joined a band of bellringers at St Giles church; the oldest and largest bell is 350 years old and weighs 700 kg.

My fellowship was joint between RAL and St Catherine's, so my job doing materials science came also with a social component, meaning free meals served by college servants for Fellows only. I was the youngest Fellow of the college, so the rule was for me to pour coffee for the other Fellows. And a strange thing about the Fellow's Common Room was the possession its chairs by absent people *"you can't sit there, that's Lord Bullock's chair!".* I think the idea was that Lord Bullock might come in at any moment and immediately need his special chair.

RAL, by contrast, was set of faceless plain boxes, its architecture was the opposite of the lovely honey-coloured stone of Oxford. As soon as I was there, I realised why the fellowship was a joint one with the

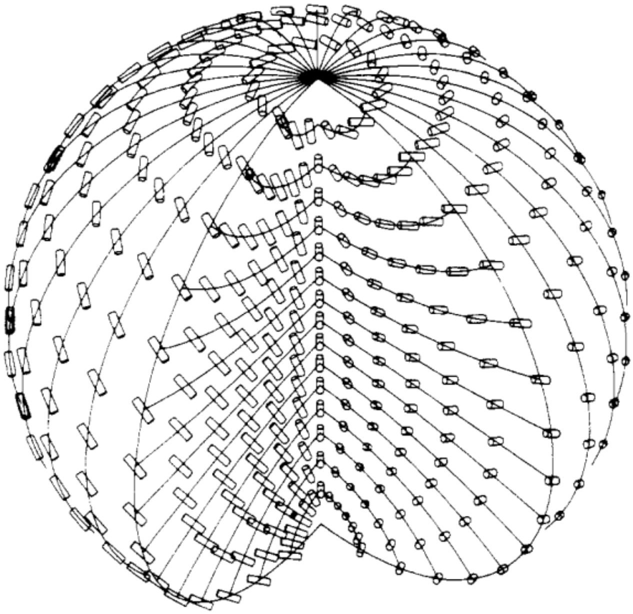
Oxford college, for St Catherine's was the cherry on top of the dull bread of RAL.

Soft matter

I studied soft materials using neutron-scattering. One such is called "micelles", or colloid, where an oily substance is surrounded by surfactant (soapy) molecules to form a droplet in water. Soap removes oil from our skin and our dirty dishes by creating billions of micelles, which are washed away by the water. The soap molecule has a water-loving (hydrophilic) end and a water hating (hydrophobic) end, so it likes to form boundary between oil and water, hence the creation of a colloid of micelles. The precise structure can be found from neutron scattering, for example if the surfactant molecule tails are ordered or disordered. The practical use would be, for example, devising mechanisms for drug delivery.

Another type of soft matter is liquid crystals, a liquid consisting of long polymer molecules. Discovered in the 1880s, these unusual substances can have multiple melting points, strange electrical properties, and they polarise and modulate light. In normal life, we all recognise the liquid crystal display (LCD) as the type of computer display that doesn't have its own light. The molecules tend to align with their neighbour molecules, so that the macroscopic liquid has this alignment; for a thin film the molecules align parallel to the film. When the molecules align in parallel, the liquid is transparent, but key point for technology is the application of an electric field, which induces a twist in the aligned molecules, making the film opaque. The electric field is created by the digital system driving the display, and the human sees the difference between transparent and opaque film. And that is the LCD display.

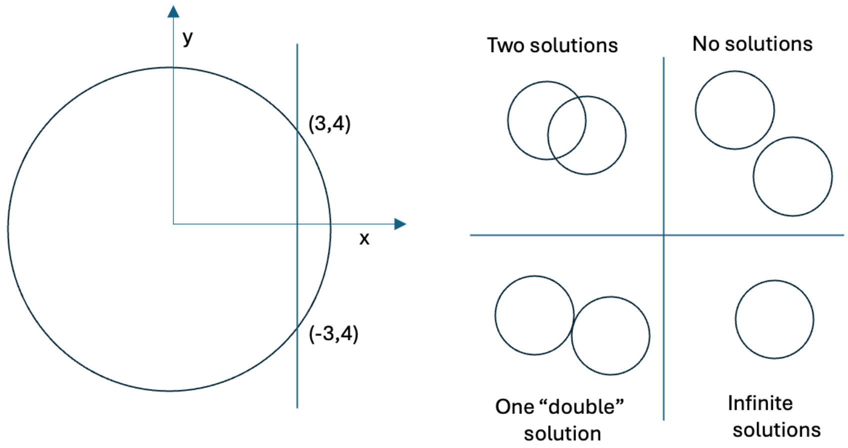
I did some nice graphics in support of liquid crystal work (Figure 5-2). The idea was to explore what a spherical droplet would look like; there are competing forces, in that the molecules want to be parallel to each other, but at the same time they must be parallel to the surface.



5-2 A perspective view of a liquid crystal droplet with a section removed to show the inside. The molecules, shown as short cylinders, are trying to be parallel to each other.

Duality

Often in mathematics and science, there are two ways to see the same thing, each providing different information. Crucially, it is often the combination of the views that yields new information. A familiar example is the idea of coordinates to describe geometry, invented by Rene Descartes in 1637. Scientists will often start a diagram in this “Cartesian” way by drawing a horizontal arrow to the right, labelled the x-axis, and a vertical arrow upward labelled the y-axis, where the lines intersect at the “origin” of the coordinates. Any point on the diagram can be expressed with two numbers: the distance along the x-axis and the distance along the y-axis.



5-3 Left: a circle with Cartesian (xy) axes, and the solution of its intersection with a straight line. Right: Intersection of circles with two, zero, one, and infinite solutions where the circles are identical.

So one way to see, for example, a circle, is just to draw it on the diagram (Figure 5-3). Suppose we say it has radius 5. By Pythagoras' theorem, the equation of the circle is

$$x^2 + y^2 = 5^2 = 25$$

Two views: one is a picture, the other is a formula. Some questions can be answered more easily with one view, and some with the other view. Let us ask for, example, which points on the circle are also on a straight line distance 4 from the centre? If we draw a line on the diagram at distance 4 from the centre, we immediately perceive that there are two solutions, because the straight line crosses the circle twice. But it is difficult to get an accurate result even with a precision pen and precision measurement. And even then, the result would not be exact, since nothing measured is exact in our imperfect physical world. But the same question is easy using the Cartesian view: we want to find where x=4 and also the above equation is true. Find y such that

$$4^2 + y^2 = 5^2$$

which leads to

$y^2 = 9$

and hence the y coordinate is 3 or -3, so the intersection points are (4,3) and (4,-3).

Another question is easier to answer geometrically: "At how many points may two circles intersect?". If we try to do it algebraically we get two coupled quadratic equations, where the centres are (p_1,q_1) and (p_2,q_2) and the radii r_1 and r_2:

$$(x-p_1)^2 + (y-q_1)^2 = r_1^2$$

$$(x-p_2)^2 + (y-q_2)^2 = r_2^2$$

Looks like a nasty problem to solve that system. But geometrically, we can picture two circles with either no intersection points, or two intersection points. (There are also edge cases are where the circles barely touch – one tangential intersection point – and where the two circles are the same, with an infinite number of intersection points.)

Fourier and Radon Transforms

Another kind of duality comes from the discovery, by Joseph Fourier in 1822, that every waveform is a sum of harmonics. We can think of Fourier analysis like the duality between a recipe for a fruit smoothie and the smoothie itself. Suppose the recipe specifies a quantity of berrys, of oats, milk, and honey. This can then be converted to a smoothie. But the key fact here is that an analysis system (i.e. a discerning taster) can convert the smoothie back to a recipe, working out the quantities from the product. In just the same way, a note from a clarinet can be converted to frequencies, harmonics, and overtones; and that same information can be converted back to the original sound.

Fourier's discovery is everywhere in modern life, because of the computer scientists Cooley and Tukey in 1965 who found the Fast Fourier Transform (FFT) that is well-fitted to the binary code of modern computers. Because of this, images, sound, and video can be greatly compressed for storage and transmission, then quickly reconstructed for listening and viewing. The JPEG and MPEG standards, in particular, are based on transforming the original data by FFT, then removing the harmonics that cannot be heard or seen, thus compressing the data; see Chapter 8.

X-ray crystallography also relies on the Fourier transform: this is how the structure of DNA was elucidated by Rosalind Franklin's brilliant technical imaging, from which the DNA model was built by Franklin, Watson, and Crick.

Closely related is the Radon transform. Instead of splitting the data into harmonics, it is represented by projecting it at different angles. An X-ray image of, for example, a hand, is a projection from the three-dimensional hand to the two-dimensional X-ray film. If a finger has a ring, the density of the ring is added to the density of the finger bone inside it. X-ray tomography is where a whole sequence of such images are taken, with the camera rotating around the body part. One might think that information is destroyed by doing this, like blurring a picture: the density of the ring and the density of the finger are added, and you can't tell which comes from where. But in fact the Radon transform is reversible with no loss. Given a set of projections (Radon transform) of the data at different angles, a computer, and the magic of the FFT, the doctors can see a full three-dimensional picture of the inside of a human body. The same technology also works for MRI (Magnetic Resonance Imaging). There is more on the Radon transform in Chapter 9.

6. Parallel Computing

Back to Pasadena

After two years in Oxford, I was ready for a new challenge, and found a job offer doing "parallel computing". Really, that pure science thing seemed to involve a lot of trying out somebody else's hypothesis, finding out it's not true, then writing a paper about it. But the parallel computing at Caltech was with computers – my real passion – being put together in a new way, and with just enough physics to make it interesting. So after two years in Oxford, Barbara and I moved back to Pasadena, in February 1986. It was cold and gloomy in Oxford, but the air in Pasadena smelled of flowers.

The parallel computing group all came from a physics background, and there was a sort of unspoken axiom of faith that the engineering should be motivated by ideas from physics. An example is optimisation. Many things need to be optimised: routing, for example delivery of 100 parcels with the shortest distance. Scheduling – suppose a school is building a timetable of who teaches when, with many individual preferences and constraints. An airline needs airplanes, crews, and passengers to all arrive at the right times. In each case, an optimisation algorithm called *simulated annealing* can provide answers. Start with a possible configuration (route/timetable/schedule), and a way to evaluate the "quality" of that configuration – how many miles, or how many people inconvenienced, or just cost. Now make random changes to the configuration and select the one with the highest quality. Big changes at the beginning allows an exploration of the whole space of possible solutions to get a general idea of where the best quality might be. Then the size of the random changes (temperature in the physics analogy) is reduced. At the end of the annealing, the best configuration is essentially decided, and small changes are made to give small quality improvements.

The physics analogy is *annealing*: the slow cooling of steel to increase hardness and decrease internal stress. By letting the temperature fall slowly, larger crystals form in the atomic lattice of the metal, reducing internal stress, giving toughness to the metal. The random changes of the optimisation process are like temperature, with high temperature corresponding to the big changes at the start, and lower temperatures to the small changes, later. The large fluctuations try a wide range of possibilities and find the best general area, the small variations find the best among close variations.

I think of the annealing approach to optimisation when there is a meeting of people, followed by a dinner. The ideal solution has people who might collaborate, who share interests, to sit close to each other, so they can converse. Often what happens is the group arrives and seats themselves immediately, so everyone talks to a random other guest. But a “mingling” period before seating allows groups to self-select, so when they are sat down, they are closer to optimum configuration.

By contrast, the “quenching” of a metal is like sudden seating at convention dinner: small crystals of conversation rather than the larger social networks that can happen with a period of mingling and self-selection. Generally, quenched steel is not used, until it has been “tempered” by raising the temperature just below the critical point for a certain time. Like allowing people to change places at the dinner, if they don’t like who they are sitting with!

A Team of Computers

The parallel computer that we were building was a “hypercube”, meaning a lattice of 2^N computational nodes, each communicating with a few of its neighbours that differ by a power of two. For example, with N=4, there are 2^4=16 nodes, a four-dimensional hypercube. Node 13 would communicate with 12, 15, 9, and 5. Writing it in binary makes it clearer: 13 is 1101, and there is a communication channel for each of the binary digits. Each node of the 16-node system has four neighbour, and we can flip the bits to get the partners: 1100=12, 1111=15, 1001=9 and 0101=5.

It didn’t look like much in 1986, a pile of IBM PCs, each running Linux, and a screen and keyboard. But the point is that the group of machines acts as one, configured as a control node and the 2^N client nodes. There was no custom hardware, just out-of-the box mass-produced PCs. But by 1990, our group, in collaboration with Intel corporation, had the world’s fastest supercomputer.

The Intel Touchstone Delta had 512 processors that could work together (Figure 6-1) and I had my unstructured adaptive mesh code with a billion tetrahedra computing high-speed flows.

ROSEMARY KAUL / Los Angeles Times

Guests at Caltech for the unveiling of Touchstone Delta, the world's fastest computer, are reflected in the computer's black plexiglass panels, left.

New Supercomputer at Caltech Ranks as the World's Fastest

By DENISE HAMILTON
TIMES STAFF WRITER

The world's fastest computer—which can perform 8.6 billion calculations per second—was unveiled Friday at Caltech in Pasadena by scientists eager to begin using it to study projects ranging from depletion of the ozone layer to the aerodynamics of high-speed aircraft.

Called the Touchstone Delta system, the computer uses 528 high-speed microprocessors, linked through a communications network, to break problems into small parts and process them in short order.

Just one of the processors can add the numbers in a telephone directory within 1/100th of one second, said Roy Williams, a Caltech scientist who will use the supercomputer for physics research.

Officials of Intel Corp., which built the supercomputer, said it is a prototype for the firm's next generation of supercomputers. Intel, which for years had a virtual monopoly on making chips for personal computers, is expanding into high-performance computing and is one of eight firms worldwide—seven are from the United States—that now build supercomputers.

On Friday, dozens of scientists and representatives from the 13 research institutes, government labs and agencies that make up the Concurrent Supercomputing Consortium crowded into a Caltech laboratory. The consortium commissioned the building of the supercomputer.

Please see CALTECH, B2

6-1 (Left) The Intel Touchstone Delta in 1990, a collaboration of Caltech and Intel. A mesh of 16x32 i860 processors combine to give a speed of 32 gigaflops.(image Courtesy of Gwen Bell) (Right) The report from the L.A.Times shows me wearing a tie.

Back then we thought that every large company would need supercomputing capability, just as in the 1990s every company needed spreadsheet software. We thought we were at the beginning of something big. Turned out the need was not so much for computing power, but for the communications revolution that was the Internet.

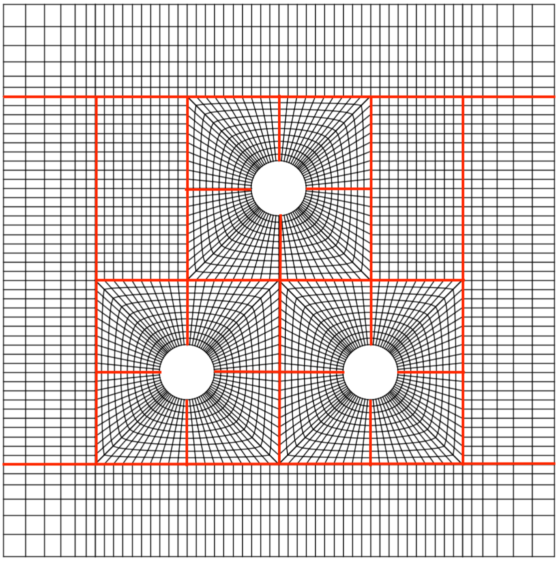
There were a number of experimental machines in our stable at that time. The *Ncube* in particular was used for fast parallel file systems, do deliver many streams of data using the many parallel nodes and disks. Eventually, *Ncube* was acquired by Oracle and became the basis of the thriving "video on demand" industry that we now take for granted every time we sit on the couch and stream a movie.

Computing Continuity

Calculus is the study of continuity. The speed of a car, for example, can be expressed in kilometres per hour, with the implied experiment of keeping the car at constant speed for an hour, and measuring the number of kilometres travelled. But in reality, as we all know, the speed of the car is not constant. So speed actually means *instantaneous speed*, with an implied limiting process. We can do kilometres travelled in a given minute. Meters travelled in a given second. Millimetres travelled in a given millisecond. The mental trick of learning calculus, the faith if you like, is to accept this limiting process without too much examination, and accept the governing rules. Bishop Berkely in 1734 was scathing in his criticism: "*And what are these Fluxions? The Velocities of evanescent Increments? And what are these same evanescent Increments? They are neither finite Quantities nor Quantities infinitely small, nor yet nothing. May we not call them the Ghosts of departed Quantities?*". But the fact was that calculus enabled Newton to compute the motions of the planets with exquisite accuracy, even the effects of one planet's gravity on another planet. A triumph of calculus in 1846 was the discovery of Neptune – by analysing tiny discrepancies in the orbit of Uranus, a new planet was hypothesised, and its position computed, then actually found at the predicted position, as discussed in Chapter 1.

Much of what is interesting to science and industry can be predicted by making differential (calculus) statements about behaviour, then using a computer to solve the system. A fluid, for example, has continuous quantities for its velocity, pressure, and temperature. Solving the fluid equations in an atmosphere on a rotating sphere gives us weather forecasts; the motion of fluid moving over a car tells us about air resistance and fuel efficiency. Add in electric and magnetic fields and we have plasma physics; add in molecular concentrations and reaction rates and we have the exploding gas inside the piston of a car engine; add in nuclear reaction rates and we can put a supernova inside a computer. Such power, but not so easy to get it right.

Continuous systems are modelled by sampling the values at specific places, or *discretising* the space, then inferring the intermediate values. If the car is moving at 60 km/h at one place, and 62 km/h further down the road, we infer that it was moving at 61 km/h at the half-way point. This discretisation can sometimes be simple, but there are often confounding factors. One confounding factor when building a mesh is geometry. Figure 6-2 shows a mesh suitable for computing in a geometry that is a square outer boundary, with three internal disks. Often, there is a lot of physics happening very close to the boundary, for example the boundary layer of air passing over an airfoil is a critical part of computing the lift and drag. So extra mesh resources may be required there. Building these kinds of multiblock meshes for a 3D geometry like an airplane can be quite complex.



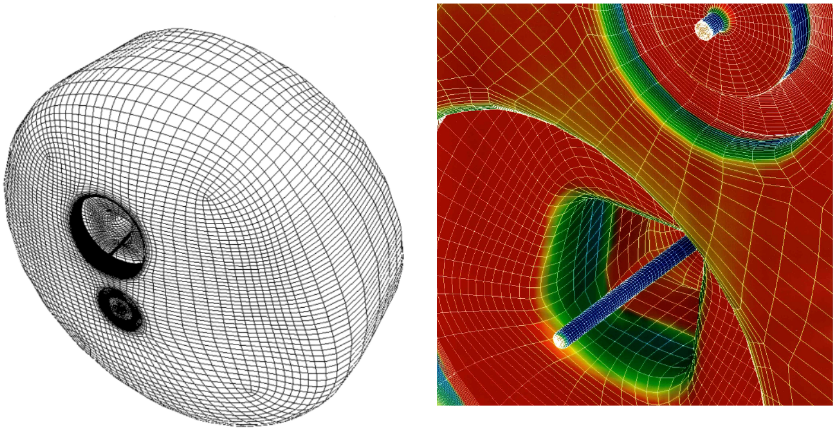
6-2 A geometry consisting of three disks inside a square, modelled with a multiblock grid of 18 blocks.

Cassini-Huygens at Titan

As a practical application of these multiblock meshes, there was a project in which I was involved the 1990s, computing the aerodynamics of a spacecraft, working with Jochem Häuser of the European Space Agency. Not a spacecraft in the Earth's atmosphere, but in the atmosphere of Titan, the largest moon of Saturn. The Huygens lander, about 1.3 metres in diameter, piggybacked on the Cassini mission to Saturn, which launched in 1997; the lander descended to a soft landing on Titan in 2005. It was the first landing on a moon other than Earth's. It sampled the upper and lower atmosphere of Titan, as well as sending back pictures of the surface.

Our project, in 1995, simulated the instruments facing into the flow, instruments to measure properties of Titan's atmosphere during the three-hour descent. The flow was computed in order to determine the effect of dust on the instruments. The worry about dust was because the combination of hypersonic heating and an atmosphere containing methane could make free carbon dust, that would attach itself to the lenses of the cameras and ruin the view. See Figure 6-3. The inhomogeneity of the mesh on the left provides very fine resolution for the instruments themselves, and fewer mesh points for the smooth body of the spacecraft where the flow is smooth and varying on longer scales. This mesh was built with knowledge of where the detail is and is not required.

The computation was carried out on the Intel Paragon parallel supercomputer at the California Institute of Technology. My part was to split the blocks of the multiblock mesh between separate processors and organise communication between them. The upshot was that even if carbon dust is produced, the flow will drive it away from the delicate instruments.



6-3 Left: The surface of a multiblock mesh used to compute the aerodynamics of the Huygens lander, not including the much larger heat shield, which descended to the surface of Titan in 2004. Right: A closeup of the instrument cluster on the descent side of *the spacecraft, as a simulation in 1995. Colour represents pressure, at a descent velocity of Mach 10.*

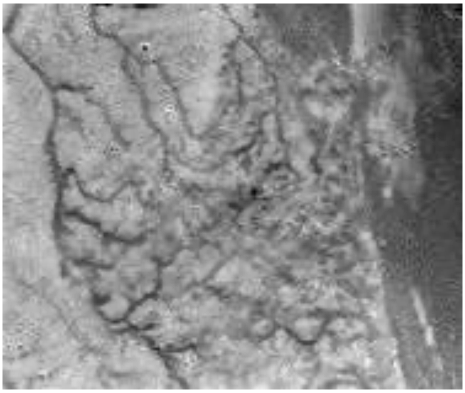
Ten years after this computation, I was waiting with bated breath as the Huygens was released, hoping the cameras and other instruments did not get obscured by soot! Reality proved the simulation correct. The landing went perfectly, and the instruments did not get clogged up with soot – as we predicted in 1995. The image in Figure 6-4 was taken by the Huygens instruments, showing that the shape of the rivers of liquid methane on Titan actually look a lot like the rivers of water on Earth.

Here is my blog entry following the landing in 2004:

Titan would be a cold and gloomy place for a human. Light gravity, like the Earth's Moon, but a thick atmosphere of continual overcast, the distant Sun providing little light. The temperature is always colder than the coldest ever recorded on Earth — by more than a hundred degrees. While the atmosphere has no oxygen, it has more pressure than the Earth's atmosphere, so a pressure suit is not needed, a Scuba mask would be enough. But a thick coat would be needed. or else the cold would kill in seconds.

Huygens found slushy ground, perhaps a mixture of sand and liquid ammonia. Rivers of methane have cut channels in the hills of ice and rock. Where Huygens landed was swampy, islands of ice in lakes of liquified natural gas. Is there anything crawling out of the ponds? Life there would be alien in an unimaginable way — because the chemistry would have to be very different, without carbon molecules floating in water, but rather carbon compounds floating in liquified natural gas.

It could be there is an expedition being mounted even now against the invader from the Sun. Perhaps the Huygens probe actually landed on the skin of a vast angry monster. In fact, I suspect the Armageddon prophesy refers to the invasion of the Titanians, should be around seven years from now, in 2012.



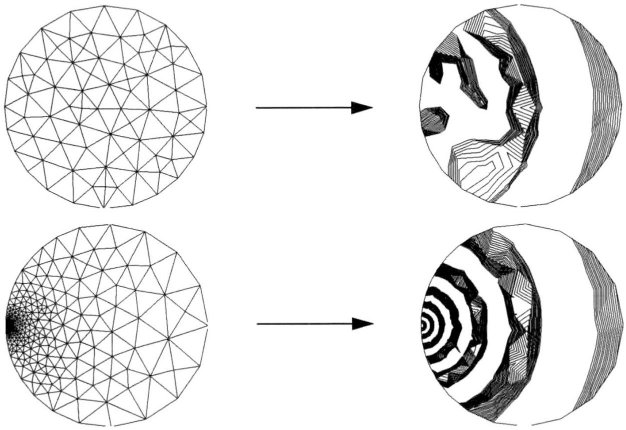
6-4 This image taken during Huygens descent stage shows drainage channels and a flood plain. The temperature is -180 C, so the land is rock-hard ice and the rivers are liquid hydrocarbons like methane. (ESA/NASA/University of Arizona)

Adaptive Meshing

It may be that the quantities being simulated have immensely different scales: imagine the fluid velocity and temperature associated with a candle flame. These quantities have high gradients at the flame itself, and a resolution of a fraction of a millimetre is needed – but a meter above the flame things don't change very much over a centimetre or so. In the example of Figure 6-5, the mesh is automatically optimised to show the details of a function that is inhomogeneous in scale.

Sometimes the inhomogeneity is not known in advance. Think of modelling traffic on a motorway as a fluid: there would be average traffic velocity and density of vehicles, there might be constrictions of the road and hills, but suddenly there is a *shock*, a wall of brake-lights, where the traffic changes from fast, low-density to slow, high-density flow. This part of the flow needs a much finer mesh, but in a place that cannot be predicted in advance.

When I arrived at Caltech – for the second time, in 1986 – I decided to work on unstructured meshes, at first triangles in 2D, later advancing to tetrahedra in 3D. I built a software platform to run physics simulations on unstructured meshes, using parallel computers. The result was DIME: Distributed Irregular Mesh Environment.



6-5 Top: an unstructured triangular mesh covering a disk, and a contour plot of a function that is samples at the vertices of the mesh. The function has much more detail at the left side, that is not captured. Bottom: A mesh concentrated at the left side

The vertices of the mesh are where the computed field values are stored: pressure, temperature, density, and so on. We can think of the elements – triangles or quadrilaterals or tetrahedra – as computational units. Each triangle pulls the three values from its vertices, then computes the changes to those values. Because there are several triangles associated with each vertex, there are changes from several triangles than must be added together to get the new values of the field.

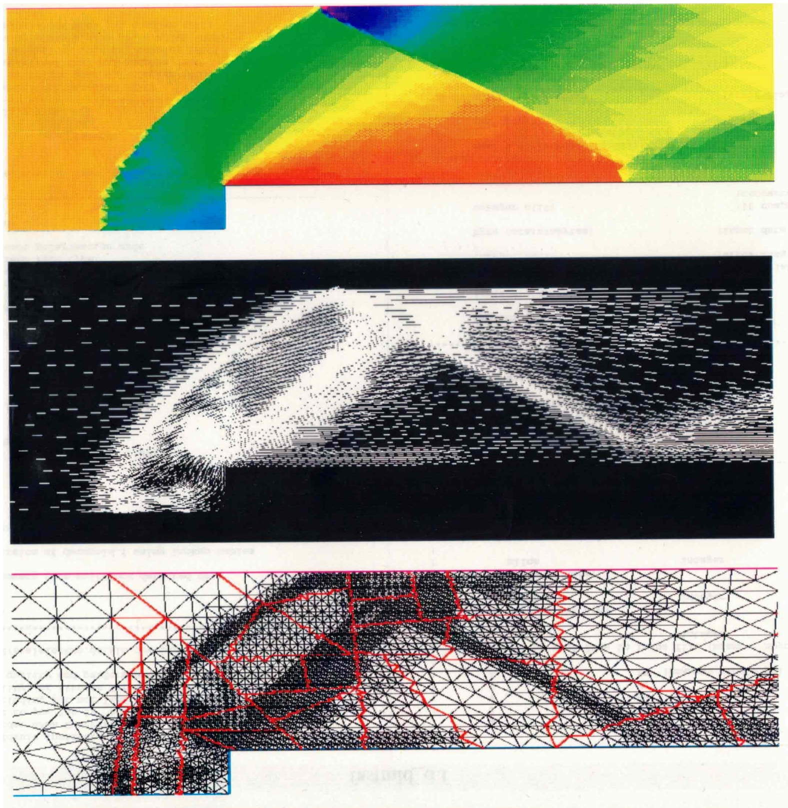
In Figure 6-6, we see the vertices and triangles explicitly. Also the mesh is split between processors of a parallel computer, meaning some vertices are copied into different processors. When the triangles make their updates to their associated vertices, the copies at the boundaries get contributions from some of the triangles, and need to communicate with each other. If a vertex is in processor A, and a copy of it in processor B, and one gets an increment of 0.1 and the other 0.2, then the communication between processors A and B adds the increments and pushes the result back to each: now processors A and B both have the value 0.3.

Another step is the mesh refinement (*Rivara refinement*) where triangles can be split so that vertices are added in the places where gradients are high. This is through a sequence of steps that ensures the correctness of the mesh. When the simulation detects that enrichment is needed, this algorithm makes it happen. This adaptive refinement has enriched the mesh at the front and back of the airfoil, as well as at the vertical shock that has formed about 2/3 back from the front.

6-6 An unstructured triangular mesh consisting of triangles, nodes, and boundaries, split between at least 25 processors on a distributed-memory parallel computer.

Figure 6-7 is an example of two-dimensional DIME running a simulation of a hypersonic flow impacting a step. A shock forms upstream, which is reflected from the top boundary. The simulation was done alternately: run the fluid simulation on the mesh until steady state is achieved, then use the gradient of the pressure to decide where the mesh should be refined; then repeat both steps until convergence. In the bottom panel, the red lines separate the domains of the 32 processors that were used.

Given that the processors exchange information only when every processor is finished with the integration step, the overall speed of the simulation is determined by the slowest processor. Therefore we should split the mesh so that each processor has approximately the same amount of mesh. Furthermore, each processor communicates data about every shared node of the mesh, so we want to minimise the boundary between processors.



6-7 Hypersonic flow impacting a step. Top: pressure, Middle: velocity, Bottom: mesh adapted by the conditions of the simulated flow.

Splitting Space

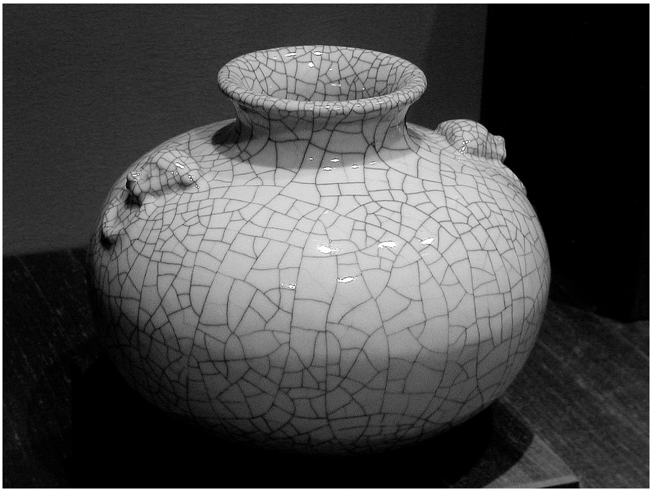
We are trying to simulate physical phenomena, and physics is generally local, so we might consider ways nature has made a way to split domains into parts. The constraint here is that nearby regions of physical space should map to nearby regions in the machine memory, and that communication between processors is a cost that should be minimised. Since that communication happens across domains between processors, it follows that we should reduce the surface area of that interface.

As I mentioned before, the Caltech parallel computing people were all physicists, convinced that models from physics could be used to optimise the computations. For example, we know the sphere minimises surface area for given volume, which is why a bubble is spherical. Converted to computation, we want a compact, roughly spherical domain for each processor. Continuing the bubble analogy, soap froth is a model for this: each internal bubble tries to minimise its surface area, see figure 6-8.

Another example is *craquelour* texture that occurs with paint and glaze (Figure 6-9), because the applied coating shrinks at a faster rate than the surface to which the coating was applied. Notice there is a hierarchical aspect to this, you can often see that at first there was a single long crack to relieve stress, then a new stress distribution, then secondary cracks appear which terminate in a 90 degree junction. Thus it is possible to see the history of a specimen through the hierarchy of cracks.

There are many other beautiful examples of splitting space into polygons or polyhedra. Figures 6-10 and 6-11 show geological polygons on the Earth, and Figure 6- shows polygon structure on Pluto, thought to have similar origins to the pingos of Figure 6-12.

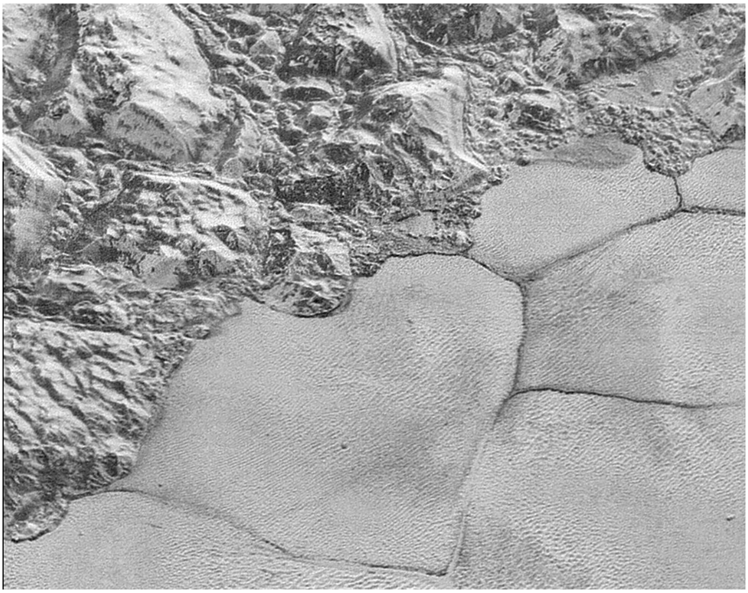
6-8 Air blown through soap water makes a characteristic partitioning of space into polyhedra, known as the Weaire–Phelan structure.



6-9 Craquelure glaze on ceramic, characteristic of Chinese Ge or Guan ware. The coefficient of expansion differs between the glaze and the body, the former contracting faster. Most of the junctions are 90 degrees.

6-10 Cracked Earth in the Rann of Kutch, on the india-Pakistan border. Most of the junctions are at 90 degrees. (Wikimedia, Vinod Panicker, CCAS2.5)

6-11 Arctic polygons and a melting pingo in Pingo National Landmark, Canada. The protected area contains a quarter of the world's pingos. (Wikimedia, Emma Pike)

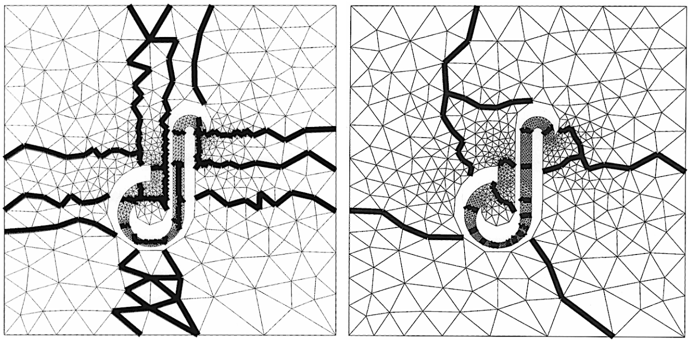


6-12 Polygons on Pluto, perhaps caused by sublimation-driven convection of water ice (NASA).

Thus we might think of using a natural process to balance the load between processors. Looking at the natural processes above, we understand there are two essential mechanisms:

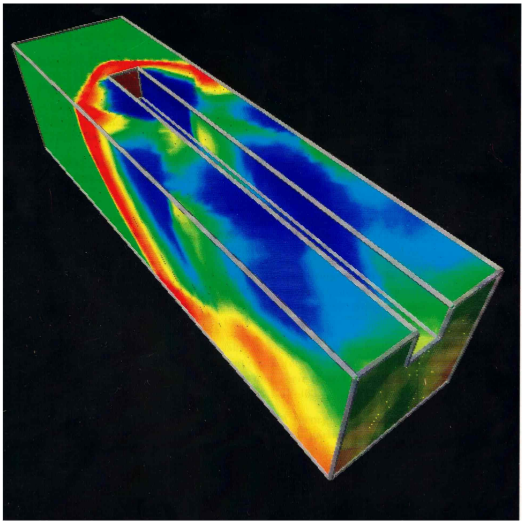
- Hierarchical, with a first split, then each side is split into to, then each quarter split into 2, and so on. This results in 90-degree junctions of the dividing lines.
- Holistic, where the splitting appears everywhere at once, which results in 120 degree junctions.

Figure 6-13 shows load-balancing strategies for a parallel computer splitting a mesh, one based on a Hierarchical approach, the other the Holistic approach.

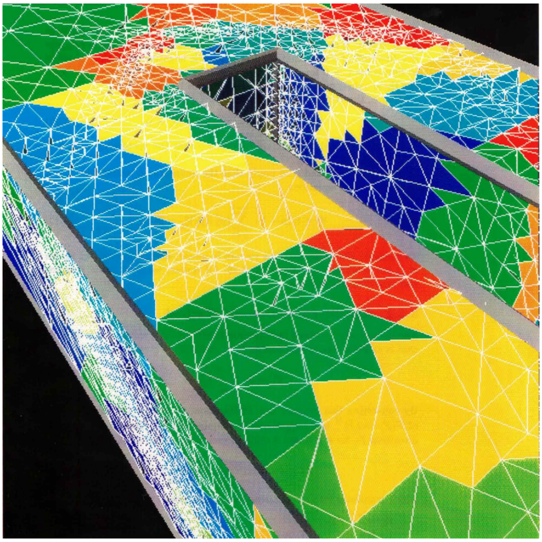


6-13 The same mesh split amongst 16 processors. Left, by hierarchical method, right by the holistic (eigenvalue) method. Note that the very concentrated belt of mesh is split more efficiently by the holistic method.

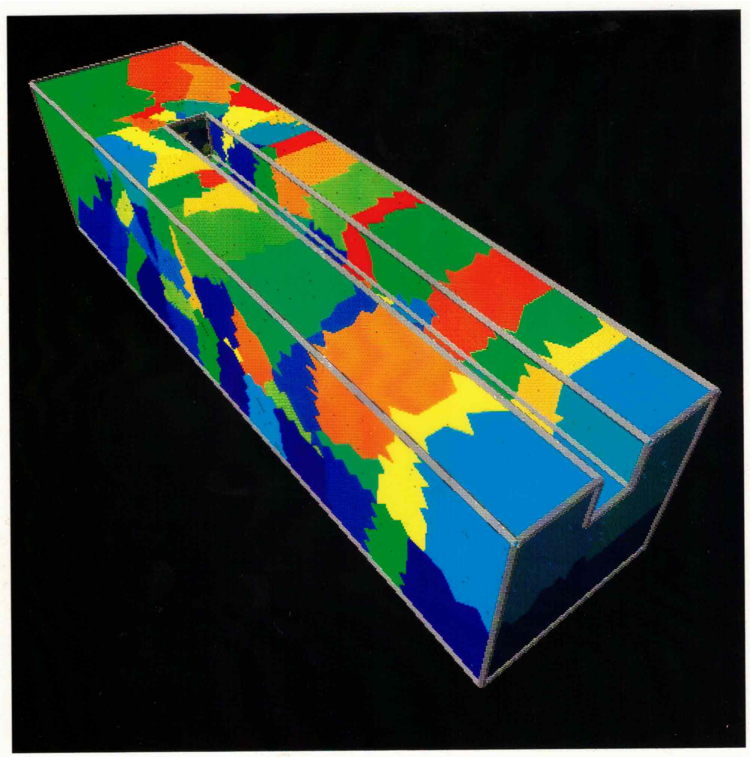
One simulation using DIME and its load-balancing algorithms was a flow simulation at Mach 1.5 in a 3D tunnel with a block placed inside. Figure 6-14 shows the result of the simulation. It's a tube of square cross-section, where the area of the cross-section is suddenly reduced by an obstruction, and shocks form ahead of the obstruction. Figure 6-15 shows the mesh of irregular tetrahedra, with a concentration of resources where the pressure gradient is greatest, at the shocks. Figure 6-16 shows the distribution of the mesh among the 512 processors. Each processor has about the same number of tetrahedra: given that the whole computation synchronises every time step, the work should be distributed equally for efficiency.



6-14 A compressible flow simulation in a square tunnel with an incised square block, using 512 processors of the Intel Touchstone Delta. Colours indicate pressure.



6-15 The adaptive tetrahedral mesh used for the compressible flow simulation.



6-16 The assignment of the 512 processors to the tetrahedral mesh.

7. Early Internet

Al Gore Did Invent It

The internet started in the 1970s and 1980s, connecting US universities and some military sites. There was a strict rule about no advertising or commercial activity. It got a big boost in 1991 with the High-Performance Computing Act, that was created and introduced by then Senator Al Gore, so in spite of the jeering of Republicans, he was partly responsible. The internet took off because of three components: the physical communication network; the TCP/IP, `http` and `html` standards to provide rich documents with links; and the Mosaic browser, introduced in 1993.

To start with, the physical links were in place, and the TCP/IP protocol allowed remote machines to connect and communicate. Soon came the domain-name system, giving human-readable names to machines. Communications always required authentication, until the arrival of "anonymous ftp", a protocol that allowed anyone to connect to a machine and get content. There was no browser, it was typing commands and receiving a file. Back in 1989 (Figure 7-1) all the sites on the internet could be printed on three pages! Many of them were devoted to Star Trek, Monty Python, or free code. The term "spam", meaning unwanted communication, is from a Monty Python script, where men in Viking uniform are in a café, and everything on the menu has unwanted spam with it.

It was the latter components, `html`,`http` and the Mosaic browser, that did the same thing under the hood, but in a much easier way that anyone could manage. Caltech, where I worked, had a website by the end of 1993, and I had made one for my organisation (Caltech Concurrent Supercomputing Facility), with some links to other places that had websites. The world-wide-web really caught the attention of the public with the world's first webcam, pointed at a coffee pot in Cambridge, England, so that people could look first before making the trip to the coffee pot and finding it empty. The bizarre thing was the state of the Cambridge coffee pot could be viewed worldwide. The camera was finally switched off in 2001, and the last coffee machine sold on ebay for £3,300. I can remember trying to convince friends that the internet will be big, but they pooh-poohed the silliness of wasting valuable communication cables on a picture of a coffee pot.

89/12/05
11:22:57

ft

From cit-vax!henry.jpl.nasa.gov!elroy.jpl.nasa.gov!decwrl!shelby!lindy!ucscb.UCSC.EDU!oc
in Tue Dec 5 11:19:59 PST 1989

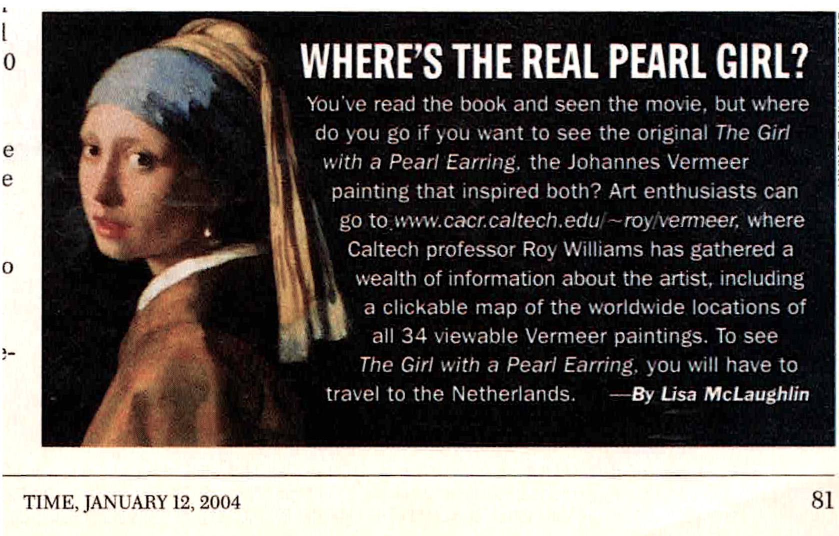
This is my list of Internet sites accepting anonymous ftp:
user name: anonymous
password: <your login name>

andy.bgsu.edu	129.1.1.2	Unix sysadm tools, Unix Vote by
andy.bgsu.edu		mail, Unix etc., College hockey
andy.bgsu.edu		stats, Monty Python scripts
anise.acc.com	129.192.64.22	Berkeley utils ported to A/UX,
anise.acc.com		Motorola DSP 56000 repository
anna.stanford.edu	36.14.0.13	Anna (Annotated Ada) softare and
anna.stanford.edu		docs
apple.com	130.43.2.2	tech-notes, worm papers
aramis.rutgers.edu	128.6.4.2	idea, RFCs
aramis.rutgers.edu	128.6.25.2	
argus.stanford.edu	36.56.0.151	netinfo
ariel.unm.edu	129.24.8.1	ethics documents
arisia.xerox.com	13.1.100.206	sunfixes, mac, LispUsers, tcp/ip,
arisia.xerox.com		IDA sendmail kit, etc.
arpa.att.com	192.20.225.1	fio library routines
arthur.cs.purdue.edu	128.10.2.1	RCS, buildtex, deTeX, mac32, Purdue
arthur.cs.purdue.edu		Tech Reports, xspeed
aurora.arc.nasa.gov	128.102.21.1	unknown

7-1 Part of a list of 600 Internet sites that supported anonymous ftp (file transfer protocol) from 1989.

As interest grew at Caltech, I was invited to give a lecture to some 1,500 people on 5 January 1995. I tried to emphasise how it could all be used for good, for empowerment and knowledge. We all had such high hopes! There were no big companies at that point, no Google or Amazon. It was all web 1.0, meaning that only experts in the arcane HTML language would create content. In 1996 there was the internet sensation of “Jennicam”, a college student named Jenny, with a webcam that recorded her apartment every three minutes. Instead of being used only by nerdy people like me, the Internet was becoming part of popular culture. Having used email for 10 years, I was astonished to be in line in the supermarket and heard somebody behind me talk about sending email. I turned around to look, and the speaker was not somebody from Caltech!

I spent several days in a coffee shop, writing notes about the paintings of Vermeer (see Figure 7-2). People were hungry for content on the new shiny web, so Time magazine would publish interesting websites. There was a big exhibition of Vermeer that opened in Washington DC in Nov 1995, and I had a work trip there, so I asked to meet the curators. I remember explaining the internet and web and HTML to their puzzled faces, and showing printouts of the pages I had made. How you could re-order the paintings by age, or by subject, or by location, it wasn't a fixed order like a book. But they were not impressed, with my greyscale printouts, did not believe when I said that there should be a website. They liked the large-format catalogue printed on high-quality paper, and there was no web page for their blockbuster exhibition. Of course, today their website[3] is wonderful expression of artistic sentiment and practicality, bursting with colour and grace.



WHERE'S THE REAL PEARL GIRL?

You've read the book and seen the movie, but where do you go if you want to see the original *The Girl with a Pearl Earring*, the Johannes Vermeer painting that inspired both? Art enthusiasts can go to *www.cacr.caltech.edu/~roy/vermeer*, where Caltech professor Roy Williams has gathered a wealth of information about the artist, including a clickable map of the worldwide locations of all 34 viewable Vermeer paintings. To see *The Girl with a Pearl Earring*, you will have to travel to the Netherlands. —***By Lisa McLaughlin***

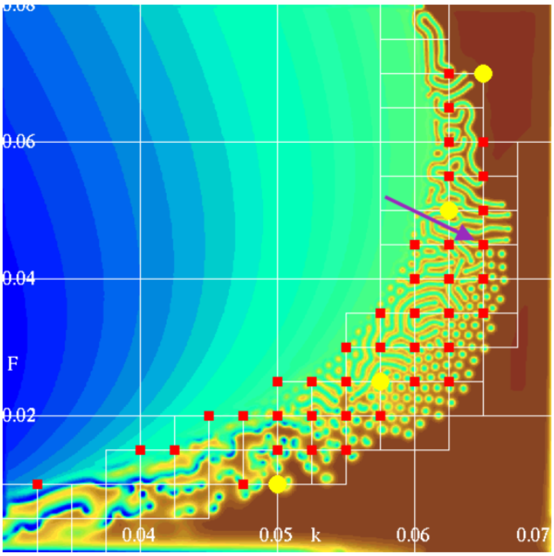
TIME, JANUARY 12, 2004 81

7-2 My web site "Paintings of Vermeer" was in Time magazine in 2004.

I built a few web pages back in the day of Web 1.0, when most people could view, but did not have the specialised skills to build web

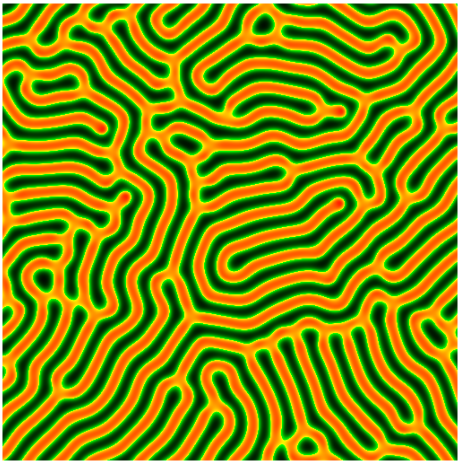
[3] https://www.nga.gov/

pages. I had become fascinated with the work of Alan Turing on how the leopard gets its spots – his masterpiece paper "*The Chemical Basis of Morphogenesis*" (1952). A simple system of differential equations that can be written on a square inch can show an astonishing variety of behaviour, much of it very biological, like the spots of a leopard or the whorls of a fingerprint. Figure 7-3 shows the parameter selection, where each red square links to a still image, and each yellow square a video. For example choosing F=0.045 and k=0.065 leads to Figure 7-4. Xmorphia has been taken up by others[4] and is a fun activity. When I was in the National Museum of Scotland recently, their "brain coral" (Figure 7-5) really reminded me of the Turing theory.

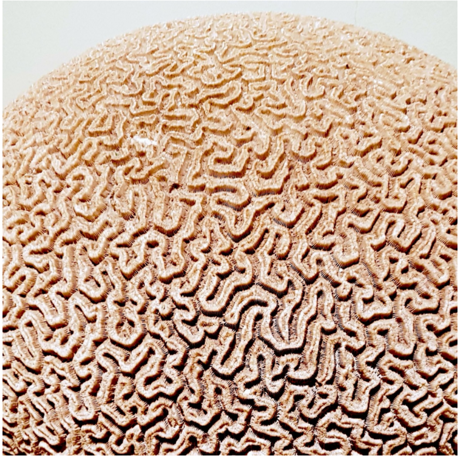


7-3 The Xmorphia selection image, representing the behaviour of a set of differential equations (Gray-Scott). (Middle) the image obtained by clicking the end of the purple arrow. (Right) a grooved brain coral from the Caribbean.

[4] https://www.lanevol.org/resources/gray-scott



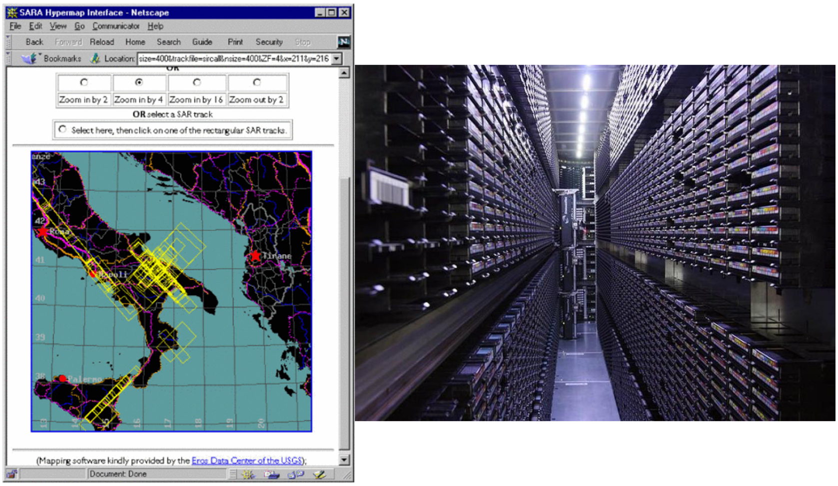
7-4 The image obtained by clicking the end of the purple arrow



7-5 A grooved brain coral from the Caribbean, in the National Museum of Scotland.

Internet for data delivery

I was beginning to have an idea that the Web could be useful for scientific data and invented a way that a machine could be programmed to fetch scientific data. The meaning of the data and how to use it would be on the web pages for humans to read, and the machines could then fetch the actual data. Of course others were having the same idea, and it wasn't my code that became popular, but rather somebody else's (`curl` and `wget`). This has been the driver for my career for 30 years now: web-based interfaces to scientific data.



7-6 (Left) User interface to the SARA digital library, shown with the Mosaic browser. You could click on the map to select which data would be returned. (Right) Storage Tek tape robot similar to the one installed at Caltech Concurrent Supercomputing Facilities in the late 1990s (image: NCAR).

With Italian colleagues, we came up with the Synthetic Aperture Radar Atlas (SARA), see Figure 7-6. The space shuttle had a program a few years before to take images of the surface of the Earth with radar, and the data was distributed on magnetic tapes for researchers. The boxes of tapes had been transferred to a storage robot – like a small room, all black, with the walls covered in cubbyholes where tapes were stored, and when you ordered up data, a machine would fetch the tape and insert it for reading. A bit like a juke box from the 1950s, where the robot arm picks out a vinyl record. There was also custom computing, combining the 8-channels of the SAR data, running contrast algorithms, and so on, before creating an image in red, green, and blue. When school parties came to visit the computing centre, they loved to click on the web browser and see the robot fetch the tape.

There was great interest in remote sensing from satellites in the 1990s, with applications from archaeology to tax enforcement. I met the Italian minster responsible for space technology, and he was excited that the satellite image could find swimming pools that were unregistered and therefore were not paying the swimming-pool tax.

The domain-name gold rush

Let me touch on the investment opportunity of 1999. In retrospect, of course, the investment opportunities were Apple and Amazon, but there was plenty of fun to be had buying and selling domain names. Companies were waking up to the idea that they needed a presence on the world-wide web, and that they needed a domain name. There is the famous story of Josh Quittner, a writer for Wired magazine, who asked McDonalds Corporation if they wanted to register the domain name `mcdonalds.com`, and they declined. He bought the domain, and refused to give it to them until they promised to provide free high-speed Internet access for a public school in Brooklyn.

I bought `escience.com` and sold it for $2,000. My friend Angus Hanton bought a lot of domain names, including lot of numbers – `five.com` and `seven.co.uk` for example – and made a packet selling the `three.co.uk` to a mobile phone company. For a Valentine's Day present, I bought my girlfriend a domain name – `softling.com` – but I don't think either of us did anything with it. It was a time when anyone with a knowledge of HTML could get a job immediately, anyone with a idea for a domain name could snap it up and sell it. A frothy time of expectation and hype, with obscene amounts of venture capital available for startups. The survivors and thrivers from that time are today's big business, Amazon and Google among them.

Worst way to calculate pi

I'll close this sub-section with a little snippet that I put on one of my early web pages, when I didn't know what else to do with it. It's not the most efficient way to calculate pi, rather it's the *worst* imaginable algorithm for pi! The algorithm starts with millions of digits of the expansion of pi, then uses them to create a *much less accurate* approximation!

Begin with the probability that two large integers are coprime (i.e. no common factor). For each prime p, the probability of both being divisible by *p* is $1/p^2$, so the probability Q of being coprime is a product over the all primes: Q = $\Pi\left(1-\frac{1}{p^2}\right) = 6/\pi^2$. Astonishing how pi gets itself into the strangest corners of maths! Anyway, my worst algorithm starts with lots of random pairs of large integers, works out the proportion Q of pairs that are coprime, then computes pi as $\sqrt{6/Q}$. But the kicker is that the source of the large random integers is the value of pi itself! If you start with pi computed to a million decimal places, you might get 10,000 pairs of big integers, the result being an approximation with 4 digits accuracy. Start with a million significant digits, end up with four. That's what I call useless.

Web-Database Systems

I have spent much of the last 30 years building web-database systems for scientific use. Finally I found out what I am good at. The first was the LIGO bakeout system, to show the engineers the progress and history of the heating of the LIGO beam tube. See Figure 11-1 to see the 4km tube covered in concrete and snow: inside is the steel tube, evacuated to ultra-high vacuum. It took 40 days of pumping to remove the air, then 30 days at 160 Celsius to drive off residual gases and volatiles, so 10 weeks altogether. I built the system to show the progress to the engineers in charge.

In such a system, are three layers: Model, View, and Controller. The model is the internal representation of the data. In the case of the bake-out it was the temperature history for each location on the beamtube and each day of operation. The View is a visual or tabular display, perhaps time-history of the temperatures at a given point, or view of the whole beamline at a specific date. The controller is the logic, built as computer software, that translates the model into the view. In addition there is an Ingestion system, where measurements of temperature are checked and formatted so they are suitable to be ingested into the Model.

I introduced the LIGO project to the use of a "relational database", a data storage paradigm pioneered by Edgar Codd, who published *A Relational Model of Data for Large Shared Data Banks.* Data is stored as tables, virtual tables called views, with ideas such as joining tables and foreign keys, systematic treatment of missing data, and not repeating data. It was perfect for the bake-out project, and I persuaded IBM to give us their DB2 product for free. The View part of the project was also new, incorporating the idea of using the HTML language that Tim Berners-Lee had built for web pages. These days, there is so much scientific data available through this combination of relational database and web server, but it was brand new back then.

Let me finish this chapter on the early internet with a couple of articles I wrote in 1995 and 1996, thirty years ago, for the Caltech “Engineering and Science” magazine. One is full of optimistic predictions of the benefits brought by the internet – but predicts “relentless corporate agendas”. The other predicts Google Maps and Streetview ten years in advance.

Infinite Information on the Internet (1995)

One might find the Internet worthwhile because it connects millions of computers, so that anyone of them can exchange data with any other; but the really important thing is that it connects millions of people, enabling them to communicate with one another in ways they never could before. Like the telephone and jet aircraft, it will change our lives. The Net is a tool you can use to enhance your professional, social, and perhaps even spiritual lives. Don't expect the miracles proclaimed by the media, but don't be afraid of it either. It's getting integrated into the fabric of everyday life – ubiquitous computing, ubiquitous Internet.

Like many new technologies, the Internet started out as something for the elite, but now it's becoming ordinary, and soon it will be necessary. The number of machines on the Internet has exploded in the last year-it's doubling every 18 months – and if it were possible for this rate of increase to continue unabated, there will be a computer on the Net for every human being in the world by the year 2010! There is no centre to the Internet, except in certain minor ways, and there is no hierarchy of control. It is a radical decentralization that works, and I think this is something unusual in human creations. Indeed, this lack of a centre is vital to the robustness of the Internet; it was built during the Cold War to resist nuclear attack, and yet it's ironic to contrast the military parent with the free-spirited child to which it has, perhaps inadvertently, given birth.

The Internet is actually a network of computer networks connected by data links, the speed of which determines the user's frustration level. A telephone line equipped with a fast modem can transmit data at perhaps 20,000 bits per second; if we take the complete works of Shakespeare (about 40 million bits) as a unit of information, then it would take less than an hour to be transferred at that speed. It may be that we only want a small quote from Shakespeare, and if we need to get the complete works before we can see the quote, then an hour is definitely too long. Furthermore, images, sounds, and moving pictures take an inordinate number of bits to represent: one full-color image can easily require millions of bits, and compressed TV-quality video needs five million bits per second.

In order to be sent across the Net, your data is broken up into “packets”, each of which carries your computer's return address and the address of the recipient computer. A router reads the address and sends the packet to another router to which it is physically connected, and which is (hopefully) closer to the destination. To make this decision about which machine might be closer, the router needs some knowledge of its local environment; this knowledge is updated without human intervention, so that when part of the Internet is damaged, data automatically flows around the crippled link. Thus two messages from one machine – or even two packets from the same message – may reach the same destination by very different routes. The Net's protocol (known as TCP/IP) also assumes that packets may get lost. So each time a destination machine receives a packet, it sends an acknowledgment back to the sender. If the sender doesn't get the receipt back, it waits a while and sends the same packet again, and again; however, the time between these repeated packets gets longer and longer, in order to avoid saturating the system with packets sent to a machine that may be broken. At the destination, the recipient machine collects all the packets and throws out any duplicates. It also puts the packets back together in the right order, if necessary, since they can arrive out of sequence after their journey.

E-mail is trendy now, but soon it will become necessary. "You don't have an e-mail address yet?" is a disdainful question increasingly heard by the have-nots. For the haves, it is increasingly difficult to bother to communicate with the one member of the collaboration who does not have e-mail. In science, at least, e-mail makes long-distance collaborations easy. I collaborate with a colleague in Atlanta, for example, and we exchange e-mail two or three times a day. It's a nice interpolation between the formality and solidity of a paper letter and the undocumented ad-libbing of the telephone. You have time to compose the message carefully, but it is then delivered very quickly. And you don't have to talk to answering machines! Soon, I think, the hyphen will be dropped, and it will be just “email”.

The Internet's second aspect, the Usenet, consists of newsgroups, also known as bulletin boards, each of which is devoted to a specific subject. You send a message to the newsgroup, where it gets posted, and anybody who subscribes to the newsgroup can read it. Your posting disappears after days or weeks, otherwise the system would fill up. Newsgroup names look somewhat like e-mail addresses, except that the words get more specific as you go from left to right-for example, “`alt.clothing.sneakers`”. This is a real newsgroup – just one of some 5,000 accessible from Caltech. There's a newsgroup for everything, it seems. There's one for The Simpsons – the cartoon-and one for 0.J., too. There's a Newt Gingrich newsgroup. There's even one called “`alt.tv.dinosaurs.barney.die.die.die`”. Readers of a newsgroup will frequently follow a “thread”, or topic of discussion, that continues through several messages; angle brackets along the left margin of a message are used to mark material that has been included from somewhere else, generally from a previous posting on the same subject, and you can get many layers of angle brackets. Also, a lot of scientific conferences are planned and advertised through the Internet: agendas are set and speakers recruited electronically – it's only when you arrive that you get deluged with paper. And the Net isn't just for grown-ups-a 10-year-old girl in New Zealand can become pen-pals with a nine-year-old in Springfield, Virginia, through a group called “`k12.chat.elementary`”.

Because they put so many people in touch at the same time, newsgroups can produce adult friendships as well in a way that paper letters and the telephone cannot. For example, two years ago, before my wife and I went to Moscow, I sent a posting to the Usenet group on Russia, “`mlist.russia`”, asking, “How do I get from the Moscow airport into the centre of the city?” The replies ranged from “You just get on the bus” to “Don't do this, you will be shot”. But one reply was from a biochemist at MIT who has a sister in Moscow, and who asked me if I would deliver some medicine, because the Russian mails are so unreliable. She sent the medicine to us in California, and we met her sister in Moscow, enjoyed her company, and made friends. This relationship was formed because of the Net.

The last, and currently the most talked-about-feature of the Internet is the World Wide Web and the similar services such as Gopher. You can look up almost anything on the Web – today's Senate calendar, how to make an origami frog, weather forecasts for Siberia, a history of the vacuum cleaner. Even the Encyclopaedia Britannica is on-line, but that you have to pay for. The Web is based on the idea of hypertext, in which multimedia documents – images, sound, and video clips, as well as text – are linked to other documents. Links can appear on your screen as underlined words or phrases, as push buttons, as icons, as images, or even as different parts of the same image. When you point your cursor and click on a link, the document you're viewing is replaced by the one at the other end of the link. A link can lead to a document on a machine anywhere in the world, which is how the World Wide Web got its name. The Web – using a browser – is growing even faster than the Net itself-doubling every three months, versus every 18 months for the whole Internet. Every human on the planet will be on the Web within five years at this rate!

To use the Web, you use a program called a browser, such as Mosaic or Netscape. When the browser is started, it brings up your “Home Page”. This is your point of entry to the Web, and you can always jump back to it by pushing the Home button on your browser, so you can't get too badly lost. You can use an institutional Home Page, such as the Caltech one, or you can create your own personal Home Page. You can put all sorts of personalized stuff on the Web, including links to whatever you think is interesting. Somebody once said, “I didn't know what to do with it, so l put it on the Web.”

In order to get connected to the Internet, you call a service provider: America Online, CERF-net, CompuServe, Netcom, Psi, or a host of others, who will charge you a fee and give you a phone number for your modem to call and software to let your computer talk to the

Internet computers. (The Net is in some sense free, but you've got to pay both for the phone use and for your service provider to connect you to it). Levels of service vary, but so far *Consumer Reports* hasn't done Internet Service Providers – I'm sure it's just a matter of time! Your service provider may offer on-line help, which the Net doesn't. Service providers may also offer extras like access to Sabre, the airline-ticketing database; legal databases such as Lexus; the world's magazines and newspapers through Nexus; specialized stock-market databases; and so on. You can get censorship from your service provider, if you want it – you can have a separate account with restricted access for the kids, which is like having the phone company prevent 976 calls being made from your phone.

Many magazines and newspapers are available on the Internet; there are even e-zines, as they are called, that exist only on the Net and aren't published on paper. A lot of publishers are transferring their paper offerings *en bloc* to the Net, even though they're not quite sure what they're doing or why. An on-line clone of a magazine is easy to read – there's no advertising, no perfume samples, no bits of paper dropping out on the floor when you turn the page – what does that tell us about how long this kind of service is likely to last?

There are books on the Internet: the great classics certainly, but also modern, copyrighted works. The publisher hopes that people who see the book on the Net will go out and buy the real thing, but cheapskates can simply read it on-line or print it on a laser printer instead of buying it. Will we wind up reading more things directly from the machine, without printing them? I think the era of the bedside computer is not so far away. Scholarly articles are not just appearing online, but their on-line “publication” is squeezing out the paper journals. Paul Ginsparg, at the Los Alamos National Laboratory in New Mexico, runs a system called “`xxx.lanl.gov`”, which contains a database of preprints of high-energy physics papers. People download about 30,000 preprints per day from his system, and roughly 10 new preprints are added per day. This is really catching on, because paper journals take a year or more to publish something, but when you send a paper to Ginsparg's system, it becomes available to the global scientific community immediately. Many high-energy physicists don’t even look at the paper journals anymore, only the Internet sources. But on-line papers are not peer reviewed, and peer review is the quality assurance of the scientific enterprise; furthermore, peer-reviewed papers are what get you tenure! The future of on-line journals is a big question – how do you combine the rigorous pre-publication scrutiny of peer review with the instant dissemination of your work? After-the-fact reviewing might be possible if it were true that the number of times a paper is downloaded is a useful measure of the quality of the paper.

You can shop on the Web, through many companies that have been set up in the last year or so for this purpose. Even the Home Shopping Network is available! You can buy all kinds of computer products, of course, you can even get free software demos. But you can also buy cookies, or even lingerie. To buy these products, a credit card is generally use, not because it's the most efficient payment method available, but because it’s the only one. Credit cards do not provide sufficient security, they don't facilitate microbilling, nor do they provide anonymity.

There's a problem with security because, unfortunately, the Internet is quite a leaky channel: the skills needed to tap into somebody else’s Internet transactions and steal credit-card numbers aren't very rare. And you might want to buy a lot of very cheap things – if you look up something in the Encyclopaedia Britannica on-line, for example, they might charge you a few cents, and credit cards don't work well with such small transactions. Anonymity will be increasingly important; the problem with electronic transactions is that people are going to figure out who you are, put you into a database, and sell you to marketers. Along with measures to make U.S. currency more difficult to counterfeit, the government is thinking of printing bar codes on our money. I, for one, don't like the idea of somebody scanning my bills and finding out everywhere I go and everything I buy. Several companies are trying to market the idea of electronic digital cash, known as cryptocash, that's secure, comes in small denominations, and is anonymous. Once there's trusted electronic cash, people will be able to start businesses on the Internet very easily, selling custom products to a global market with very little start-up cost.

Closely related to the question of security of information is the issue of encryption, which is a topic of heated discussion these days. The essential question is whether the government has the right, when sanctioned by a judge, to “wire-tap” a computer in the same way that the law allows telephone taps. We must decide this soon, because technology is rapidly taking over. Software to produce military-grade, unbreakable encryption is already available on the Internet for free. The system works like this: you make up a “private key” – a phrase that you never tell anyone, that's between you and your computer. Your computer then converts this private phrase into a public phrase, or “public key”, which is a sequence of apparently random characters. You can't go backwards – you can't turn a public key back into a private key, even using all the computing resources in the world for the age of the universe. The public keys are available to everybody. Now, let's suppose that Angus wishes to send a message to Jill. Angus looks up Jill's public key and his computer combines that with his message to produce the encrypted text. The encrypted text is sent to Jill, who uses her private key to decrypt it and get the message back. The private keys never move across the Net, so nobody can intercept them. The only way for someone to get your private key is to look over your shoulder as you type, or to steal it if you're foolish enough to write it down. The government is trying to outlaw this kind of software – it's treated as munitions under some circumstances – precisely because they can't break it. But as the more anarchic citizens of the Net like to say, “If privacy is outlawed, then only outlaws will have privacy”.

There are other legal issues as well. When Gutenberg invented the printing press in the 16th century, one of its first uses was to produce large quantities of pornographic woodcuts. The same vulgar objectives are fulfilled by any new medium, including, of course, the Internet. In October 1994, Carnegie Mellon University decided to censor the Usenet feed, some of which contains obscene material. There was an uproar in the campus community over free speech. The question here is whether the Usenet feed is like a telephone company or a television station. Ma Bell is a common carrier and isn't expected to censor its traffic – you can say anything you like on the phone. Whereas a television station partakes of the limited resource of radio bandwidth, and therefore is held responsible for its content.

It's also possible to send e-mail and contribute Usenet postings completely anonymously, which can lead to very frank discussions – people can say things on the Internet that they don't say anywhere else. People don't take any responsibility for what they say anymore, and the few are spoiling it for the many. In so-called flame wars, people try to be as vicious to each other as they can with just words. In an extreme case, last Thanksgiving two journalists on Long Island not only had their e-mail "bombed" – that is, their e-mail mailboxes were filled up with rubbish – but the attackers also got into the phone company's computers and redirected the victims' incoming calls to an answering machine containing an obscene message. The Internet operates on the honour system, and if you flout that you can do people a lot of harm. But preventing this sort of thing is difficult when there's encryption and anonymity. If you can't see what's being moved across the Net, and you don't know who sent it, how can you possibly decide whether it should be there or not, and, if not, how to stop it?

You can be your own publisher: it is becoming easier and cheaper every month to set up a Web server. The Internet will make disseminating individual artistic expression easy, and we will have access to information that can empower us. It is, perhaps, the gateway to a great new virtual culture. We can expect journalists – and ordinary citizens! – to report on their findings from raw data, rather than pre-digested information. The bright light of media attention will become more penetrating, causing honesty in reporting the facts, but also more scope for fallacious statistical arguments. Retailers will adjust to the new medium, enticing us into their cyber-stores with giveaways of information, "frequent-visitor programs", and advertisements with ever-fresh "eye-candy" pictures. The less-exciting alternative is that we'll be forced to spend time in some awful virtual space where we won't meet anybody, and every now and then our path will be blocked by an advertisement for laundry detergent or a car, and we'll have to wait for it to finish before we can continue.

Individuals who have the technical and creative abilities to do so should try to put something of themselves on the Internet, before it gets taken over by relentless corporate agendas. We must not simply cocoon behind security gates, with our computers and lots of software, having nothing to do with the nasty cold real world outside. We must use the Internet to build a virtual community and explore what it can do. But we must make sure that it enriches the physical community and real meetings between people, rather than replacing them.

Dreaming of Hypermaps (1996)

I'm writing this on a laptop on an airplane, watching the Oklahoma landscape roll smoothly by. There's a lot to see, in my opinion, from airplane windows – geological faults and rock unconformities; the way water erodes rock, and how roads, farms, and cities form themselves around the resulting watercourses. Sometimes, with raking sunlight and a dusting of snow, you can see ancient villages or medieval agriculture (though not in Oklahoma). I love the nonstop from Los Angeles to London, looking over endless, endless Arctic Canada and the mountains, like broken bottles, that cut through the Greenland ice sheet.

With a little imagination, such a landscape springs forth from a paper map, especially a finely detailed, large-scale one. For this reason, I have always enjoyed armchair traveling with the aid of a map; this is especially fun before a trip, and occasionally more fun than the actual trip. Thus, the favourite maps in my own collection are those that represent lands far from my own experience. I like to say the names to myself, to wonder what happens when that tiny road simply ends in the middle of a jungle, to speculate about who uses the quay that gives access to a tiny, Atlantic battered island. There are others who share this passion: in the meat-market district of Manhattan there's a cafe whose walls are covered in old street plans of cities from around the world, stuck up with thumb tacks. (The time I was there I walked from map to map, peering at them over the heads of other customers, who had to lean aside to get out of my way.) A map can add colour to a book or newspaper story by showing where a battle was fought, or a train derailed; where the world's rice is grown, or the territory of a vanished empire stretched.

Old maps can be a lot of fun too. A few years ago, I was living in Oxford, quite close to Holywell Church, which is on Holywell Street. After a few exploratory sessions, I could find no evidence of a well, holy or otherwise. This seemed like a challenge, since it must have existed at some time, so I decided to try to nail it down. I wheedled my way into the Map Room of Oxford University's Geography Department and dug up some town-planning maps from 1862. These maps were at the scale of 1:1250- in other words, an inch on the map equalled roughly a hundred feet on the ground. At this scale you could see everything! Next to the church, at a distance equivalent to perhaps 20 feet, the map was annotated "Ancient Well" in gothic characters. I rushed back to the church to check it out. The ground showed no evidence of anything ancient, just a compost heap. But the churchyard wall contained some extra angles, implying that the builders had been making space for something-presumably the well. It was quite satisfying to feel that a tiny scrap of very unobvious history had been unveiled.

I have dreamed for years of an electronic alternative: owning a map whose centre, scale, and content are determined by me, not by the maker of the map, that I call a hypermap. A hypermap would integrate data from many different sources, making the information available to professionals and amateurs alike. I would enter the map by selecting a point and having the scale double. After ten iterations, my point of view will have been reduced from a flyover of the globe to a ramble in the countryside. Inevitably, my Net provider would sell this information to a database company specializing in travel-related matters. Over the next few days, junk mail (both electronic and paper) would arrive, advertising the joys of an adventure vacation in Australia. Through automated database correlation, yet another company would have narrowcast to me with piercing accuracy, simultaneously helpful and eerie.

Can the concept of an expedition into the wilderness survive? We zoom in closer and closer and the map scale gets larger and larger: what happens when we let our imaginations play with this system? Suppose the map becomes three-dimensional, like those plastic maps that you can run your fingers over to feel the mountains. Perhaps the viewer will use a virtual-reality helmet to land an airliner, or loop-the-loop in Arches National Monument, or fly through the glacial canyons of Antarctica and over Everest.

An existing system of some interest here is the "Virtual Los Angeles" project – created by William Jepson and the rest of the Urban Simulation Team from UCLA's Department of Architecture – a virtual model of large tracts of the city, complete with trees and graffiti.

Graduate students shoot video footage of the streetscape, which is fused with satellite and mapping data into a seamless, realistic, textured urban landscape. The user can then “walk” around within the model by means of a mouse, a joystick, or a virtual-reality helmet.

And the data aren't just digital photographs, but the outputs of other sensors that have nothing to do with visible light and work instead in infrared or microwave frequencies. There's a strong analogy here to astronomy, which was confined to optical observations until the arrival of radio telescopes. Today, a burgeoning family of telescopes observes the entire electromagnetic spectrum, neutrinos, and soon even gravity waves. The invisible emissions captured by these instruments have provided a new view of the universe, revealing it to be a violent and capricious place, in sharp contrast to the quintessentially perfect “music of the spheres” of the medieval imagination. In the same way, the wide availability of high-resolution geographic data will change our view of Earth, making it at the same time more familiar, more mundane, more complex, and more precious.

One source of such high-volume data is Synthetic Aperture Radar (SAR). Parts of Earth that had previously been difficult to see with visible light because of almost continuous cloud cover, can now be seen clearly by radar. SAR can see through clouds, vegetation, and sometimes even a few meters of sand. In the Andes, volcanoes have been discovered that were previously unknown, due to their inaccessibility at ground level and to being shrouded in clouds and fog when looked at from above. SAR can see deep enough into sandy desert to discover an ancient ghost city on the Silk Road and can espy eco-friendly farming taking place beneath the canopy of the Amazon rain forest. SAR can measure the moisture content of Kansas cornfields and differentiate spruce from birch in the Russian taiga. SAR can trace the movement of Chilean glaciers, document the destruction of African gorilla habitat, probe the geology of Hawaiian volcanoes, determine the vintage of Antarctic sea-ice, and monitor the recovery of Yellowstone from forest fires.

In order to see so much so clearly, there is of course a price to pay – the raw data from the satellite are not directly visible, but need to be processed by a supercomputer before becoming intelligible. Such a project has been under way at Caltech and JPL for two years now, and it's a massive endeavour involving many people. We feed Intel and Cray parallel supercomputers with tapes of raw data and receive multichannel colour images in exchange. Every pixel in a colour image conventionally represents three data channels, encoded in the colours red, green, and blue, which correspond to the three kinds of receptors in the human eye. But SAR takes data at eight or more channels, leaving a choice of how to throttle the flow down to only three. Such filtering choices can be made to emphasize different aspects of the terrain; for example, to identify types of trees or the composition of volcanic lava, or to gauge the quality of potential ski slopes.

For the armchair explorer, part of the exhilaration of this remote-sensing data is that it has not been processed and digested by a human, only by a computer. There may be the ruins of a hidden city, barely visible without contrast enhancement, or a lake forming where none was known before. "Traveling" by means of SAR data carries the possibility of discovery, much like that offered the patient comet-watchers, roving the sky with binoculars. By comparison, making the trip by paper map will feel like looking at a star atlas instead of looking at the sky. SAR data are not simple pictures to be examined but can be reprocessed in many ways – just as statistical data can be massaged and processed to highlight, emphasize, and maybe even cheat. When we combine SAR data with conventional maps, we can see correlations and associations that were previously hidden, thereby creating knowledge, and – who knows? – perhaps a scrap of what we all crave: insight.

With my colleagues Thanh Phung and David Payne of Intel Corporation and Ellen O'Leary of JPL, I am developing a pilot version of a World Wide Web-based hypermap, called SARA, for Synthetic Aperture Radar Atlas. SARA actually lives on a supercomputer here at Caltech, but you can get to SARA's Web site via an ordinary Web browser, such as Netscape. SARA welcomes you with a map of the world, on which you click in the general area you wish to visit. This in turn brings up a closer view of that region, in which the available SAR data sets are highlighted in red. Clicking on a red zone brings up a "thumbnail" black-and-white SAR image. These images are compressed eightfold from the actual SAR data, meaning that the smallest details one can see in the thumbnail image are eight times bigger than the smallest details one can see in the actual data; similarly, the colour channels of SAR data are replaced by one channel rendered in shades of grey. Thus the volume of data to be transferred to your computer is a mere 1/512th of the actual SAR data set – a necessary concession to the speed of the average modem. If you're directly connected to the high-speed part of the Internet, you can then call up the real SAR images, set the three colour channels to show you what you want to see, and zoom in to bring up the spatial data that got compressed out of the thumbnail version. This is currently unrealistic for home or high-school use, but soon, we netmongers hope, higher-speed networking and even-faster cheap computing will make its way to the domestic hearth. SARA was demonstrated at the Supercomputing '95 conference in San Diego last December, running on an Intel Paragon machine, and, I think, was received with some interest – we had people standing in line to see our show.

The technical path to the hypermap is fairly well marked at this point. There are enormous challenges awaiting in database management, and in designing means of exploiting and presenting the data that will be useful to the expert and novice alike, bur the routes to solving these problems look reasonably clear. I hope that these issues can be resolved, and the hypermap brought to fruition, because it will deliver information that is not only useful but inspiring, enabling individuals to see the world through their own uniquely personal maps. In earlier times the translation of the Latin Bible into the languages of the common folk allowed fresh views on Christianity; now hypermaps will allow people to choose how raw data are to be processed and delivered to them, thereby minimizing the distorting lenses of those who would "interpret" the data for us.

Social Media

The internet started, in my mind, with Tim Berners-Lee building the HTML language for writing documents with hyperlinks and the HTTP protocol for requesting and receiving such documents. It was a real advance to have a complex system documented this way, so you could just click to get more information, rather than the old way, of looking it up in the index, pulling down another ring-binder from the shelf and finding the page number. By 1993, the Mosaic browser was available, and the HTML files were not just text but also images. In 1994, I was thinking about the potential of the internet for distributing scientific data, that would become a mainstay of the rest of my career. While people loved clicking on pages in a browser, it was clear that there should be a way for a machine to fetch data from another machine, and I was part of a group making such tools.

Companies like AOL attempted to make a sanitised internet, a "walled garden" so that people could stay safe by staying the AOL space. Wikipedia started up, and we found that we no longer needed six feet of shelf space for the encyclopaedia. People started communicating by email, and the hyphen disappeared from the word "e-mail". The movie *You've Got Mail* came out in 1998. In 2000, I remember an article in the New Yorker called *You've Got Blog*, about a new kind of web site (Blogger) where non-technical people could upload thoughts, their "web log". I started using a blog system called Livejournal, writing about life and exchanging thoughts with others. At first my friend network was just people I had met in person, then it grew to include friends of friends. What a joy it was to meet somebody in person, but to already know so much about their inner thoughts.

In 1999, recently divorced, I was wondering about taking dance classes in order to meet women, and a friend suggested using the internet. It turned out to be the perfect medium for my search – at that time. There were not that many people using this strange new idea, and there was a pejorative sense that meeting through the internet was not like a "real" meeting. But also it was like an exchange of letters, not today's swiping at Tinder. It was like meeting somebody as a pen-pal; further it was only the clever and technology-savvy who were using the internet, and of course that's exactly the sort of partner I was looking for.

In 2005, the term microblogging was coined, the idea that people want to say tiny snippets of a thought instead of a paragraph. Then came Twitter and Tumblr and all the rest. Facebook was started for college students only, but soon opened up to everyone. When every Facebook friend was a real friend, it was a marvellous place to keep up with people, see their photos, keep the relationship alive. Soon enough there were more advertisements than real content, and then came the "influencers", people paid to advertise products, and the friend content started to be drowned out.

There was a feeling that the internet empowers citizens, that the extra communication it affords – over and above the telephone and newspapers – would transform the way people interact and render them proof from crackpots and authoritarians. Society holds together through shared values: social trust, strong shared institutions, and shared stories. But during the 2010s, people were less sharing experiences in the way that strengthens friendships, but rather building an imaginary "influencer" world of performance. Then came the "algorithm".

In the beginning, there was a chronological sequence of the posts from friends, then came the ideas of "share" and "like" to amplify content into friends-of-friends. The tech companies were able to gather data about what its users wanted to see and replace the chronological feed with what they are most likely to like/share. It turned out that generating emotion is what the algorithm found: especially anger, especially anger at out-groups like immigrants and civil servants. Users learned to make a post that "goes viral", and the way to do that is dishonesty and enraging the mob. People started to coalesce into "bubbles", each with their own truth that the other bubbles cannot comprehend. Democracy has become attacked by lies and nonsense that drives voting. It is no longer a matter of citizens thinking through policy positions, rather they respond to three letter phrases like "Get Brexit Done" and "Stop The Steal". Democracy depends on shared trust in institutions, but social media cuts that away. Evey election is a desperate struggle to stop the hated other side from taking over.

Twenty years after the start of microblogging, people spend time scrolling through tiny micro-videos one after another in endless sequence; they go on holiday and take pictures of themselves in bucket-list spots in order to show off on their insta feed. People with very strange views easily find community with others of the strange views. In the old days, there was just one account of the news of the world, and an honourable spirit of integrity in the journalism profession, a search for truth rather than clicks. But increasingly, everyone gets the version of the news they want, and even facts are impossible to agree on. These trends exacerbated by the smartphone so everyone can be online at all times. Society has not been educated by the internet, but rather split into self-satisfied, warring factions.

8. Representing Information

A World Made from Bits

Computers operate with binary bits, 1 and 0, it is all they know. Or on and off. Or yes and no, whatever. But these things mean nothing without context. The yes and no only come into focus when we know the question: “Will you marry me?” or “Shall I supersize that for you?”. So how can computers deal with this context, this ***meaning*** and ***knowledge*** that humans want along with their bits?

At school, when we learn about binary, it is in terms of numbers. That the number 5 in binary is 101 or 1x4 + 0x2 + 1x1. We might learn how to do logic gates – AND, OR and NOT applied to binary bits, and how to combine those operations into arithmetic, how to add a pair of numbers expressed in binary, to give a result, also expressed in binary. The reason computes use binary is another kind of duality: between a switch and a number. The electrical idea of a switch can represent the mathematical idea of a number, 0 or 1.

Electronic components such as capacitors and transistors can be combined to make the logic gates. The next step is the so-called floating-point number, like the scientific notation 2.99×10^8, which a computer might express as 299×10^6, a combination of the integer 299 and the exponent 8, together with a sign + or -. Computers actually use binary numbers for the integer and exponent.

Arrays

When I started computing, in the 1970s, the Fortran language was the high-level language, and it offered not only numbers, but the very powerful ability to utilise *arrays* of numbers. If you are simulating, for example, the heating of a 100-centimeter metal bar, the array can represent the temperature at each centimetre along the bar, and the time-evolution of the system would be implemented by changing the values in the array. At the top of the program, it would say

```
DIMENSION TEMP(100)
```

In those days, computer code was in capital letters, and the statement doesn't say if TEMP is integer or float – that is implied by the first letter, where `I`, `J`, `K`, `L`, `M`, `N` are presumed integer and others float.

But it was a breakthrough: the language gave control over an arbitrarily large amount of memory, to run a large simulation. Also the semantics of the array have been used for centuries as the calculus of finite differences, written by the English mathematician Brook Taylor in 1731, about how to approximate continuously varying quantities, such as the temperature of the bar, using only a finite number of samples. The binary bits of the computer are regimented together, using the array concept of Fortran, into a form that mathematicians and engineers can use. Finally, we have computers expressing meaning and knowledge.

Fortran assumes that the computer running the code is a single processing unit, doing instructions one by one. Suppose we want to compute something based on the simulation temperature in the bar, let's say the sum of all the elements. In Fortran we would write:

```
      TOTAL = 0
      DO 30 I = 1, 100
      TOTAL = TOTAL + TEMP(I)
30    CONTINUE
```

You notice the sequence of the operations: first add `TEMP(1)` then `TEMP(2)` and so on, an inherently sequential operation. But we all know from elementary school that things can be added in any order and get the same result. Imagine we had four processors instead of one – then we could have one processor add the first 25 elements, while simultaneously another processor adds up the next 25, than at the very end add the four totals from the four processors. This way, the computation is four times faster. Modern machines often have many processing units (cores) to boost performance, so this inherently serial processing of the code above is now replaced with something like the Numpy library; in modern Python code the above would be:

```
import numpy
total = numpy.sum(temp)
```

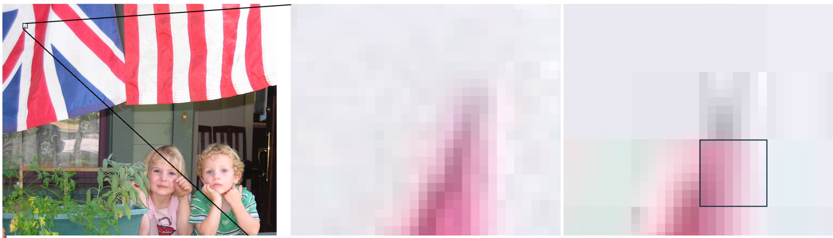
The calculation is not written explicitly and sequentially as it is in the old code, rather the optimised code from Numpy looks at the length of the array, and may parallelise the task. Numpy is of course much more sophisticated than this, and an excellent way to build high-performance numerical simulations such as the fluid flow described in chapter 6.

Back in those ancient times, there were only 256 characters – the so-called ASCII character set. Besides the upper and lower-case alphabet and the digits, there were ! and # and ()[] and * and several more “non printable” characters like “line feed” and “escape”. One we discovered at school was the “BEL” character that simply makes a bell-like noise, to the annoyance of anyone else in the room.

In the old days, computers were used for computing – numerical formulae for simulation. Ballistics of things shot out of guns was a popular application, where the air resistance plays a crucial role in the range of the weapon. Because the application domain was so blinkered, the chairman of IBM, Thomas Watson, declared in 1943 that the world market was “maybe five computers”. But then people discovered a whole new world of data outside numbers: databases, images, video, structured documents.

Images, Music, and Movies

Much of the data that fill our modern day computers and data streams is images, music, and movies. Let us start with representing an image to a computer, since the other two have the same characteristics. The image begins life from an array of light sensors, invented by Willard Boyle and George Smith in the late 1960s. These “charge-coupled devices” have replaced analogue photography in the way that vehicles with engines have replace vehicles with horses. When exposed to light, each of the sensors might provide a 4-byte number for the brightness in each of three bands, red, green, and blue (RGB), so that an image 4096x4096 would be stored in a 200 megabyte file. But the key observation is that a compressed image looks just as good for most purposes, and the magic of the Fourier transform allows great compression to be done quickly.



8-1 (Left) original image (Middle) High quality JPEG magnified (Right) Low quality JPEG showing 8x8 blocks.

The idea of JPEG compression is to remove the data-heavy parts of the image that the human eye doesn't need. First there is a transformation from the RGB planes to a monochrome brightness plane and two chroma planes: the eye is very sensitive to errors in the monochrome plane, but a lot less sensitive to the chroma planes, so they can be compressed more. The image is split into 8x8 blocks of pixels; The 8x8 pixel block is Fourier transformed, meaning that smooth variation is separated from sudden variation. Usually, the sudden variation components are small and can be simply ignored without changing the perception of the image by the human eye, since in a natural scene, each block will vary smoothly in brightness across the block. This means the compression will be very effective. When an image is saved as JPEG, there is selection of a "quality" number: if it's 100 there is no loss of information and the file size is not reduced, but if the quality is low, so the compressed image file is very small, at the expense of quality. Figure 8-1 shows the effects of the JPEG compression. The low-quality cutout on the right clearly shows the 8x8 blocks of the algorithm, but the higher-quality cutout in the middle is much smoother. The higher quality image is eight times larger than the lower-quality image.

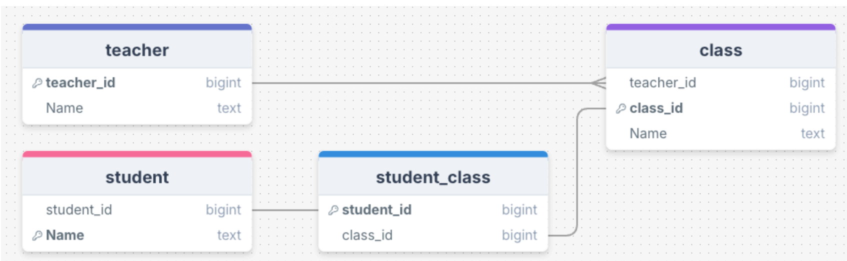
This kind of Fourier-based compression is a crucial component of our culture of sharing images, listening to music and facetiming on the go, streaming movies, and so on. Imagine if Jean-Baptiste Joseph Fourier could be shown the results of his ground-breaking mathematics, 200 years later!

More About Databases

These days, computers do much more than compute with numbers. Non-numerical data had already been stored in various ways for many years, when Edgar Codd described the relational model in 1970 – a breakthrough in the efficiency and effectiveness of storing a wide variety of data. Suppose a school has a list of classes, each with a teacher, whose name is written into each record in the list. If a teacher has two classes, and somebody updates that record, mistakenly changing the name, then a teacher can have two names! Similarly, if a teacher is teaching no classes this semester, then that person's record disappears completely. The answer, according to Codd, is to have two lists: one of teachers and one of classes, and neither of those problems can happen again. In the list of classes, there will be a *reference* to a teacher, what Codd calls a *foreign key*. Whenever there is a foreign key, there must also be a record that it points to, in other words, each reference to a teacher in the classes list must have a unique corresponding entry in the teacher's list, the so-called *referential integrity*.

These days, relational database software knows all about the foreign keys; if the teacher is deleted, then all the classes that teacher teaches can also be deleted. This latter behaviour is when you set `ON DELETE CASCADE` when setting up the database. Sometimes it makes sense to delete all the "downstream" records, sometimes not. Perhaps this is not the right behaviour for the classes/teachers application, since when a teacher leaves, the class will survive, but a different teacher teaches it. But when you try to delete a teacher record, the database will tell you about all the classes that need a new teacher – so it can satisfy its foreign key constraints.

The relational model includes a powerful language: SQL or Structured Query Language. SQL allows a wide variety of queries and updates for tabular data. Suppose for example there is a table of students, a table of classes taken by the students, and a table of teachers who teach classes. Each student can take many classes, and each class has many students, so there is a "helper" table also expressing if a given student takes a given class. This can be expressed with an "entity-relationship" diagram, see figure 8-2.



8-2 Relationships between tables in the SQL model. A teacher can teach many classes, each student can take many classes, and each class has many students.

Each teacher can teach many classes (one to many relationship). Each student can take many classes, and each class can have many students (many to many relationship). The latter is implemented with a

JOIN table called “student-class”. The SQL language can be used to extract useful information, for example how many students are there in total:

```
SELECT count(*) from students
```

Or what classes does each teacher teach:

```
SELECT teacher.Name, class.Name from class
NATURAL JOIN teacher
```

I don’t want to go into a long tutorial on the SQL language. It is a rewarding language to learn, and `khanacademy.org` has a marvellous free course. SQL and databases together power the modern economy, where tables can have a billion entries. The SQL language is an intuitive, easy to learn language for manipulating data, as well as a good way to describe the structure of the data through schema. SQL is *interoperable*, since every database administrator uses the same basic language. It a *transactional* system, so either the whole operation completes or none of it – you wouldn’t want money taken from your account but the goods not purchased! Constraints and built-in checking ensure accuracy and reliability, and the data is at all times consistent.

Databases for astronomy

Databases and SQL were a marvellous achievement of the 1970s and 1980s. They started to be used in astronomy in the 2000s, with the first of the big sky surveys, the Sloan Digital Sky Survey. Up until then, astronomers just downloaded the whole dataset to their own machine to mine it, but SDSS was very large, about a terabyte (10^{12} bytes) for the first data release, up to hundreds of terabytes for later releases. Giving astronomers meaningful access to this massive dataset was unknown territory. It was clear that a new model would be used: “bring the computation to the data”. There was a centralised data server, with a website interface, but the question was how it would work in practise. There was the “persistent object” model, favoured at the time by another big data community, high energy physics, where your favourite computer language (C++, Python, Java etc) would access data “objects” and work with them, where an object could be a star with its properties, its lightcurve, its spectrum all wrapped together into a sort of portfolio.

We have already mentioned the *atomic transaction*, the idea that a sequence of operations either fully completes, or the entire transaction fails, and the database is left unchanged. My colleague Jim Gray was a pioneer of modern databases in the 1970s, and recognised the importance of transactions. If you are at a cash machine, the sequence may be to deduct the money from the account, then issue the cash. You don't want this half-completed, where the deduction happens but the cash is not issued! Suppose there is a joint bank account with two debit cards, and the two owners are at different cash machines simultaneously. If one takes out all the money in the account, then the second one should not be able to do the same. The withdrawal of money from the machine either fully completes or nothing happens. Jim defined the "ACID" rules that all modern databases achieve:

- Atomicity, as defined above, where the transaction is not left half done,
- Consistency, so that constraints are preserved: for example if every user record has a separate profile record, then deletion of the user record should also lead to deletion of the profile record.
- Isolation, meaning that if several transactions happen concurrently, the result should be the same as if they occurred in sequence.
- Durability means that once a transaction happens, it stays done, even with a subsequent system failure.

The ACID rules were a real advance on earlier databases, since they allowed programmers to see the database as an abstraction with a well-defined interface. The database is like the car, presenting a steering wheel, accelerator and brake, without the user needing to know how it all works under the hood.

Public Private Cryptography

For millennia, people wanted to communicate in secret, so that anyone other than the intended recipient would not understand the message. Politicians, spies, and countries at war prevented their enemies from knowing what they are doing. Cryptography is the study of how to do this, from the greek work "crypto" meaning secret, and the word "graph" meaning writing. In the information age, we all need cryptography, since it keeps secure our passwords and thus bank accounts and so much more.

One of the earliest encoding method is called the “Caesar Cipher”, used by Roman general Julius Caesar 2000 years ago. It is a substitution of one letter for another, based on shifting the alphabet. Suppose we shift by 4: this number is called the key:

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ
WXYZABCDEFGHIJKLMNOPQRSTUV
```

Therefore the encoding of the word `HELLO` is `DAHHK`. Notice that whoever has the key can both encode and decode the message. The secrecy of the message depends on the secrecy of the key. Centuries later, in the 1970s, the US government published the Data Encryption Standard (DES), where blocks of 64 bits are encoded using a 56-bit key. There were all sorts of worries that the US government had a backdoor, and that the key was too short, leaving the algorithm open to a brute-force attack. But of course the biggest problem was the key itself: if Alice wants to communicate secretly with Bob, how can she give him the key without a third party getting it? In order to communicate secretly, Alice and Bob first need a way to communicate secretly. Its amazing how it took until 1976 for this problem to be solved!

The answer is "Diffie–Hellman key exchange", also known as public/private key pairs, later made practical by Rivest, Shamir, and Adleman. Instead of a single key used to both encrypt and decrypt, there are two keys. If Alice wants to communicate with Bob, then Bob makes up a public-private key pair (details below). Bob keeps the private part secret, and lets Alice – or anyone else – know his public key. Alice can encode her message with the public key, and send it to Bob, who can use the private key to decode it. Notice that no secrets need to be communicated in plain text, only encrypted data is communicated. In order for Bob to reply, Alice needs to make her own key pair and advertise the public key.

The algorithm is based on the idea of a “trapdoor function”: something easy to compute in one direction, but very difficult to invert, unless you have some extra information. RSA relies on the difficulty of factorising large numbers. The computational complexity rises fast with the number of digits in the number: in 2019, a 240-digit number was factored using 900 CPU-years of computing power, and the team estimated that a standard 1024-bit RSA encryption would take 500 times as long.

The algorithm relies on finding three integers *N, e, d* so that for all integers *m*:

$$(m^e)^d \equiv m \quad mod\ N$$

so that knowing *e* and *N* (the public key) makes it very difficult to find *d*. The way this is done is to find two large primes *p* and *q*, then multiply them to form *N = pq*. The key-length of the cipher is the number of bits in N, let's say it's about 1024. The choice of *e* is somewhat arbitrary, it can be 3 or 65537, or some convenient prime between them. If the factorisation of *N* as *pq* is known, then it is easy to compute *d*, the private key. Once we have the ingredients for the above formula, and the other numbers can be discarded, the *p* and *q*.

The symbol *m* above stands for Bob's message. It is converted to a number, or a sequence of numbers, each of 1024 bits. For each of these, Bob multiplies *m* by itself *e* times, modulo *N*, using the information in Alice's public key, then transmits it across the wire to Alice. Now Alice does the same operation, multiplying the encrypted message by itself *d* times, modulo *N*, and then she has the decrypted message.

It turns out that all these multiplications and powers can be done in seconds, rather than taking the age of the Universe, as would be required to decrypt by brute force. But even so, the RSA algorithm is not nearly as fast as a substitution cipher such as DES described above. So what really happens is that the symmetric DES key is transmitted by the unbreakable RSA cipher, which can then be used for the current communication session, then choosing a new DES key for the next session.

Billions of Galaxies

In the old days, people thought the Universe was human sized. A flat Earth with a dome of sky over it, created a few thousand years ago. The size of the Universe was a small multiple of the distance from London to Rome, and the age of the Universe a small multiple of the age of the oral history passed down the generations. It was James Hutton, and later, Edwin Hubble who broadened our Horizons – see Chapter 12. But in addition to deep time and deep space, there is big data. Each astronomical observatory produced photographic plates in the past, kept them in an archive, and visitors could come and examine the plates with a magnifying glass and a light table. But now the sample telescope has a digital sensor, and produces gigabytes of data. This is then available on the internet for fruitful joining and matching with data from other observatories.

The Virtual Observatory

In 2001, the Virtual Observatory came to be, the idea of standardising and unifying the way astronomical data is stored and presented, part of an attempt by government to understand how all digital data can be made more useful to society. So there are billions of catalogued stars and galaxies, and millions of digital photographs.

You might think it's a simple question – find all stars or galaxies that are near to a given point in the sky, or find an image of that part of the sky. But then its like ordering a ham sandwich in the USA: what kind of ham, what kind of bread, do you want cheese, what types of cheese do you have, and so on ad tedium.

8:14 am - I say Redshift, you say Banana. Sitting in a classroom in redbrick building at Harvard, discussing interface protocols. Trying to get agreement on a standard about how to ask questions of an astronomical database. The extra-galactic people want the query word to be "redshift" and those that study our home galaxy want it to be "doppler velocity". Then somebody says lets generalize and suddenly let's just allow the customers to ask any question they like and the poor old implementers will just have to make it all work. Then somebody asks about if the computer understands the meaning of the word "redshift", and so the discussion falls once again into the black hole of semantics and can the computer be said to understand anything at all. Happy days!

The Sloan Digital Sky Survey

It was in 2000 that the Sloan Digital Sky Survey (SDSS) begun: a real revolution in astronomy. A telescope in Arizona made a wide, deep survey of the entire sky with digital, rather than photographic, results.

Astronomers could find the data online, wherever they were, no more travelling across the world and using a jeweller loupe to examining photographic plates. SDSS also ran source detection algorithms on the pixel data, to make a calibrated, high-quality catalogue of stars and galaxies. But with all this data, there was a need for backend technology to handle its delivery to the users. Alex Szalay was the pioneer in the field of data-driven astronomy, one of the first to see the so-called fourth paradigm of science[5]. The first three paradigms are experiment, theory, and computation, with data exploration as the fourth. There were a lot of trial runs and experiments with different data technologies, ways to allow scientists great flexibility in the questions they could ask of the dataset, while at the same time being secure, efficient, and reasonably easy to learn. Astronomers learned to build SQL queries which would run in the background, with an email notification when it was finished. Then further queries could select on that table, until the refined product was small enough to download. I think of the process of mining diamonds: hundreds of tons of rock crushed and processed through multiple refinement stages until all that's left is a few precious gems.

Jim Gray was a titan in the world of databases, he advanced both theory and practice, and made the first successful commercial database, pioneering e-commerce, online ticketing, and cash machines. Many of us in the early 2000s were building web systems to cover the Earth – my small version was SARA in 1997 (Chapter 7), while Jim and colleagues built the far more capable Terraserver a few years later, using the Microsoft SQLserver technology, enabling the zooming and panning over much of the Earth; it became Bing maps, while Google developed their own mapping system launched in 2005. I made a system called Virtual Sky that utilised much of the technology from Terraserver, and in the process became great friends with Jim. The underlying image data was a scan of the Palomar Sky Survey called DPOSS, and it was the first time I started working with a terabyte sized data set (10^{12} bytes). I didn't have the smooth zoom that Google bought for its mapping service, just a click to double the resolution; but still it was astonishing, a sort of vertigo, to start with an entire constellation, and click down and down and down to

[5] https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf

a lensed galaxy that is just a few arcseconds across. Previously I had seen pictures of the 20 most famous galaxies, but Virtual Sky let me see how there are millions and millions of galaxies. When I was young, I saw dinosaurs as skeletons in a museum, but when the film *Jurassic Park* was released, how amazing it was to see “real” dinosaurs moving in a herd and feeding. That is the change of perception that you get from a system like Virtual Sky.

Jim Gray and Alex Szalay became firm friends and a very productive partnership. Jim marvelled that the astronomical data was “worth nothing” (by which he meant no pesky security regulations), and Alex was happy that the SQL language was the bridge that was needed to connect science users to the dataset. The SDSS database ran on Microsoft technology.

Astronomical data facilities across the world saw the success of SDSS and started putting their own data into a similar framework. In the UK, the Wide Field Astronomy Unit was started by Bob Mann and Andy Lawrence to serve up the digitised plate libraries of the Royal Observatory Edinburgh (ROE). In 2017, I would become part of ROE (see Chapter 12).

Citizen Science

I was at a meeting at Johns Hopkins University in 2007, when Alex Szalay came in full of anxiety, to say that Jim Gray had gone out to sea on his boat the previous day, to scatter his mother’s ashes, but had not returned. Jim was a keen and competent sailor, and we all assumed he would be back soon. It was a clear day and a calm sea. But the day wore on with no news, and we wondered how long Jim could survive in the open ocean. Because Jim was so well-known in Silicon Valley, there was an unprecedented effort. Thousands of human volunteers signed up to the Mechanical Turk system to scan satellite images for evidence of the boat. The Digital Globe system made their data public, Amazon servers held and distributed the satellite data, with Microsoft, Google, and Oracle working hard to get the data collected and available quickly. The Mechanical Turk team worked tirelessly. But in the end, Jim was never found.

It was just a few months later, based on the immense effort that so many people made, Chris Lintott at Oxford worked with Alex and the Baltimore-based SDSS team to build “Galaxy Zoo”, a way for ordinary people to contribute to the scientific enterprise from their own internet browser by evaluating the shapes of the millions of galaxies in the SDSS catalogue. It was tremendously successful! In that same meeting room the next year, I remember Alex exclaiming “the servers are melting!” about the project. The key observation is that for the first time in the history of science, there is more data than an individual scientist can examine, due to the automated nature of the data collection and delivery. As Lintott put it "One advantage is that you get to see parts of space that have never been seen before. These images were taken by a robotic telescope and processed automatically, so the odds are that when you log on, that first galaxy you see will be one that no human has seen before."

One of the signal results of Galaxy Zoo was “Hanny’s Voorwerp”, or Hanny’s object in English, found by a Dutch schoolteacher Hanny van Arkel. It is a green blob close to a spiral galaxy, and was quite a surprise to the astronomical community. It is now thought to be a kind of light echo from intense activity in the associated galaxy, and there are now several more of these, known as Voorwerpjes.

The galaxy zoo of astronomy evolved into a large number of “citizen science” projects in all fields of the arts and science – history, medicine, literature, climate and many others. It was the loss of Jim Gray that showed how much the citizens want to contribute in a worthwhile way. A worthy tribute to Jim.

Glass Ceiling

In the old days of Caltech, it was a professor whose job was to teach undergraduates and also did research with a handful of graduate students, perhaps a postdoc or two. The small team would run experiments, analyse results, and write the papers. Following the model of the ancient Universities, the professors together – the Faculty – would run the whole place, with help from the Staff, who maintained the equipment, administered the computers, and swept the floors. Throughout my career there, I was a staff member, not a professor.

But by the turn of the millenium, there were much larger collaborations with multiple institutions, and my boss at Caltech, Paul Messina, had spun off a quasi-independent, research-focused organisation, the Center for Advanced Computing Research, with its roots in the parallel computing group of the 1980s and 1990s. While there were many fruitful collaborations with professors, the person in charge did no teaching, did not fit the old model, and was not a professor. By 2009, I was what they called “entrepreneurial staff”. For ten years, I had been running the Caltech part of the Virtual Observatory project, writing grants and getting awards to support myself and others. But those awards had to be signed off by a professor before Caltech would submit them to the agencies. I worked with a professor to hire people, paid by the Virtual Observatory money, let's call him Professor X. But there was a glass ceiling for me, which was that you had to be a professor, and teach undergraduates, in order to rise further.

The wind changed direction with a change of Provost, who decided to put the entrepreneurial staff back in their place, and go back to the old cottage industry, where there must be a professor in charge of everything. Professor X declared himself in charge of my award money, and decided to get his own people in. It was an anxious time, I was scrambling for a job, and went on a few interviews: Berkeley, Seattle, etc, but no offers. My rescue came from my old friends at LIGO, who I had worked with ten years before (see next Chapter). The National Science Foundation, who funded LIGO, were demanding an "Open Data" system, and my experience fit well with that.

A few years afterwards, Caltech resolved the glass ceiling with a new type of position called a “Research Professor”, which gives a place to the entrepreneurial staff member who does research but not teaching.

9. LIGO: An Observatory Based on Faith

I worked for LIGO at Caltech between 1998 and 2000, then again from 2010 to 2017. An astonishing enterprise: searching for astonishingly tiny signals from astonishingly distant collisions of black holes. Many people saw it as a fantasy, a waste of money, but it was sustained by faith in technology, faith in science, and by faithful disciples in the US government. Then after 40 years of effort, the signal was detected, a miraculous annunciation from a billion light years distant. Our faith was vindicated, and our understanding of the Universe expanded. I was so happy to be part of the epic quest! Direct detection of gravitational waves by the LIGO project is one of the signal scientific accomplishments of the century. And I'll tell you why it's awesome.

The Hill of Difficulty

The greatest Scottish scientist was James Clerk Maxwell, who in 1865 formulated the beautiful and symmetric equations that control electromagnetic fields. He discovered wave-like solutions to these equations, and the speed of these waves turned out to be the same as the speed of light. Maxwell hypothesised (correctly) that light is a wave the propagates through alternating electric and magnetic fields. He also wondered if gravitation also travelled through a medium, and if so, would have wave-like solutions. Oliver Heaviside in 1893 wrote a field theory of gravitation with wave-like solutions, followed by Max Abraham a few years later.

On the experimental side, there was a peculiar problem with the orbit of Mercury around the Sun: the elliptical orbit did not repeat (as Kepler and Newton predicted), but rather rotated at 43 arcseconds per century. Such a tiny modification of the orbit, but Newton's theory had been paramount for 200 years, and the motions of the planets precisely followed its predictions. Astronomers thought back to the discovery of Neptune in 1846, predicted from perturbations of the orbit of a known planet, Uranus; therefore they hypothesised a planet inside the orbit of Mercury, called Vulcan, that was perturbing Mercury, and there were numerous false detections – because people see what they want to see.

Einstein is the superhero for everyone working in LIGO. He took experimental evidence that the speed of light is always the same, and built the theory of special relativity, then completely rebuilt the theory of gravity in 1916. Instead of Newton's immediate action at a distance, where mass attracts mass, the Einstein theory has the geometry of space itself as an intermediary. Mass distorts space, which causes another mass to move because of that geometry. The big differences from Newton's theory occur when very large masses move very fast, such as two black holes orbiting. Because Mercury is moving quite fast and the Sun is quite big, this is where Einsteinian gravity makes a difference. As soon as the theory was coherent, Einstein tackled the problem of the perihelion of Mercury, and the theory precisely predicted the 43 arcseconds per century. His arcane theory was vindicated! I'm sure he had a huge grin on his face for the rest of the week! However, the rest of the world didn't find out until another prediction was tested: the bending of light by a big mass. It was a solar eclipse in 1919, when stars could be photographed very near to the edge of the Sun (without its blinding light), and Einstein has predicted a distortion of 1.75 arcseconds. He was of course correct: the *New York Times* blared "Lights All Askew in the Heavens!" Since that time, the name Einstein has been a byword for incredible intelligence. People buy toys for their child with names like Baby Einstein in hopes that some of the glorious intelligence will flow into their child.

Gravitational waves were also predicted by Einstein's theory of gravity, but there was a lot of confusion. In 1936 he wrote "If you ask me whether there are gravitational waves or not, I must answer that I don't know. But it is a highly interesting problem". Debate raged on for 20 years as to whether they exist, and if so, whether they carry energy. By the mid-1960s, Joseph Weber made a detector from a 3-ton block of aluminium, and unfortunately made claims of discovery that dogged the field for the next 30 years.

By 1972, a brilliant engineer named Rai Weiss, from MIT, was thinking about using lasers to detect gravitational waves. There would be two mirrors separated by a long distance, and any small change in the distance between them could be detected by optical interference in the reflected light. Weiss wrote the famous paper "Electromagnetically Coupled Broadband Gravitational Antenna", which contained the basic design that would win him a Nobel Prize 45 years later. There was full analysis of all the possible noise sources: laser fluctuations, thermal noise, radiation pressure, seismic noise, electric and magnetic fields noise, and many others.

At the same time, a brilliant theorist named Kip Thorne was starting a research group at Caltech about relativistic astrophysics, prompted by the discoveries in the 1960s – pulsars and quasars – showing the Universe to be much more energetic and violent than the old idea of the timeless, eternal, fixed stars. Thorne was investigating how black holes would form, how they would vibrate; this was well before anyone actually believed the existence of black holes!

There was a dispute about Einstein's equations, which appeared to predict gravitational waves, that could carry energy over great distances. But these are complex nonlinear equations in four dimensions, with a lot of transformations of coordinate systems, and some thought the waves to be merely an artifact of the maths used to solve them. But the world of massive compact objects was given a huge boost by the discovery of pulsars in 1967, by a graduate student at Cambridge, Jocelyn Bell. Radio pulses were shown to be from a very small object, rotating very fast, now known as a neutron star, the collapsed remnant of a massive star. The frequency at which a pulsar rotates is very stable, and therefore they can be used as a clock. More pulsars were discovered in the next few years, one notable example being PSR 1913+16, found by Russel Hulse and Joseph Taylor, using the enormous radio telescope at Arecibo. The pulsar signal is at 17 Hz, but strangely the pulsations are *modulated*, going up and down in frequency every 8 hours. This was interpreted as a *binary* system, meaning two neutron stars orbiting each other very very close – about a million kilometers – much less than the distance at which the Earth orbits the Sun. But there was more: the modulation period was slowly decreasing, as if energy were being radiated from the system. The idea that gravitational waves were being emitted was compelling, especially when further measurements showed other general relativistic effects on the orbit. In 1993, Hulse and Taylor were awarded the Nobel Prize for this indirect detection of gravitational waves. But it also spurred the development and long-term funding of LIGO.

In 1975, due to a shortage of hotel rooms, Thorne and Weiss shared a room, and spent the whole night talking about Weiss' ideas for detecting gravitational waves. Both of them knew of the tantalising discoveries of Hulse and Taylor, and were convinced the waves were out there to be found. Thorne started an experimental gravity group at Caltech to make the interferometer, in parallel with MIT, Glasgow in the UK, and Garching in Germany.

Ron Drever was a brilliant Scottish scientist, intuitive rather than analytic. As a schoolboy he built a television from army-surplus parts, and watched the Queen's coronation on it in 1952, surrounded by family and neighbours. He was head-hunted by the LIGO project in 1979, and made great improvements on the original design of Weiss. He would have shared in the Nobel Prize in 2017, but passed away just months before.

In 1990, Weiss, Thorne, and Drever persuaded the US National Science Foundation to fund the construction of LIGO – The Laser Interferometric Gravitational Wave Observatory. Notice the cheeky use of the word "observatory" – its not just a detector, but it will be able to observe the Universe. Richard Isaacson at NSF was instrumental in the funding of LIGO, a project searching for something undetected and that many people thought was a big waste of money. Astronomers were particularly outraged, thinking it would suck funding from their own projects. One of my colleagues at Caltech said that searching for gravitational waves was "like looking for fairies at the bottom of the garden".

Rai the zealot

After I returned from Oxford to Pasadena in 1986, things settled and I did high-speed flow on parallel computers, but no actual science. The parallel computers were fun, but by the mid-90s it became clear that they were just more powerful ways of doing computation, not exactly a new paradigm. In 1998 there was this wacko project LIGO happening behind the car park at Caltech, where a priesthood of laser engineers were building something that everyone knew would never work. Kip Thorne at Caltech was the theory lead, Rai Weiss at MIT the engineering genius, and Ron Drever the eccentric Scotsman who, like Scotty from *Star Trek*, could build anything electronic. One morning I ran into Rai Weiss in a Starbucks, and it changed my life, got me out of the doldrums and shining with enthusiasm. Rai told me about how gravitational waves would one day revolutionise astrophysics and cosmology, and the LIGO equipment, once scaled up, would detect black holes and neutron stars – the most exotic things ever – not just themselves, but colliding, crashing, coalescing, in a magnificent burst of energy. And the very early Universe (first microsecond) could be detected with its echo still rumbling about billions of years later. The world's most stable lasers, the recycling of light to make the antenna a hundred times larger than its actual size, measuring deflections the size of an atomic nucleus. The possibility that the gold in my wedding ring was forged in the collision of two neutron stars. The incredible energy release of a gravitational wave event, in contrast to the incredible difficulty of detecting it. To a boy from South London, it all seemed so adventurous. Rai was able to make it a valiant endeavour, like the St Crispin's Day speech in Shakespeare's *Henry V*:

From this day to the ending of the world,
But we in it shall be remember'd;
For he to-day that sheds his blood with me
Shall be my brother;
And gentlemen in England now a-bed
Shall think themselves accursed they were not here.

So I started with LIGO in 1998, part of the band of heroes, putting their faith in their leaders, and making their purpose the detection of gravitational waves. Under construction, in Louisiana and Washington State, were two detectors. There were two detectors because an astrophysical signal would be seen by both, whereas terrestrial noise would be only at one detector.

Each observatory has two arms at right angles 4 km long. These huge structures are enclosed in concrete, and inside is a metal tube containing ultra-high vacuum, the largest high vacuum system ever built. Laser light is bounced back and forth along the length of the tubes, reflected from mirrors at each end. The light interferes with itself, so that a change in length – from a gravitational wave – would change light to dark on the photo detector. A simple interferometer would detect length changes whose magnitude is the wavelength of light compared to the 4km length of the beam tube, a strain of less than one part in a billion or 10^{-9}. But this is not nearly enough, the expected strain from a cosmic source is more like 10^{-21}. There is a lot of very clever engineering to make this huge increase in sensitivity over the naïve detector. Much of this came from the ideas of Rai Weiss and the engineering of Ron Drever. LIGO is in many ways the most sensitive instrument in history, and so it must be insensitive to everything that is not a gravitational wave. The tubes are insulated from all kinds of noise: seismic noise, coating noise, changes in gravity from the movement of Earth's atmosphere, noise from the suspensions holding the mirrors, and several other sources.

One source of noise that must be minimised is from the movement of air inside the vacuum tubes. During construction, in 1998, emptying the tubes of air demanded forty days of pumping. The result was one of the purest vacuums ever created on Earth, a trillionth as dense as the atmosphere at sea level. Part of this was the "bake out". Nothing to do with cakes and cookies! Rather it was heating these 16 km of stainless steel tube to 150 Celsius to evaporate the volatiles, so that a high vacuum can be achieved inside. My job was to use the new technology of the internet to build a website and database to show the entire team how it was going at each point along the tubes, and if it was following specifications. This was the first "web-database" system I built, the first of many of the following 25 years. It's such a common design pattern now, the web-database, so common that we hardly remember when it was new.

The Celestial City

Astronomy has traditionally been all about light from the stars – we do not use our ears or noses to find out about galaxies. Telescopes allowed huge steps: Galileo seeing the moons of Jupiter, Hubble understanding that our vast Milky Way galaxy is just a dot in a huge Universe, the images of Pluto from 2015 showing complexity even in the great cold. When I say light, I mean the whole spectrum of light, much of which our eyes cannot see, including radio, X-rays, and infrared, that can show us so much more than just optical wavelengths. Recently astronomy has flourished by adding computers to all that light. For example, computers are much faster than people are at finding differences between images of the sky from different times, exposing transient sources such as supernovas and comets. Launching satellites into space lets us get the light without the intervening atmosphere. The old picture had a tranquil Universe of the "fixed stars" with exactly seven wandering planets; but from modern astronomy we now know the Universe to be a dynamic and violent place full of movement and explosions.

Gravitational waves offer a new way to understand the distant Universe, like a new sense, in addition to our eyes and telescopes. Like all waves, gravitational waves have a frequency – the rate at which the wave goes up and down, measured in cycles per second, and abbreviated as “Hz”. It's the same language as music, where A is 440 Hz. Indeed, the analogy goes further: the frequency band for which LIGO is optimized is just about the same as the frequencies that the human ear can hear. Thus it is simple to convert the LIGO signals into an audio feed, with the detector in Louisiana through one ear and the detector in Washington State through the other ear. The LIGO signal sounds like a lot of noise, mostly, and LIGO has recorded years of it. The computers work hard to find signal in all the noise. When two black holes coalesce into a single entity, their last furious orbits happen at hundreds of cycles per second, generating the signals measured by LIGO. Faster and faster the orbit goes, until each collapses into the other in a great burst of energy, the combined black hole spinning and shivering in a “ring-down” like a sustained note of a piano. When converted to audio, the signal is a chirp, rising in frequency, sounding rather like a bird call.

The analogy with sound carries further, because now, in 2016, we only have two gravitational wave detectors, just as we only have two ears. Humans have a sense of direction based on where the sound came from, which is actually a measurement in the brain of the time difference between the left and right ears. If a signal comes first to the left ear and a millisecond later to the right ear, we instinctively look left so our eyes can try to see where the sound came from. LIGO does the same with computers and telescopes; by comparing widely separated detections, a map of the sky can be made showing where to look for the elusive optical counterpart of the gravitational wave source. Because of modern communication and cooperation, telescopes all over the world can be turned to that probable area to look for transient sources, just as our two ears hear sound waves and we turn our eyes to find it precisely with light. Originally there were only two detectors, a third started observing in 2016: the Italian-French-Dutch Virgo detector, located near Pisa, Italy. A

Japanese detector is coming, and another one in India. If all of these are observing, the sky localization will be much better, and it should be possible to see if there is any light coming from a gravitational wave event.

The more bizarre predictions of Einstein came to be believed, during the 1970s and 80s, including existence of “black holes”, where space-time is bent to an infinite degree. Gravitational waves were also predicted by Einstein; the idea that space itself can be jiggled like jelly, and everything in that space jiggles with it – think suspended fruit. Gravitational waves, unlike jelly waves, travel through space at the speed of light.

We all know one of Einstein's equations, $E=mc^2$, meaning energy = mass multiplied by c^2 (speed of light squared), predicting that mass can be converted to energy. The multiplier, the square of the speed of light, is very large: a single gram of mass, the weight of a paperclip, converts to as much energy as was released from the Hiroshima atomic bomb. Another Einstein formula, less famous, tells how much bending of space is caused by mass: curvature = mass divided by c^2. Here the very large number is going in the opposite direction, so that huge amounts of mass must be concentrated to get appreciable curvature, and that mass must move at huge speeds to generate gravitational waves, such as when two black holes get very close and dance cheek to cheek.

LIGO is not just talking about a black hole, but of two of them getting together. Two black holes joining into one is the most energetic kind of event in the Universe, and yet the theory says no light is produced! There is no “matter” there at all, no dust or rocks or atoms or anything, because the black holes have been spiralling in for millions of years, and have either eaten or thrown off any surrounding matter. Without matter there can be no light or heat, and what is left is just the singular-infinite curvatures of the black holes. That is the theory – but of course the astronomers want to look because if there is an optical counterpart, it would smash the theory and thus be very exciting. The energy release from an event like this is truly titanic – a million times the light and heat from a supernova – and yet invisible to our eyes, except by distorting the positions of background stars.

I should make it clear that there are basically two kinds of black holes: stellar-mass and galaxy-mass, the former being from one to a hundred solar masses, and the latter in the range of a million to a billion solar masses. The stellar-mass black holes and their collisions are accessible to LIGO, with their chirps in the hundreds of Hz, however each galaxy-mass black hole influences an entire galaxy of billions of stars. Other experiments are trying to find gravitational waves from pairs of these 'supermassive' black holes, using radio astronomy or space-based lasers (LISA).

My role in LIGO

I joined LIGO in 1998, working at Caltech in Pasadena. The technologies – using lasers and mirrors to detect gravitational waves – were already 20 years old. Much of the development was with a prototype on the Caltech campus, a 1/100 scale model of the eventual LIGO. The US Government chose to fund the big effort, and by then the two observatories were built, each an L-shape with arms four kilometres long. LIGO has spun off techniques to make very stable and high-power lasers, very smooth sapphire mirrors and special coatings, suspensions to isolate and reduce noise by factors of billions, even quantum manipulation of photons to reduce noise. There are two detectors: Hanford, in the desert of eastern Washington, and Livingston, in the bayous of Louisiana. You can see them both on Google Maps. As noted above, the time difference gives a clue about the location of the gravitational wave source in the sky; but more important is that any astrophysical signal will happen at both detectors, whereas almost all the noise effects are independent and uncorrelated between the two, thus boosting the signal relative to the noise. Another essential component is software – first to simulate the complex black-hole motions to find the precise waveforms that are expected, and then to efficiently detect these waveforms if they are in the detector data.

9-1 One of the 4 km beam tubes at LIGO Hanford. A strange alien structure in the middle of nowhere.

It has been a strange place to work. People were driven by faith – faith that eventually LIGO would detect something. Some people have been working on LIGO for 30 or even 40 years, always knowing that it might end up as a complete dud! Great structures of theory provided estimates of the rates of measurable events, estimates that had massive uncertainties – there might be a hundred events, or there might be just a small chance of a single event. If Nature were unkind to us, we imagined just one stingy, dubious detection in years of running, and destructive squabbling amongst the collaborators about whether to publish. There were early attempts to detect gravitational waves, with false claims of detection, resulting in a contrasting culture of great care and deliberation fifty years later. Of course everyone's heart goes pitter-pat at the thought of a real detection, and so it is especially important that we only find what is there, not what we want to be there. LIGO has elaborate protocols to ensure this.

But Nature was kind to LIGO, and gave us a signal as soon as Advanced LIGO began operation in September 2015. The coalescence of two black holes that was detected was an event of unimaginable power. One was 36 solar masses, the other 29, and it was over a billion light years away. If the gravitational wave power were converted to light, it would outshine the entire Universe! That light, seen from Earth at a billion light-years distant, would have outshone the full moon – the moon that is less than a light-second from us. Just as the atomic bomb energy is the result of converting mass the size of a paperclip to pure energy, the LIGO event converted three times the mass of the Sun to energy! Ten-to-the-33 atomic bombs! And yet that energy is not in the form of light but gravitational waves, making it much more difficult to detect. It took the immense technological achievement that is LIGO, with hundreds of people over decades, to detect and identify this vast energy.

Imagine that our Sun went supernova, a type of huge explosion of a star, where ten billion years of solar output happens all at once (but don't worry, the Sun will not really go supernova). At a distance of eight light-minutes the Earth would survive, but would be blasted to a crisp, all air and water gone, death to all. However, if we are eight light-minutes from the coalescence of black holes, it is quite different; even though there is 100 to 1000 times the energy release of the supernova, the outgoing gravitational waves would have much less effect, more like standing in front of a loudspeaker at a rock concert when the drummer hits hard. Actually the supernova is also putting most of its energy into something we cannot feel – neutrinos, the ghost-like particles that go right through the Earth without stopping. But it also puts out a tiny bit of its energy as light and heat, and that is what we do feel, that is what blows us away. So the supernova neutrinos carry a thousand times the energy of the light – and the energy of the LIGO event is a thousand times that! These huge, invisible energy releases somehow remind me of when two people having a secret love affair – great storms of emotion so powerful that the outside world is nothing by comparison; and yet nothing visible to anyone else.

This detection is not just checking a box on the last unverified prediction of Einstein, it is not just a titanic explosion as big as a thousand supernovas, it is not just incredible technology, but also a new window on how the Universe works. Black holes can spin, and there is information in the direction of those spins, as well as the orbital spin of the binary system. Multiple detections by LIGO will yield the origin story of black holes, if and how black holes evolve and merge to make the supermassive black holes that are in the centres of galaxies, and how our own solar system, our own experience, is influenced and shaped by a population of black holes in the Universe, as yet hardly explored.

The LIGO project is, like all big science, an international collaboration, including major contributions from Britain, Germany, Australia, and India. But most of the cost came from the taxpayers of the United States, through the National Science Foundation, proving US accomplishment and collaboration in yet another area of high technology. It is refreshing to see Government funding devoted to this high-risk, long-term search, that has inspired a generation of our best and brightest, of all genders and races, to give their best, build advanced technology, and create the future.

This enormous achievement came about because of the unique nature of the American science enterprise. All brains are used, including women and non-white people; people are promoted because they do a good job, not because of nepotism or political views; and people feel free to speak up when things aren't right, everyone listened to with respect. Here we have the goose that lays the golden eggs of scientific and technological breakthroughs.

I'll just mention my own contribution, which was (of course) to build an Active Digital Library. The funders of LIGO required that the data, for which they have paid so much, should be released to the world, not just for research, but also for very detailed study by sceptics, to ensure the detections are real. With a small team, I made the website with all the data, together with Jupyter notebooks that allow people to make the computations that make the data meaningful. In other words, an Active Digital Library where the active part is on the user's own laptop using the software we provided. It was quite an advance on the traditional release of data only. On the day of the press release, in February 2016, let me just say that the open data web site went from zero to hundreds of hits per minute, and nothing crashed. So I'm proud of that.

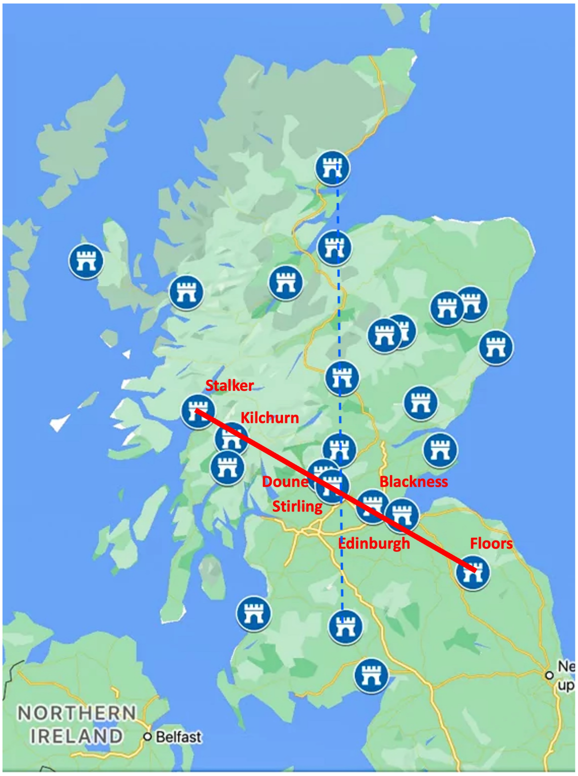
Needles in Haystacks

I'd like to talk about the process of finding jewels in data. The LIGO detectors are extraordinarily sensitive, but the signals they are measuring are extraordinarily tiny. If we didn't know what we were looking for, it would be impossible to detect it through the inevitable noise that comes with such sensitivity. Think of being in a large group of people talking, and as soon as somebody says your name, you hear it. It's as if we all have a special brain circuit matched to our own name – technically known as a *matched filter*. The signal that LIGO wants to detect is a *chirp*: a rising frequency, like a bird call, as the black holes rotate faster and faster around each other, giving off gravitational waves, losing energy, and thus rotating faster. The theorists have computed the precise waveform of the chirps of binary black holes inspiralling and merging, and they are good at matched filtering. So, you might think, we just push the data through the filter software and see what comes out. Not really!

It's a common paradigm in science, to try to find something interesting in a big pile of data. But the real work comes after the discovery – when you need to evaluate the *significance* of what you have found, so that your sceptical reviewers will believe it. While on the subject, LIGO and gravitational wave detection was plagued for years by the published results of Joe Weber from the 1960s, claiming discovery, which were discredited in the following years. Therefore the LIGO people wanted to be incredibly careful. It turns out that vast amounts of computing are used in this process, much more than what is needed to run the matched filter. The following is this scientific, justifiable way of deciding whether to believe something or not.

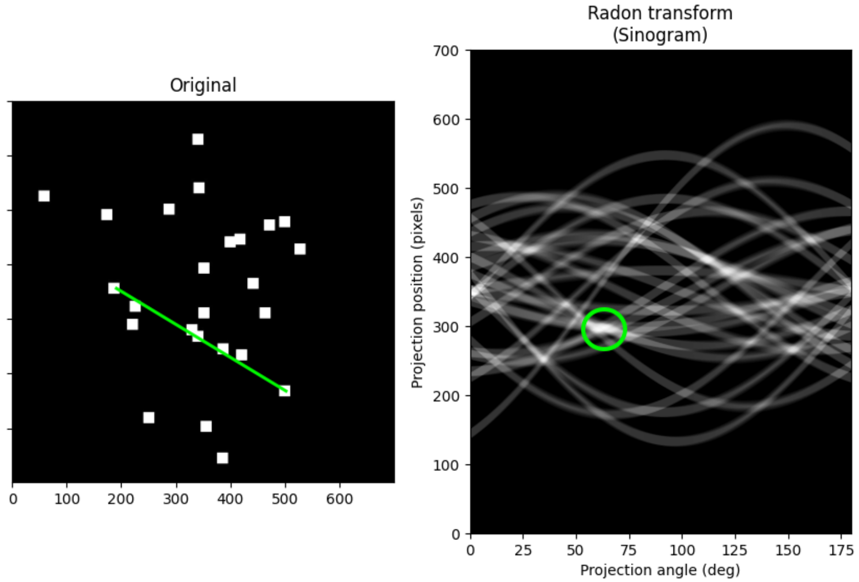
Ley lines and tomography

As an example, I'd like to talk about "Ley Lines": straight alignments of historical sites, prehistoric monuments, and other significant landmarks. The idea was developed 100 years ago by Alfred Watkins, who theorised the straight lines were ancient trade routes. Ley Lines were revived in the countercultural 1960s with talk of earth energy, dowsing, numerology, and the age of Aquarius. Let us start with the "best-of" list shown in Figure 9-2. Suppose we perceive the red line showing a "discovery" of an alignment. It is not right to simply announce this as something astonishing, rather we must prove that it is astonishing. Let us then ask if a random configuration of blobs also has these straight lines: if so, it means that our "discovery" is actually just by chance.



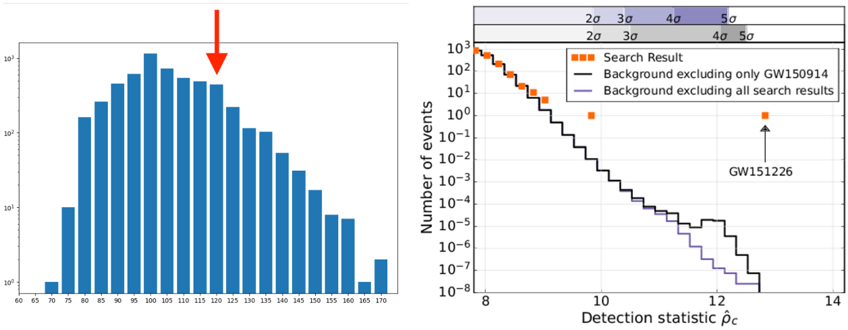
9-2 Given 25 of the "best castles of Scotland", we notice a near-straight alignment of 7 of them: Stalker, Kilchurn, Doune, Stirling, Blackness, Edinburgh, Floors. All of them worth visiting if you are touring Scotland! (From chasingthelongroad.com).

Let's start with the 25 "best" castles in Scotland, according to the internet (Figure 9-2). We can see that several of them lie in a straight line, shown in red. Do we just announce the "discovery" of the supposed alignment and hope for the best, or do we take a more scientific approach? We will need to run many trials to assess if our discovery is real. This can be done with the Radon Transform (Figure 9-3). Each white blob in the left side of the picture makes a sine wave on the right, and the place where most sine waves add together is the best line. Another way to think of the Radon Transform is that (1) each point corresponds to a line and (2) the value of the transform is the number of white pixels that line passes over.



9-3 Computing Ley Lines with the Radon transform (see section on tomography above). At left, the castles of Scotland, each shown with a 20 pixel white square. At right, the Radon Transform maps straight lines to points by projecting the castle image in every direction; the maximum value of this is the best straight line (green circle), also shown on the left figure.

Now we run thousands of trials, each is a random set of 20-pixel white squares on a black background, then finding the best straight line



9-4 (Left) Histogram of the detection statistic for the Ley Line search, and (Right), histogram of the detection statistic for the 2nd gravitational wave event, the so-called "boxing-day" event[6].

The results are to the left in Figure 9-4, along with a real significance plot from LIGO, to the right. Each shows the histogram of the detection statistic from a large number of random trials, with the purported discovery shown with an arrow.

The red arrow in the left panel shows the detection statistic for the Ley Line we saw, connecting seven Scottish castles, it has a score of 120, meaning the line passes across 120 white pixels of the map. However, the trials show that this is not unusual at all! There are many trails with much higher values when the random arrangements are considered. Therefore we should believe that the Ley Line we discovered is actually a chance alignment.

However the right panel, the LIGO result, shows what is almost certainly a gravitational wave found in the signals. The point marked GW151226 has a statistic of 13, much higher than the ~10 found in millions of random trials. Notice that the number of events with a given detection statistic goes well below 1, which seems odd; but this is done with a technique called "time shifting" that I won't describe here.

[6] From https://arxiv.org/abs/1606.04855

The upshot is that it's a high-significance event, and would occur by chance only once in 1000 years of running. Therefore, we choose to accept this as a genuine detection.

But we are not done yet, there is the human side of all this. The people running the computers and everyone they know is *desperate* to find that gravitational wave, knowing that their future in science depends on it. What we should not be doing in the above process is changing the way the detection is validated after the computations are done. In the Ley Line example, we might see that there is another possible straight line, shown dashed blue in the map of castles, and that line is north-south. Having found that in the data, we invent some story about neolithic civilisation worshipping the north and south, then declare the discovery. Or we adjust some parameter of the search and the computation of significance so that "our baby" suddenly looks significant. This is not how to do science! The right thing to do is the "box opening" – see below.

The Miracle Detection

First there was the prototype, the 40-meter detector, then the long years of building and commissioning what was optimistically called a gravitational-wave "observatory" with the 4km arms. The initial runs didn't find anything, and there were mutinous rumblings from the astronomical community about wasting precious funds on "looking for fairies at the bottom of the garden". The final push was called Advanced LIGO, a $600 million upgrade to bring all the instruments to maximum sensitivity.

My job at this time was leading the Open Data initiative. After being burned by the false detections of the past, LIGO was very close with their data, not allowing anyone outside the collaboration to analyse the data and possibly find and announce a false discovery. And even within the collaboration there were elaborate safeguards against changing the nature of the search to convert a nearly-significant event into a significant one: after a search had run over the data, they had a "box opening" meeting, where all the results of the computation are kept under wraps until everyone is in the same room, then the big reveal to see if anything is found. There were also blind injections into the system – before the first discovery – where very few people knew that an artificial signal was being added to the noise, just to make sure the systems were working properly.

Anyway, the National Science Foundation took the view that any future discovery must be properly exposed to public scrutiny to be accepted, and furthermore, the US taxpayers had paid for the project and had a right to see the results. So we designed and built the software and website for open data: the Gravitational Wave Open Science Center (gwosc.org).

Discovery Day!

It was 15 September 2016. I rode my bike down to Caltech, walked up the stairs to the office, and immediately felt an emotional crackle in the air. Soon I found that there had been a detection the previous evening, with great excitement as if Jesus had revealed himself *in corpore* to his disciples. As scientists, we are trained to be sceptical, not to believe what we want to believe, but to work hard at gathering more information to prove or disprove. The most recent blind injection had actually been a bit of a dud, because even the experts were not able to get the injected signal precisely right, given all the procedures of scrutiny. We could all see that GW150914 wasn't some marginal candidate event, but strong and clear. So everyone assumed it was an injection, but later than morning the director of LIGO looked us straight in the eyes and said no it wasn't an injection – that he would have known and he did not. There was a scramble to validate the data, to run the search codes with greater capability, and of course to keep it secret until everything was ready. It was exhilarating to be in possession of a secret known only to a few, the idea that 40 years of work and a billion dollars had found the astonishing, the amazing, the annunciation from the angels!

The signal of coalescing black holes lasted a quarter of a second; and is in the audible frequency range and resembled the chirp of a bird. The frequency increases because as the black holes spiral inwards, each orbit is faster than the one before, leading to the increasing frequency. The event was caused by coalescence of two black holes, masses 30 and 35 solar masses, and was about a billion light years from Earth. During the last 20 milliseconds, the power radiated in gravitational waves was 3.6×10^{49} watts – many times greater than the power output of all stars in the observable Universe.

After the discovery, there was a lot of questioning if it could be a false alarm, and the most likely theory was the “disgruntled postdoc”, somebody who was able to inject a very precise signal at both detectors, with very precise timing. But thinking it through, that just doesn’t work. It would have to be either one person with an impossibly broad knowledge, from electronics to data mining, interfering somehow simultaneously at two locations 2000 miles apart; or a much bigger group conspiring together, but without any hint of the secret coming out.

The Open Data team moved into high gear in the following months, building and testing. The strong cybersecurity team headed by Stuart Anderson enabled us to maximise the ability of our testers to see the website, but not to let anyone see anything if they were outside the collaboration. By Christmas, rumours were flying, but the LIGO-Virgo collaboration of 1000 were able to keep it quiet. I rather messed up at one point, where a page with the text “GW150914” was open to public view for 3 hours or so before I closed it down. No data breach, thank goodness, but in that time, the Google robot had indexed it and cached it, and I discovered how difficult it is to make Google forget something it has seen!

One innovation of the LIGO Open Data was the provision of not just data but also the software to repeat the analysis. We built Jupyter Notebooks, a way to run code in a browser window, code together with documentation and plots and images. Just like a web page, except that you could also run, edit, and run again to try different parameters, and extend and combine with other code. In other words, not just data, but a workshop and a toolbox to understand and test the LIGO analysis.

The press conference in February 2016 was a time of high excitement, with attention from the world’s media – the highlight, I think I can say, of my scientific career. What I was most proud of was that the GWOSC website didn’t crash, even though it was used for the first time that day, and was deluged with hundreds of requests per minute for gigabytes of data. In May 2016, the full collaboration, and in particular Ronald Drever, Kip Thorne, and Rainer Weiss, received the Special Breakthrough Prize in Fundamental Physics for the observation of gravitational waves. As part of the full collaboration, I was happy to receive 1/1000th of two million dollars, and there was a big smile on my face for a long time.

Gravitational Wave Counterparts

The first detection of gravitational waves was incredibly exciting: a whole new way of getting information about the Universe in addition the electromagnetic (light/radio/Xrays etc) and the particles like cosmic rays and neutrinos. There were more detections, also binary black holes, but these are not expected to produce any light. More interesting, astronomically, would be a coalescence of very massive and dense objects that are *not* black holes, because that would create light, and telescopes might be able to see it, and get much more data by taking a spectrum of the light. Before the first detection, I was part of the team working on this so-called "EMfollow" effort. We mapped the probability density of possible events, both on the sky and in terms of distance from Earth, and found all the galaxies from which such a signal might come. The possible galaxies were ordered in mass, since the more massive the galaxy, the more stars it contains, and therefore the more likely that the counterpart was in it. We had a system to coordinate all the observations to maximise collaboration and the eventual scientific impact of the event.

The event that made all this worth it happened in August 2017, two months after I left the LIGO project and moved from California to Edinburgh. It was a coalescence of two neutron stars that was seen by dozens of observatories in wavelengths from gamma to radio. The location of the optical signal was well-predicted and was in the third galaxy of the list I had made ordered by mass. I wish I could have participated in the excitement at Caltech. When I look at my wedding ring, I think of how it is probably formed in the ashes of a neutron star coalescence a billion years ago.

10. Science with Children

I waited a long time to have my children, and I was very happy that they came out as really nice people. What an opportunity to have a bit of science fun!

Little Ones

Let us begin by teaching them that unification and duality are driving forces of science.

One of the simplest science experiments is a matter of shaking up cream. You put double cream in a jar, shake for a long time, until it separates into butter and whey. Now you can spread the butter on bread. It is a story of transformation, that you thought there were two things – butter and cream – but actually one is made from the other. Then you and your child can look up the words “butter churn” and see how they did it in the past. Water and ice is another example of transformation, where the hard, sharp ice and splishy splashy water are two views of the same thing. If you live in a place without too much cloud, you can point out the morning and the evening star to your child, and tell the story of the genius from ancient times who worked out the two are actually one thing – Venus – seen in morning and seen in evening.

Rainbows are always fascinating. The bow is always 42 degrees from the shadow of your own head – about two spans of your hand. Teach your child where to look for a possible rainbow. Another route to the rainbow is CDs and DVDs, just reflect sunlight from them and there are fabulous rainbows. Now you can ask the (deep) question about why tiny droplets of water in the sky, and closely spaced grooves on the silver disk both produce the same kind of pattern. From where does this duality arise? The answer is that sunlight consists of all the colours of the rainbow, combined, and the light can be split into its colours in two ways. A prism or raindrop has glass or water, where light is split by differential *refraction* that depends on wavelength; a silver disk’s fine structure splits the light by interference, or *diffraction*, of colours according to wavelength.

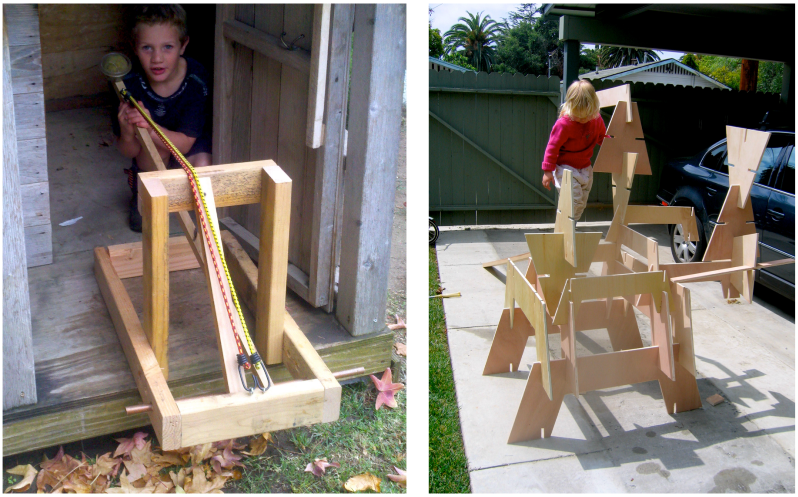
Another science project is to visit the kind of market where you can buy chicken feet. Chinese and Mexican supermarkets have worked out well for me. You do not need to eat the chicken feet! Rather get pliers or a vice-grip so you can hold the three white tendons where the leg was chopped off. When you pull on the tendon, the foot clenches, it shows exactly how tension from muscles is converted to movement of the body. Look at the tendons on the inside of your own wrist, that make your hand move.

A hard disk drive is a great place to find powerful magnets. Inside you see the platter that spins, and an arm, like the stylus of a vinyl record player, and the magnet is the actuator behind the arm, in the shape of an arc of a circle, see Figure 10-1. If you don't have an old disk drive, try to get a neodymium magnet, but don't spend too much time on the rather weak "toy" magnets with red paint on them. Your children ***must not*** eat a powerful magnet, that is very bad! So they need to be older, or alternatively, you can also attach the magnet to a piece of wood with duct tape or epoxy, so they can't possibly swallow it. You also need to keep a powerful magnet away from your phone or other delicate electronics. The children can go around the house testing what is magnetic and what isn't. "Is your mother magnetic?".

Figure 10-1 How to extract the powerful magnet from a hard disk drive (Wikimedia Alessio Sbarbaro)

The next couple of projects involve a bit of woodwork, but not difficult. If you have a sawhorse and a horse-mad child, you can add a horsey head from plywood, and a horsey tail at the back, then cover in stuffing and cloth. It's a project you can do with your child, and they can be shown how to hammer the nails. Another one – my other child – was not interested in horses, but rather swords and shields. We had visited a lot of castles in the UK that year. You start with a long stick for the sword and short one for the crosspiece. The shield can be made by painting a large round lid, and adding a string handle. A longbow is not too hard to make either, the main thing you need is a springy green stick as long as the child is high. Continuing the theme of medieval weaponry, the next thing was the trebuchet, an elastic-powered catapult that could throw a tennis ball 100 yards.

Figure 10-2 shows two more wood projects. A trebuchet was used for attacking a besieged castle in medieval times, and I made a small version powered by bungie cords. On the right is an erector set made from plywood sheet that slot together; I was lucky enough to have a wood shop with a table saw and a dado blade set up to make the wide cuts for the slots.



10-2 (Left) Trebuchet (Right) Homemade erector set from plywood.

High School

Here is something you should *not* do with your older child. When I was 8 years old, my father did a brilliant science demonstration with some old lead pipes. First we dug up some of the abundant London clay, and made some moulds by pushing our fingers in it, then we built a brick firepit with a grating so an old saucepan could sit on it (do not take a working saucepan from the kitchen). Once the fire was going, the lead went in the pan, and nothing happened at first. But suddenly there were brilliant drops of liquid metal in the pan, more and more as the lead melted. The deeply dull grey pipes became magical glittering quicksilver! We poured it in the moulds, and forever after there was a "lead thumb" in the garden shed to recall that afternoon.

I bought some magic substances from Amazon. You can buy a little glass tube filled with gallium metal. It has a melting point of 30 Celsius, meaning it is solid at room temperature, but if you hold it in your hand it becomes liquid. Quite a bizarre feeling! It is the same chemical family as aluminium, so not toxic when touched, but obviously it would be very nasty in the eyes or mouth, and obviously wash your hands after touching. Another strange substance is ferrofluid, also available on Amazon. Its really really dirty, so it must be kept in a container, without any chance of the children opening the container. But it makes the most lovely patterns in a magnetic field. See neodymium magnets above.

There were experiments on electrolysis: you just drop a 9-volt battery in salt water and see the hydrogen and oxygen bubbles at each end. You can see immediately that there is twice as much H as O because it is H_2O. You can capture the gas and light it with a satisfying POP.

Another thing you can do uses a 9-volt battery and steel wool. Make sure you are in an outdoor, fireproof area first! When you hold the steel wool against the terminals of the battery, its burns and sparks in the most beautiful way. If you have a sensitive kitchen scale, you might want to weigh the steel wool before and after the combustion, because it gets heavier by addition of oxygen from the atmosphere. You can also show how a substance finely divided and in the air burns very differently from normal. Throw sawdust or flour over a fire and it makes a flash of flame in the air. If you don't want to do this with your own fire, search

Youtube for “flour fire”. You can then talk about dust explosions through history, for example the Great Mill Disaster in the USA in 1878.

A slinky spring is great for demonstrating properties of waves. Once the spring is stretched out a bit, you can do compressive waves (like sound), and transverse waves (side to side), and even measure the difference between the speed of each type, which then relates to the S and P waves of earthquakes, how one kind arrives before the other.

For people particularly interested in Einstein’s theory of gravity, a sheet of spandex fabric and some heavy steel balls works well. Stretch the spandex over a frame, so that when you put the ball on the surface, the geometry is distorted by the weight of the ball. Now a smaller ball can “orbit” the larger one because of the potential well of the bigger ball. You can also shake the spandex sheet in various ways to show gravitational waves.

11. Software and Hardware

Languages

Programming when I was young was so different from now. It's as if there was nothing in the world but integers and floats (i.e. numbers with a decimal point), together with arrays of those that can be rectangular blocks. It was fine for computing ballistics tables, that task first assigned to electronic computers. But soon programmers wanted to do more, to work with strings of characters and utilise flexible data structures, to build code that is understandable and reusable. I'd like to summarise how writing code has changed in the 50 years since I started.

The first and most obvious difference is the time it takes to get the code to the computer. When I was a schoolboy, the computer ran every Tuesday and Thursday, when it was a games afternoon ("Computing as a Sport"). So the slightest syntax error meant a delay of days, and we carefully scrutinised our cards before the run, but even so, the human eye sees what it wants to, so I and 1 get mixed up. At Imperial College, where we took the card decks, there was a much more modern machine (CDC 6400), and the undergrads could queue up to put their cards in the reader. It was the "Instant Turnaround" queue. Sometimes we schoolboys would sneak into that queue, feeling very grown-up.

Fortran programs often have a spaghetti-like flow control. While the loop is simple enough – do this 100 times for example – it's the GOTO statements that make it difficult to understand, modify, and verify correctness of code. A particularly difficult statement is this:

```
IF (X) 10, 20, 30
```

which changes control flow to statement 10 if X < 0, to statement 20 if X = 0, and to statement 30 if X > 0. As soon as you see it, you know it's going to be really difficult to figure out what the code does. As you write more and more code for a given project, it definitely needs to be understandable: not just so somebody else can modify it, but also so the author can understand it in six months, when they have forgotten why it was written that way. So making code understandable leads to it being durable.

Modern languages are more explicit about scope – who sees what. In the old days of Fortran, things could get really out of hand: there's an array called X that is 30x40, and another array Y that is size 1200, then the `EQUIVALENCE` statement to say they inhabit the same memory, and a `COMMON` statement so that any subroutine can access any element. So if X doesn't seem right, it could be because somebody else's code changed an element of Y in some distant part of the code. These days, though, all the variables and arrays get passed around, so there is an explicit trail of breadcrumbs about where it came from and who had access to it.

Modern Coding

But now in the 21st century, the coding environment is so much easier than the old card punch. For one, you see the code itself instead of seeing a deck of cards, and there are visual editors and development environments with colour coding, so editing is easy, and syntax errors stick out. Multiple windows allow one for editing and another to run. Jupyter notebooks make the code into a document explaining itself, that can be a tutorial for somebody else, and a starting point for them to modify and extend a program,

Let me mention how code is made by teams, and how code can be modified without accidentally breaking it. The team uses a system called 'git' which holds the code and all be variations that people have made. Git was built by Linus Torvolds in 2005 of Linux fame (see below). There might be two machines running the same code, but one has the extensively tested 'release' version, that is known to work; the other has some new feature that is being tested and debugged, that will only be merged into the main branch after the team is sure it works. When a team member wants to put their new code into the main branch, they create a 'pull request' which is an invitation for review, for the rest of the team to scrutinise and test the new feature. It also means the rest of the team understands what is happening, in case the original author falls under a bus. There may be 'continuous integration', so that each time new code is pushed to the main branch, a set of tests is automatically run to ensure that things that used to work still work with the new code.

Another modern emphasis is encapsulation and classes, where data and the associated functions are bundled together, separating the builder of a piece of code from a user of that code. It's analogous to ordering a coffee, where the user asks for a cappucino, which is translated by the barista into a set of steps involving grinding beans and getting milk from the fridge. Data is hidden from the user: whether there is already enough milk, whether the steamer is already hot, and so on. This makes life easier for user, who doesn't need to think about what they don't need to know.

Modern languages have a richer variety of structure, not just rectangular arrays of integers and floats, with character strings. Lists and dictionaries play a big role in modern languages, as a way to organise data. A list is just an ordered set of other data structures, and a dictionary is a set of key-value pairs, where the values can be lists or other dictionaries. For example a list of people could be written:

```
[{'name':'Jane', 'age':20, 'gender':'F'},
{'name':'Sam', 'age':33}]
```

The square brackets indicate a *list*: and ordered sequence; the curly brackets indicate a *dictionary*: a set of keyword-value pairs. There can be lists of dictionaries (as above), dictionaries that contain lists, and so on.

Computers for Everyone

By the 1990s, I was still using the Unix operating system that I had learned in years before, with the command line and pipes and redirects, the well-trusted systems built at Berkeley in the 1970's. Absolutely super for running big computations, for machines talking to machines, for all that serious system-level work. But the new windowing systems were attractive too: Windows 3.1 had the advantage of the Microsoft Office suite: Word and Excel and Powerpoint. All through the 90's it seemed that anyone who knew Office could get a job right away. But Windows 3.1 had a big problem – all the processes ran in the same memory space, so that when one thing crashed, the whole machine would likely crash with the dreaded "blue screen of death".

In 1991 a young Finn named Linus Torvolds rebuilt Unix as his own, and called it Linux – a combination of Linus and Unix. It took off in popularity, because it was open source, meaning it's highly customisable, and fixable, in the sense that anyone can build new features and fix bugs, then request for their code to be pushed to the main branch. It was –

and still is – free to download and very stable, the harbinger of a whole new software business: instead of selling proprietary software, rather the product is free to use, but you can buy consulting and services to optimise it. Now Linux is the basis of the Android operating system for phones, running in 70% of the world's mobile devices. So good job, Linus.

Meanwhile in the late 1980s, Steve Jobs had argued with Apple computer, left them, and built the NeXT operating system and the hardware to run it. Innovative and modern, it was the vanguard of object-oriented programming and the windowing user interface. By the late 1990's, NeXT had not sold enough of the new machines, and was acquired by Apple for half a billion dollars, and brought Jobs back with it. Of course it was his magic touch, along with the British designer Jony Ive, that built the sleek iPod for music, then the idea of using a touch screen for a mobile phone. I remember the fashionable phones of the early years, the Nokia and Blackberry with their tiny keyboards with tiny keys – what a relief to use a touch screen instead. But the apps on those phones were the same for everyone; another big innovation from Jobs was the app marketplace, a place for third-party developers to make and sell code to run on the Apple iPhone.

As for hardware, that too has moved on quite a bit from the IBM 7094 that I ran my first programs on, even at that time ten years out of date. The program I punched in 1974 to compute pi to 5,000 decimal digits took special permission to run the computer for two hours. And I just ran the same calculation on the 2022 Macbook Pro and it took a fraction of a second. The first the size of a car, the second the size of a book.

Programming in the old days was, of course done with paper, with cards, with paper tape, and with the output printed on paper form a "lineprinter". It was about 1982 that I first started using what they called a "visual display terminal", meaning green text on a black background, a cathode ray tube in place of paper. A few years later it became black text on a white background, then colour and sound were added. I remember a discussion of the circumstances under which a colour display is necessary for the scientist programmer; and what was the purpose of a computer emitting sounds. We were all stuck on the idea that computers were meant to be for *computing*.

But in 1980, people started getting “home computers”. In the UK the Sinclair ZX80 was launched, and in the US the IBM Personal Computer. Apple had released the Apple II in 1977. Enthusiasts loved them, but most people wondered what it was for. Children already liked the arcade games like Pong, which were ported to the new hardware. I remember a recipe program, where you put in the number of people to be served, and the quantities were multiplied out. Not surprisingly, most cooks didn’t bother to switch on the computer for this simple task.

In my opinion, the “killer app” for the personal computer was the spreadsheet. All that tedium with the adding machine and its roll of paper, copying numbers by hand, so easy to make mistakes. But with a spreadsheet, you can not only do the arithmetic without mistakes, but also organise the calculations to show the logical flow, and add explanations so it becomes a report. You can do a scenario – suppose we changed things in a certain way, how would the tax bill change? Of course, like all innovations that make work easier, the result is increased demand for complexity, differing rates depending on circumstances, demands for more accountability and information. And soon enough a business cannot compete unless they have a computer with a spreadsheet. Videogames became a huge industry unbeknownst to me; I was never much interested in either playing solitaire with a deck of cards, or playing solitaire with a virtual machine-gun on a screen.

On the data storage side, the punched cards and tape were replaced by floppy disks, at first the size of a vinyl record, and eventually just three inches on a side. And by the early 2000s, floppies were only seen as icons – the glyph that tells us “click here to save”. It was the time of the flash-memory chip, or thumb drive. There was also the dot-matrix printer, with paper that came in rolls and had holes punched in the sides. By 1985 Apple released the LaserWriter, with the Postscript page description language – which evolved into the pdf standard we all use today. That allowed desktop publishing, the next killer app, with people talking about fonts and their semantics as they never did before.

One problem with these new peripherals for the home computer was that they required the user to open up their IBM PC and use a screwdriver to insert an “expansion card”. There was one for the printer, one for gaming, one for the floppy disk drive, and so on. In the late 1990s, the USB (Universal Serial Bus) arrived, so that all you needed is to plug in the new device, and it just worked.

From Programming to Software

At first, it’s just write a program. Then it becomes a longer program to do something more complicated. Then you want to work with a colleague on a program, and you look back to what you did six months ago and cannot see what any of the code means or what it does. This is when you start understanding the complexities and start calling it software engineering. Yes, ChatGPT can write you a bit of code to read a file and reformat the results, but it won't tell you why you are reading that file, who wants the reformatting, or why the data is full of mysterious numbers like -99.

Lets start with the -99, what happened here? I heard a story in 2002 about a large astronomical dataset called the Sloan Digital Sky Survey; essentially a huge table of stars, each with the magnitude in different filters (i.e. colour measurements). A group of statisticians was brought in to look for unknown patterns in the data that might indicate new kinds of science. They worked on the data for a few weeks, then came back with an exciting result: sometimes the magnitude is between 12 and 22, but a lot of stars have a magnitude of -99. The astronomers just laughed – they all knew that the -99 value simply means “unknown”, that the magnitude in that filter had not been measured for one reason or another. The real problem here is a failure of the *abstraction*, that when modelling the data of the sky survey, magnitude could only be a floating point number, like 15.77 or 21.3, and somebody decided that “everyone knows” that -99 means “not measured”. What is actually needed here is a concept of “NaN”, or Not a Number, to be added, to express the idea that the magnitude is not known. This is part of the IEEE 754 specification for floating point numbers from 1967: other strange beasts include infinity, negative infinity, and even negative zero! A related concept is “None” meaning nothing there. Subtly different – you can add a number to NaN, the result being NaN, but if you try to add a number to None an exception is thrown.

Then you start a job as part of a team writing software, and there is unimaginable complexity, where each team member understands one part of the edifice quite well, has a vague idea about the other parts, and some of it is mysterious because that team-member went elsewhere! It can only be made to work with abstractions, interfaces, and layers. Take for example driving a car: the driver learns about the steering wheel, the accelerator, and the brake, simple abstractions that allow the car to be driven. But behind these is complicated technology; think about how the accelerator is implemented very differently for internal combustion and electric vehicles. The internet is built on a four-layer transport model: (1) Network Access is the physical layer of wires and fibres and wifi, (2) Internet layer makes controls flow and routing, splitting and reassembling data packets, (3) The Transport layer provides reliability and acknowledgements, then (4) The Application layer is what the end user sees, with email, whatsapp, netflix, etc. Each layer has its own experts, vocabulary, messages, syntax, and semantics.

An important word for software developers is “idempotent”, meaning that repetition does not change the answer. A device with a “Stop” button is idempotent, in the sense that the first push will make the machine stop, but subsequent pushes have no effect, because it is already stopped. If you ask a webserver for a simple page, then ask again, it gives the same result. However, if you order an item from a shopping portal, and do so again, it is different: you get two of the items sent to you! If you fill in a web form and submit it, then try to refresh the page, there will be a warning, something like “do you want to resubmit the data”, because your browser knows the submission is not idempotent. Suppose you want to process a large number of data files, pushing the results into a database. The simplest way to do this assumes that all goes well from start to end, with each file creating one database entry. But what if something goes wrong in the middle of the job, and only half the files are processed? You have to delete what you did from the database and start again. A better strategy is to keep a list of what has been processed, and only process what isn't on the list. Now the entire operation is idempotent, and a failure simply means a restart without redoing work.

Another place of darkness in software comes from identifiers and names. I recall a long meeting about how to represent people’s names in a database. Names change, there are several names, people use multiple names at the same time, the first and last names can be the same, they can be all lower case or all capitals, of course they use non-English character sets, they are not unique, they can go by a one-word name, they can contain hyphens and apostrophes. A baby might not even have a name until it reaches a certain age. Even the phone number has complexity and hidden rules: are spaces allowed in the input, and if allowed, must they be kept in the system? The phone number can have a local version, without the area code, and a national version, to which a country code must be added for international use. The UK phone number starts with a zero for domestic use, which is replaced with country code +44 for international use.

An identifier, by contrast, is for use by the machine to fetch data, and may have no meaning to a human. While a relational database can handle very general queries – for example select all records from last month from a specific user – it is limited in scope by the need to support that generality, and difficult to scale up to very large databases. By contrast the non-relational databases, such as Cassandra or MongoDB, do not allow general queries, but are more like a key-value retrieval. Given an identifier, the database produces the corresponding record; but none of the flexibility of a relational database, no JOINs or foreign keys or data-based selection. In exchange for this restriction on retrieval, the database can scale up to much larger sizes than a relational database. Before computers, a library would use card catalogues, and there would often be two of them side by side: the author index keyed by author alphabetically, and the subject index keyed by something like the Dewey Decimal or Library of Congress system of subject identifiers. In astronomy, data on sources (stars and galaxies) would be keyed by catalogue identifier, but of course there is also a requirement to search on sky position – “what sources are near this source?”. In the world of transient astronomy (flames and flashes), a third requirement is to search on time.

Predictive text and Machine Learning

Recently, the ChatGPT arrived, and the world loved it and hated it. It bears out Arthur C. Clarke’s words from 1962: “Any sufficiently advanced technology is indistinguishable from magic”. So let me show a little tiny version of ChatGPT, that shows the basic principles. Let us start with seven works of Charles Dickens, the Victorian novelist, each over 300,000 words, and try to imitate his style.

First we strip out all the punctuation to leave just words: in these seven weighty novels there are 1,931,000 words chosen from a vocabulary of 34,265 words. The simplest way to imitate the style of Dickens would be to blurt out his vocabulary at random, there result being like this:

```
his used witness a your every she it the light a
bloody been fellow
```

This obviously makes no sense and really doesn't look like Dickens. In the next stage, we list all the *pairs* of words. We know that Dickens would never write "been fellow" as appears in the above fragment. As an example, we find that the word "stationery" can be followed by only three unique words:

```
stationery → and*2,business,consumed,
```

where the "*2" means the pair "stationery and" appeared twice. So now lets build a sentence where we use the last word of the sentence to predict the next one, according to the pairs. Now it looks a bit more readable:

```
he came neither ignorant of stationery consumed
in her blue
```

the next step is to make a list of all the triples. Given the pair of words "stationery consumed", we find the only possible next word is "in", then we look at what can follow "consumed in", and we find several possibilities:

```
consumed in → care, the, incessant, getting,
fruitless, the, it
```

Finally, we make a list of all the quads: given three words in a sequence in a Dickens novel, what can the next word be. But now we run into a new problem, that now the chatbot is simply quoting long sections from one of the novels, rather than making up text in the style of Dickens. If, for example, we start with "The assurance he", then a long quote appears from Chapter XXIII of *Little Dorrit*: "The assurance he now had, that Blandois, whatever his right name, was one of". The problem is that in all the novels of Dickens, there is only one word that follows the first three, and again, and again. Only when we get to “was one of”, do we get some variation, where the possibilities are:

```
was one of → Mr, her, my, our, the*6, them,
those*5
```

And so we veer off the direct quote, and actually end up with nonsense, and yet nonsense with a distinctly Dickensian look:

```
The assurance he now had that Blandois, whatever
his right name, was one of the party came in to
breakfast with no better company anywhere, but
he put so plainly at sea on this part of the
parallel stand. Arthur had no choice but to do
his utmost at all times stuffed and close as if
it . . .
```

The Large Language Models of ChatGPT are obviously a lot more sophisticated than just word sequences. But the point is that it just copies its input and attempts to regurgitate that input. Artificial intelligence is not really intelligence at all.

12. Edinburgh

Escape from Caltech

In the early days of the Virtual Observatory, there was a strong collaboration between the US and the UK, and I did my best to take trips from my home in Pasadena to Edinburgh, where the UK meetings were concentrated. It also meant that I could go to Aberdeenshire to visit with my nephews Josh and Dan, who were charming toddlers. My brother Fred, like me, waited a long time to have children, but certainly my eternal thirst for community and family was slaked by these trips. My new wife Jessica came along in 2002 and she loved Edinburgh, and we explored the splendid north-west coast of Scotland. She declared that if we moved from Pasadena, perhaps we could move to Edinburgh.

By 2016, I had been at Caltech for 30 years, first with simulations of fluid flow on parallel computers, then building web-based science archives, the standards process of the Virtual Observatory, then building the open data system for LIGO. The big discovery had been made – gravitational waves – and my 60th birthday approaching. Definitely time for a change! I was looking at college costs for my two children, then 11 and 12. Something else had been in my mind since 2002, talking on the phone with Fred, when he said “We could do with a bit of global warming here”. I already knew that climate change would become more and more significant through the time of my children’s lives, and decided that the soft drizzle and cool summers of Edinburgh would be a much better place to live than the increasing heat, drought, fire, and flood of the American southwest. Besides, the children were having “active shooter drills” at school, and Trump was rearing his ugly head.

So I was over in Scotland in summer 2016 (see below), at my brother’s house, talking about the possibility of this move. My sister-in-law, Dominique, said “Roy, if God wants you to move to Scotland, He will give you a sign!” I had already looked up my friend of many years, Andy Lawrence, told him I was doing school visits in advance of a possible move, and we had dinner together. What a surprise it was to be invited by him and Bob Mann to apply for a job at the Royal Observatory … I think that must have been the sign that Dom talked about!

That visit to Scotland was also to be at a LIGO meeting, it was in Glasgow, and over lunch I had been chatting with “SG”, who had posted on Facebook a picture of her summer intern giving a talk, with the caption “You go girl!” I remarked that as a man it would not be appropriate for me to say the same thing about my own (female) intern, who had also given a talk. SG looked horrified, as if I had said something terrible. Her friend caught wind of this, they whispered, and then the friend was also horrified. I kept asking “what did I say?” and “why are you offended?”, but they would not clarify. Later that afternoon, it turned out that SG had complained to my manager, who took me aside and warned me about my bad behaviour. He would not say why SG was offended, and I still don't know why she was upset, maybe SG did not hear what I actually said, but imagined something terrible? Anyway, the fact that SG was *very offended* was the important thing.

A year later, it was just days before I quit Caltech, to move myself and my family to Edinburgh. There was a nice lunch of people saying goodbye to me, and I went back to the office in a glow of good feeling. However, early that morning had been another Facebook exchange, which unknown to me was visible to SG. The post was shared by a mutual friend, about Professor Ott who had invented a fake female name as a co-author on his paper. I commented that there have been fake author names before, quoting some famous examples, and perhaps Ott felt that a female name would increase his chance of a good referee report. I knew that Ott is the deadly enemy of SG, but I did not know that she would see my comment, nor that she was already an author on this paper, nor that she would become enraged. Anyway, there were screams and shouts coming from my manager’s office that morning, before the lunch. After lunch another “little talk”, where he accused me of harassment, told me I would get fired from my Edinburgh job if I continued this way, told me about others “like me” who had been knocked down by their “bad behaviour”. He also indicated his personal dislike of Ott, as if that’s relevant. My Facebook comment was based on a simple reading of the post, and I was full of anger that it could be weaponised in this way. Why did she again decide to attack me? Did SG have undiagnosed mental illness? Perhaps she thought I was impugning her own scientific ability? Who knows. She was *very offended*, that was the important thing.

It has been lovely in Edinburgh, in spite of the long dark winters. My family have taken to the new place like fish to water, without any complaining. Three years attending a Scottish secondary school means free University tuition for them, so that’s a lot better than the $50,000 - $80,000 per year charged by a good American University.

Nobel People

For the record, I thought through all the 12 Nobel prizewinners with whom I have had some sort of conversation, mostly during the 30 years I worked at Caltech.

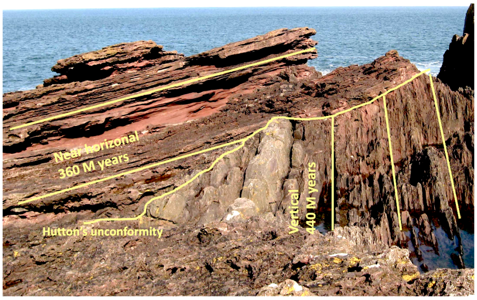
- Eugene Wigner, Physics 1963, shared a taxi in Israel in 1981.
- Richard Feynman, Physics 1965, taught me quantum mechanics in 1980, told me at a lunch table that play is essential in science.
- Hans Bethe, Physics 1967, had dinner with him in 1983 at a professor's house, stories of Los Alamos.
- Willie Fowler, Physics 1983, wrote a paper with him about the isotope Rhenium-187.
- Rudi Marcus, Chemistry 1992, meetings about parallel computing for chemistry, during the 1990s.
- H. David Politzer, Physics 2004, mostly random chitchat over lunches, forgot how we met.
- Barry Barish, Physics 2017, stayed at his house when my friend Jeremy Bentham was housesitting, plus other business.
- Kip Thorne, Physics 2017, talked about wormholes at a party at his house, plus other business.
- Rai Weiss, Physics 2017, spent an hour inspiring me about LIGO at a Starbucks in 1998.
- Frances Arnold, Chemistry 2018, talked at a party, always admired her from afar.
- Andrea Ghez, Physics 2020, worked with her on a software project in 1988.
- John Hopfield, Physics 2024, talked to him about optimisation with multiple parallel actors in 1986.

Deep Time and Deep Space

Sometimes something is discovered that is so astonishing that your head rings like a concussion. Your whole world view is broken and needs to be rebuilt. We all remember the wondrous childhood moment when we realized that every star in the sky is as big and as hot as our own Sun. People instinctively see that which is close and familiar as the template for everything, they saw the clouds and the stars moving in the sky over the fixed surface of the Earth, and made the obvious assumption that clouds and stars are about a mile away and moving. There were fanciful dreams of crystal spheres and platonic solids and epicycles and the stars were pinpricks allowing through the light of Heaven. But then science and logic and experiment came to the rescue!

People found that the Earth moves around the Sun rather than the other way (1543); that the force controlling the fall of the apple is the same force that controls the movement of the moon and planets (1666); the Earth was older than it is possible to imagine (1785); the Universe is larger than it is possible to imagine (1924); the Universe is expanding from an infinitesimal point (1930s); there are places where space and time are infinitely curved (1971).

James Hutton was born in Edinburgh in 1726, made his fortune from chemistry, and took up his scientific interests. His breakthrough principle was that the natural processes seen today are the same as those that formed all geological phenomena: that a volcano of the distant past would look like a volcano today; that rocks smashed to sand on a beach happens now as it happened in the past. From this principle, he build a workable systematic theory of geology. By the late 18th century, there was a lot of observational knowledge about different kinds of rocks and their strata, and plenty of fossils. However, the Bible was being taken as literal truth, meaning the Earth was forced to be 6,000 years old, the sedimentary rocks remnant so the great flood, the fossils placed by the devil, and so on. People knew that rocks eroded and calcium shells of animals dropped to the seafloor, but Hutton realised that heat and pressure are the agents converting these sediments to hard rock; he recognised that the heat of the Earth's core was a crucial part of geology; he broke from his colleagues with scientific, rather than Biblical explanation.

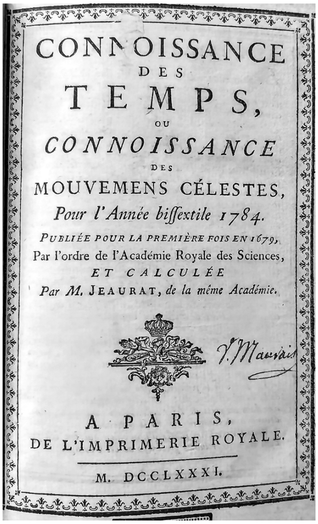


12-1 (Left) Siccar Point with Hutton's Unconformity separating rocks that differ in age by 80 million years (Wikimedia Dave Souza CC BY-SA 4.0). (Right) My own painting of Siccar Point.

Soon after my family and I moved to Scotland, we took a pilgrimage down the coast to Siccar Point (Figure 12-1), the place most associated with Hutton's theory of geology, the place associated with the idea that the Earth is actually a million times older than human artifacts. You can see sedimentary sandstone that must have taken a long time digesting deep in the Earth, that has been pushed back up and is now vertical; then over that further sandstone is laid down and converted to rock; and that the whole long process is exposed and has been subject to erosion for further millions of years.

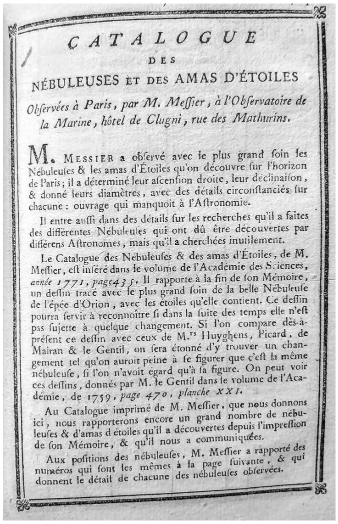
Imagine, I said to my children, that you have lived 80 years near a deep river cut valley, and you have not perceived any further cutting in that time; now you realise that to cut down 200 meters, the river must have flowed for a million years. This is "deep time", where you look back further than your great great great grandparent, further back than the earliest stone of the earliest King of Kings, a million times further. It takes your breath away.

Now let us think of deep space. At the same time Hutton was developing his theory of geology in Scotland, astronomers were making better and better telescopes, and Charles Messier in Paris was obsessed with finding comets: the king of France called him the ferret of comets. A comet looks like a fuzzball, a diffuse source of light quite different from a point-like star or disc-shaped planet. Messier realised there were fixed and transient diffuse objects – the latter being comets – so he published a catalogue of the fixed ones so that other comet-hunters can discount them (see Figure 12-2). There were 100 or so in the catalogue, which I found in the library of the Royal Observatory Edinburgh. Quite astonishing to be taken back in time to that era of powdered wigs and kings of France.



CONNOISSANCE
DES
TEMPS,
OU
CONNOISSANCE
DES
MOUVEMENS CÉLESTES,
Pour l'Année biſſextile 1784.
PUBLIÉE POUR LA PREMIÈRE FOIS EN 1679,
Par l'ordre de l'Académie Royale des Sciences,
ET CALCULÉE
Par M. JEAURAT, *de la même Académie.*

A PARIS,
DE L'IMPRIMERIE ROYALE.
M. DCCLXXXI.



CATALOGUE
DES
NÉBULEUSES ET DES AMAS D'ÉTOILES
Obſervées à Paris, par M. Meſſier, à l'Obſervatoire de la Marine, hôtel de Clugni, rue des Mathurins.

M. MESSIER a obſervé avec le plus grand ſoin les Nébuleuſes & les amas d'Étoiles qu'on découvre ſur l'horizon de Paris; il a déterminé leur aſcenſion droite, leur déclinaiſon, & donné leurs diamètres, avec des détails circonſtanciés ſur chacune: ouvrage qui manquoit à l'Aſtronomie.

Il entre auſſi dans des détails ſur les recherches qu'il a faites des différentes Nébuleuſes qui ont dû être découvertes par différens Aſtronomes, mais qu'il a cherchées inutilement.

Le Catalogue des Nébuleuſes & des amas d'Étoiles, de M. Meſſier, eſt inſéré dans le volume de l'Académie des Sciences, *année 1771, page 435*. Il rapporte à la fin de ſon Mémoire, un deſſin tracé avec le plus grand ſoin de la belle Nébuleuſe de l'épée d'Orion, avec les étoiles qu'elle contient. Ce deſſin pourra ſervir à reconnoître ſi dans la ſuite des temps elle n'eſt pas ſujette à quelque changement. Si l'on compare dès-à-préſent ce deſſin avec ceux de M.rs Huyghens, Picard, de Mairan & le Gentil, on ſera étonné d'y trouver un changement tel qu'on auroit peine à ſe figurer que c'eſt la même nébuleuſe, ſi l'on n'avoit égard qu'à ſa figure. On peut voir ces deſſins, donnés par M. le Gentil dans le volume de l'Académie, de 1759, *page 470, planche XXI.*

Au Catalogue imprimé de M. Meſſier, que nous donnons ici, nous rapporterons encore un grand nombre de nébuleuſes & d'amas d'étoiles qu'il a découvertes depuis l'impreſſion de ſon Mémoire, & qu'il nous a communiquées.

Aux poſitions des nébuleuſes, M. Meſſier a rapporté des numéros qui ſont les mêmes à la page ſuivante, & qui donnent le détail de chacune des nébuleuſes obſervées.

12-2 Publication of the Messier catalogue in Connoissance de Temps in 1784 (collection of Royal Observatory Edinburgh)

The astronomers of the time did not conceive of deep space, that these Messier objects were so much further than the comets they seek. At the speed of light, a comet is just a few hours away, but the furthest Messier object is sixty million years of light travel time! As telescopes improved, the spiral structure of many galaxies was discovered, but their nature remained a mystery. The biggest of these mysterious spirals was Messier 31, in the constellation Andromeda.

While the discovery of deep time was in Scotland, the discovery of deep space was in Pasadena, California. The Mount Wilson Observatory was built in the early years of the 20th century, by George Ellery Hale, and its 60-inch and 100-inch reflecting telescopes were the best in the world. In 1917, Harlow Shapley established the size of our own Milky Way galaxy, and that we are about 15,000 light years from the centre. But there was still the debate about spiral galaxies – perhaps they were outlying parts of our own galaxy?

Hale hired Edwin Hubble in 1919, who went on to discover the place of mankind within a Universe of unimaginable size. It was Henrietta Swan who perceived the usefulness of Cepheid variable stars for establishing distance: she measured thousands of lightcurves of Cepheids, all the same distance, and demonstrated the period-luminosity relationship: the longer the period, the greater the intrinsic brightness. The distance to Swan's stars, all in a satellite galaxy of the Milky Way, could be established by the parallax – changes in position through the year as the Earth orbited the Sun. But it was Hubble who turned the enormous 100-inch telescope on the Andromeda spiral galaxy, looking for variable stars. He found the characteristic light-curve of a Cepheid, and was able to calculate the intrinsic brightness from Swan's work. Distance is derived by comparing intrinsic with apparent brightness, yielding two million light years! Over a hundred times the size of the Milky Way galaxy.

As astronomy advanced, our horizons increased by many orders of magnitude; the centre of the Universe moved from the Earth to the Sun, to the uncounted billions of stars in our Milky Way galaxy. Then it turns out that our Galaxy is just one of uncounted billions of galaxies. Our minds quail at trying to know these numbers and sizes! Today the picture is that there is no actual centre of the Universe, that deep time and deep space are connected though the speed of light, going back to a Big Bang moment 14 billion years ago when space and time started to have meaning.

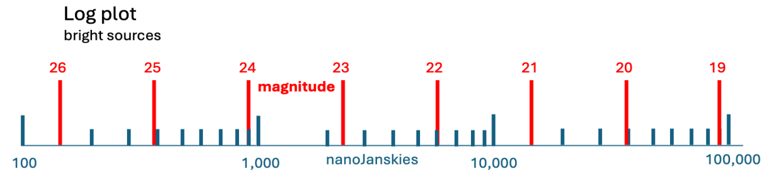
There are many kinds of galaxies in addition to the pretty spirals. The very largest – ellipticals – have been built by successive collisions and lost their delicate structure; they are often the centre of a galaxy cluster, with thousands of orbiting "ordinary" galaxies. These very massive galaxies distort the very geometry of space: the view of what is on the other side is warped and twisted by their gravity. There is a light show when two galaxies collide, and material is pushed together. Clouds of hydrogen and dust form new brilliant stars, that often do not last long and the supernovae crackle like exploding fireworks.

Magnitudes: The Astronomer’s Burden

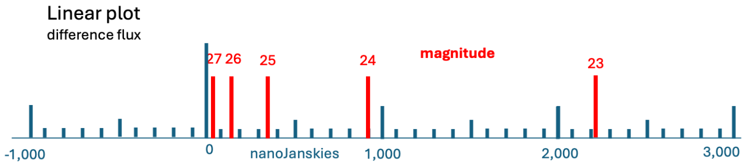
Astronomers are interested in brightness of stars. They have been for thousands of years. In Ancient Greece, two thousand years ago, the range of brightness visible to the eye was divided by Ptolemy into six magnitudes, and this has led to the system in use today. An alternative – flux – is more rational but not well-known. On the one hand we have tradition (think shillings and pints and feet), and one the other hand we have the rational system (eg: 1000 mm in a meter).

The magnitude scale is logarithmic, 5 magnitudes is 100 times brighter. The Richter earthquake scale is also logarithmic in the energy released. The reason that a quantity is primarily logarithmic is that it has great range of values, which in the examples above is testament to the quality of human senses. We can hear the quiet sound of a person breathing while asleep, but also hear the sound of a rock concert, the latter being a million times the energy. Similarly, the logarithmic nature of the magnitude scale is because of the great flexibility of the eye, which can operate over a range of a trillion in light energy! Magnitude 6, the faintest for the naked eye, is 10 Janskies in flux units, meaning trillionths of a watt per square meter – whereas our Sun on a sunny day is more like 10 kilowatts per square meter. So you see the utility of a log scale.

Magnitude is an old unit, still in use because of culture rather than logic. Astronomers have been here before with sky coordinates. There is an analogue of latitude and longitude for points in the sky, and up until the early 2000s the sexagesimal scale was used. The declination was degrees, minutes and seconds, and the right ascension was hours, minutes, and seconds. Points just south of the equator are particularly tricky to deal with for computers. A declination of -00:37:23 is likely to get wrongly converted, because the programmer splits it into tokens, then mistakenly converts each token to a number. This means the “-00” becomes simple zero, and the point below the equator becomes a point above the equator! Anyway, awful system. But progress: for the last 20 years, astronomers have weaned themselves off sexagesimal, started using degrees with a decimal point.



12-3 A plot of magnitude as the logarithm of flux. Works well for positive flux.



12-4 A linear plot of flux, with the magnitudes failing to show negative values.

The problems with the magnitude scale start with the fact that it runs opposite to common sense: as you decrease the magnitude number, you increase the actual magnitude! It's logarithmic scale, so that an increase of 1 magnitude means the brightness (flux) changes by a factor of $10^{2/5}$ = 2.512. Nobody understands astronomical magnitude except astronomers! See “Log Plot” in Figure 12-3: the flux is in blue with logarithmic tick marks, and the corresponding magnitudes in red. Two powers of ten are five magnitudes. People who come from physics, rather than observing, would rather use flux, as that’s what physicists call brightness. But tradition wins, like so often in life.

However, a new problem for the beleaguered magnitude scale looms on the horizon: transient surveys, a type of data that will be especially exciting when the Rubin Observatory starts up. With a transient survey, the software looks for differences in flux between the reference image, taken at the start of the survey, and the current image from tonight. So its no longer just brightness, which is always positive, rather *difference* brightness, which can be positive of negative.

In real life this happens with time: think of the words “February” and “fortnight”. The first is a marker of absolute time, the second measures difference between two fixed times. This time, the difficulty for supporters of the magnitude scale is not its inherent quirkiness, but rather inflexibility. But the flux scale can handle differences like a duck can handle water! Looking for a faint pre-explosion surrounded by noise uses a linear scale, not logarithmic, and the data can be negative as well as positive (see “Linear Plot”). Anyway, it is very awkward dealing with difference flux when expressed as magnitudes.

Billions of Flashes from the Sky

I have been part of a team making a “community broker” for the Rubin Observatory[7] LSST survey (`lasair.lsst.ac.uk`). The telescope finds changes in the night sky – variable stars, supernovae, black holes eating stars, and many other transient phenomena. The word “lasair” means flame or flash in both Scottish and Irish Gaelic – because the team is from Edinburgh in Scotland and Belfast in Ireland. Gareth Francis, Ken Smith, Stephen Smartt, and Dave Young are my trusty team-mates, who make up for my failures and work all hours of the day to make Lasair a reality. Once again, I am making an Active Digital Library, where computing and databases and the web browser come together. Once again, I am building a platform for science, not doing the actual science. Like those selling shovels in the gold rush – who did much better than those digging for gold.

A lot of science can be achieved by studying the variability of celestial sources. Particularly for gravitational lensing, searching for supernovae, determining the physical properties of gamma-ray burst sources, discovering gravitational wave counterparts, probing the structure of active galactic nuclei, studying variable star populations, and discovering exoplanets. Variability can also enable understanding of exotic phenomena such as neutron stars and black hole binaries, novae and stellar flares, gamma-ray bursts and X-ray flashes, and stellar disruptions by black holes. It could also discover new classes of transient, as yet unseen, such as mergers of supermassive black holes.

The LSST survey will be carried out at the Rubin Observatory in Chile, taking wide-field images of the sky in six filters, and looking for changes. The Rubin observatory has the largest ever camera used for astronomy, at 3,200 megapixels, that can detect sources as faint as

[7] https://rubinobservatory.org/

magnitude 27. It will take hundreds of images each night (about 20 terabytes) and compare each image with a previous reference sky. Each difference (at the 5-sigma level) will be converted to an alert packet, which will be sent immediately to seven “community brokers” for storage, added value, so the brokers can forward rare jewels to the scientists. There are three such brokers in the USA, one in each of Chile, France, Germany, and the UK, the last being the Lasair project described here.

The survey will produce millions of alert packets per night, each about 80 kilobytes, so the data rate will be 200 Mbps (megabytes per second), potentially with burst up to 5000 Mbps. Note that a domestic video stream is about 8 Mbps, going into special hardware to render it into video. This 200 Mbps is a lot of data if it must be processed by software and stored in a database. And remember this isn't a unique demonstration of an hour at these speeds, it will be all night, every night, reliably.

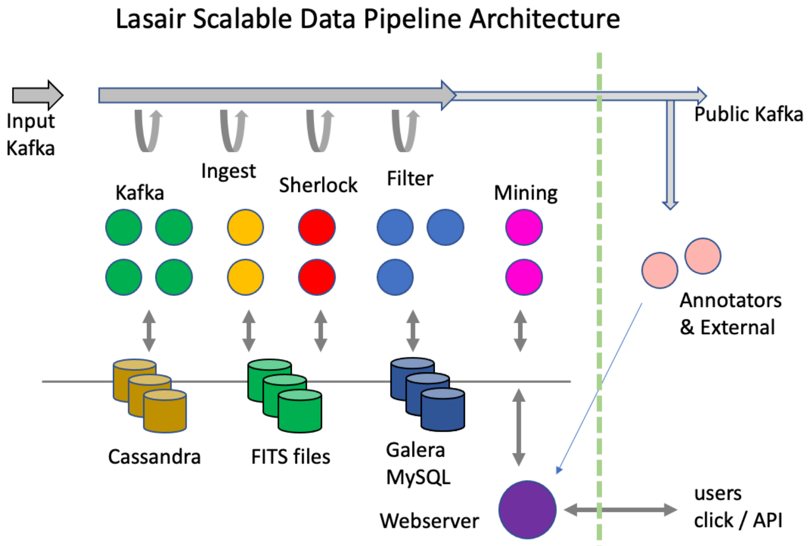
The trick of distributed computing is to minimise the need for consensus, minimise one node waiting for another. It’s like employing people: if you are able to give each a self-contained task that doesn’t require working with others, then each employee can work faster and more efficiently. The high-speed fluid flow computations from an earlier chapter are not like this, because each forward timestep requires synchronisation and data exchange. Thus the overall computation runs at the speed of the slowest node, hence the need to balance load.

Lasair runs on an Openstack cloud “Somerville” at the University of Edinburgh Advanced Computing Facility. Lasair ingests data with a pipeline of scalable clusters, as shown in Figure 12-5: Kafka, Ingest, Sherlock, Filter. Each cluster does a different job, some more compute/data intensive than others. It is difficult to know *a priori* how much resource should be allocated to each cluster, so our design gives flexibility: each cluster can be grown or reduced according to need. Also, there are persistent data stores (Cassandra, MariaDB via a Galera cluster); again, each is a resilient cluster architecture that can be grown or reduced according to need. The diagram shows the concept: data enters the Kafka system on the left and progresses to the right. The Kafka cluster (grey) consumes and caches data from Rubin and from the other pipeline clusters; the ingest cluster (yellow) splits and redirects,; the Sherlock cluster (red) crossmatches with known catalogues; the filter cluster (blue) runs user filters and watchlists, then stores the alerts to the Galera database cluster and sends user output to public Kafka. We also include the web and annotator nodes in this picture (bottom and right), as well as the mining nodes, although they are not part of the data ingestion pipeline. The web server supports users by delivering web pages and responding to API requests. The annotator nodes may be far from the Lasair computing centre and controlled by Lasair users, but they are in this picture because just like the others, they push data into the Lasair database.

Being a software engineer, like so much of life, means a history of regrets and mistakes, understanding how it should have been done, and doing it right the next time. And its easier to get it wrong several times, than think through how to get it right the first time. Furthermore, in scientific software (my life), even the requirements are not known initially, and only when the data starts coming in does it become clear what is wanted from the software system.

Sometimes it is best to make a single piece of flexible code that can do many things, with switches and parameters to modify the behaviour, a sort of swiss-army knife multitool that can do lots of things. Of course every line of code means maintenance. Some future person (or a future yourself!) will need to understand what it does and why it’s common code, but doing just one job well. Like a knife that is only a knife, and a pair of snippers that just snips.

Smaller chunks of code are easier to handle than a sophisticated, complex application that has its own documentation. The smaller codes are also easier to test, easier to modify, and easier to delete and rewrite. Much of the trouble with a bigger code is the “sunk cost fallacy”, where a serious investment has been made, and nobody wants to admit that it needs to be rebuilt, even if that is the optimal solution, because it makes the previous effort look worthless.



12-5 Architecture of the Lasair system

there. So we want to re-use existing code as much as possible. But sometimes it is better to make several software tools, each with a bit of

We mustn't be overly enchanted by new ideas. Python, MySQL, git, etc are all well-understood by the software team, and equally important their failure modes are understood. Really there should be just a few real innovations in any project, if there are too many it will be impossible to maintain.

Anyway, lets get to the processing of enormous amounts of data. The innovations of Lasair are:

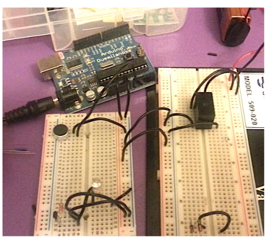
- Letting the user write SQL queries that wil be executed by Lasair, allowing the flexibility and expressiveness of SQL, but carefully checking to make sure it does no harm.
- Lasair is architected in several functional stages, data flowing downstream from one ot the next; but each stage can be implemented by multiple nodes, so that a slow stage can be accelerated simply by adding new nodes to that stage.

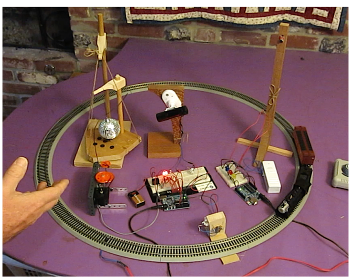
- When data is being ingested, the databases write that data, and nobody is reading from them at the same time. This allows asynchronous writes, so the database is doing the work, while the processing node does not need to wait for the database.

Lasair divides the day in the natural way. At dusk, lookup systems are prepared and cached for watchlists, annotations, etc, so they are ready for the processing nodes. During the night of course, is the very fast flow of data from the telescope, putting it into the databases and running user filters so that they get rapid notification of interesting phenomena. The morning is for catching up from the night's data, then during the day the databases can run follow-up activities to get more detail on that which was interesting the previous night.

Arduino and Pi

We nerds, like everyone else, are desperate to make our children into clones of our wonderful selves. So there’s all this science in their lives, and attempts to get them to learn coding for computers. Sometimes it takes – like a fish to water – and sometimes it doesn’t. I spent quite a time playing with tiny computers called Arduinos, making them do fun things. I have also worked with Raspberry Pi, which seems much the same thing, but is a rather different architecture inside. The difference is that the Arduino can only do one thing, while Raspberry Pi is a Unix machine, capable of multiprocessing. The Arduino is great for a simple logic, where the computer is watching a sensor, and as soon as





12-6 (Left) Arduino with breadboard (Middle) Model railway with sensors (Right) Lights around the door.

the signal comes in, it does something on its output, so it's just Input + compute → output.

In Figure 12-6 are three photos of Arduino projects. In each case, there is a hardware and a software side: the hardware (left) might be a button microphone and some LED lights; then you connect the Arduino to your computer and use the development environment to write code and push it to the Arduino. Once the code is in there, it stays, even without power. You can power the Arduino with a 9V battery or by a USB cable. You don't need to solder anything unless you want to graduate from the breadboard to a compact, hardy version of a project.

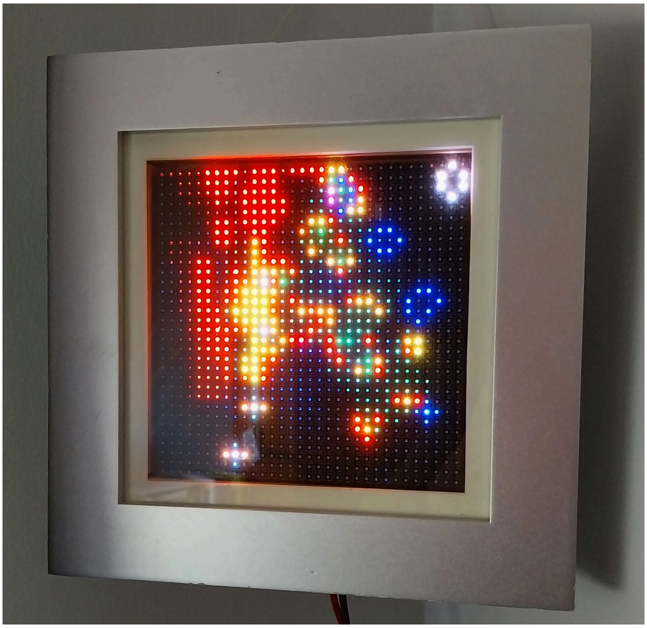
The model railway project had several kinds of input (magnetometer and microphone), and several kinds of output (motor and lights). When the train goes past the magnetometer, the motor pushes the disco ball and the torch turns on, so that lights dance on the ceiling. Also the LED lights flash on an off if somebody in the room starts talking.

The lights over the door utilise an “Addressable LED Strip” that has full colour LEDs, maybe 60 to a meter, that can be chained together. I added a microphone, and tacked the strips around a door. The code responds to sound by initiating a wave that moves around the door and reflects off the end of the strip. You can have a lot of fun choosing how the lights behave! Note that an extra power supply is needed for the LED strip, a 12 V supply that is plugged into mains power. The little dribble of power from a 9V battery is not enough.

Finally, the Raspberry Pi. It's a real computer, you can log in, use your favourite language and IDE, fetch libraries, and connect to is remotely. I got a 32x32 RGB array from Adafruit, and put it into a box frame, so there is room for the Pi in the back. The Pi connects to my house wifi, and read the data ingest status from my astronomical data feeds at the University of Edinburgh. It makes the coloured LEDs dance, showing me as I wake in the morning that the data is flowing in distant California. The parts list, instructions, and code are in in github[8].

---

[8] https://github.com/RoyWilliams/RGBMatrix.



12-7 The RGB matrix with data flowing in from the ZTF survey from Palomar in California.

13. Relaxing

One thing I have learned is to have a try at as many things as possible! I have learned to put down the urge to complete the work project and have time off. After all those years in the USA, where you are expected to be always on, all the time, it is great to be in the European culture, where work-life balance is respected, and it’s bad form to send work emails on the weekends.

I am a maker, not a real scientist. I like to build things: there is nothing as satisfying as making a tight dovetail joint or cutting intricate curves with a scroll saw. All the planning and drawing and sourcing the wood, the careful thicknessing and cutting and mortices and tenons, then it comes down to putting all the pieces together, which is where the heart is in the mouth. If the fit is too loose, then the furniture comes apart or wiggles about; but if the fit is too tight, the pieces may split and whole thing is ruined.

In Figure 13-1, from left to right is some furniture that I have made:

- A blanket chest. Mahogany with Penrose-tiled panels.
- Candlesticks in the style of Rohlfs, walnut.
- Curlicue corner chair in the style of Rohlfs, dyed maple.
- Drop-leaf circular table, mahogany.
- Corner chair with padded seat, cherry.
- Digital clock with decorative computer hardware, using rare large Nixie tubes sourced from 1950’s Russian submarines.
- Sofa, Stickley style, oak.
- Hall chair, Greene and Greene style, cherry.
- Picture frames, curly maple.

Figure 13-1 A selection of furniture from the last 30 years. Everything except the television!

I have enjoyed a bit of painting with acrylics, with a particularly active period during Covid lockdown in 2020/21: see Figure 13.2. On the left is an illustration of Morley's Theorem, a lovely bit of mathematics: given an arbitrary triangle (grey), trisect the angles and create a new

Figure 13-2 Two paintings: left, Morley's Theorem, and right, girl in a bikini with a urostomy bag

triangle from the intersections (red). The theorem simply states the astonishing result that the red triangle is always equilateral. On the right is a picture of a woman with a bikini and a urostomy bag – something that has been part of my life since an attack of bladder cancer in 2018.

Stained glass has also been a hobby for many decades, see Figure 13-3. On the left is a picture of a whale and its child, on the right a

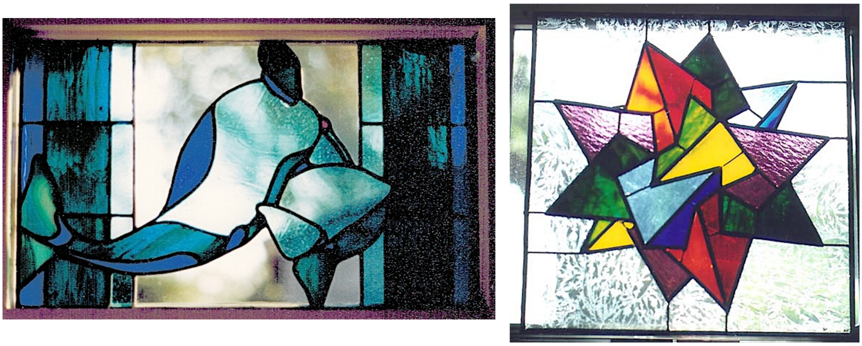


Figure 13-3 Stained glass: left whale mother and child, right, compound of five tetrahedra.

stellation of the icosahedron, also known as a compound of five tetrahedra. There are more examples of woodwork, painting, and stained glass at my website[9].

While on the subject of relaxation, I should mention cooking and weed. I like being in the kitchen because it's an escape from the screens: the computer screen of the data and software, and the phone screen of Facebook and Reddit and Bluesky and all that. Like the other crafts above, its hands-on, it's a learning process, it's a way to learn new ways to do things. And the weed, well that started many decades ago; I find it best to have none for a long time, then when I get some, just enjoy the freedom from anxiety and freedom of thought. When I have been thinking hard about something for weeks, it's as if the weed opens the gates to fresh ways of seeing the goal and how to get there. I should say, by the way, that doing weed all the time just dulls the mind and there is no creativity at all! I have been a regular visitor to the Netherlands for 30 years, not only because weed is easily available, but because I like being in Europe, because people speak English, because the Spring arrives a few weeks before it does in Scotland, because there is great art. I like to walk along the sides of the canals (*grachten*) and imagine the women of

[9] https://roywilliams.github.io/

Vermeer behind the windows, the bustling traffic of boats with merchandise being lifted by ropes to the warehouse doors high above. I like the way the transport systems are properly integrated, and people travel by bicycle, I like that the government is there not for themselves, but to serve the public interest. I always feel relaxed and tranquil in Amsterdam or Leiden or Haarlem.

14. Perspective

Deep Time, Deep Space, and Deep Data

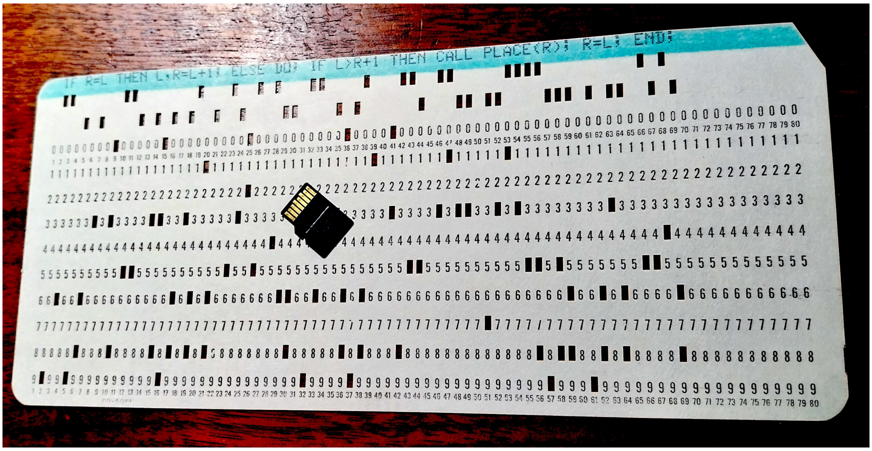
I have written already about James Hutton and his realisation in the late 1700s that the Earth is much, much older than the human-centric counting of generations from the holy book. But making that quantitative took quite a bit longer. Researchers considered the rate at which salt was eroded from rocks and how long it would take for the sea to become salty; but they did not know that salt is also added from deep ocean vents and removed from seawater through subduction. Scientists measured the rate of sedimentation and compared with the thickness of the sedimentary rocks; but the answers varied widely. Lord Kelvin used the internal temperature of the Earth, and asked how long it would take to cool down; but did not account for the heating due to natural radioactivity. It was Bertram Boltwood in 1907 who used the radioactive decay of Uranium to Lead to compute the ages of rocks, and found the answer – billions of years – that Hutton had intuited 150 years previously.

The study of deep space involves the "distance ladder". Each step is a different kind of measurement, and the final result is a product of many factors. Eratosthenes, 2000 years ago, found the diameter of the Earth, by looking at the shadow of a vertical pole at two places of different latitude: Alexandria in Egypt at latitude 31 degrees, and Aswan, far to the south, at 24 degrees latitude. By combining the distance (800 km) and the angular difference (7 degrees), he figured the size of the Earth. Giovanni Cassini and Jean Richter, in 1672, computed to the distance to Mars by parallax: Richter, in Guyana, saw Mars at a slightly different place in the sky than Cassini, in Paris, and the angular difference, combined with the distance from Paris to Guyana, gives the distance to Mars, and hence the scale of the solar system itself. Parallax was used again, by Bessel in 1838, to get the distance to stars, because of their small change in position over six months; the diameter of the Earth's orbit, combined with the angular change, gives the distance to the star. The next rung of the distance ladder is though "standard candles" variable stars whose luminosity is determined from the way the brightness varies. The classic standard candle is the cepheid variable: we see a nearby example, and know its distance from Bessel's method, then see a very distant example, and deduce the distance. In this way, Henrietta Swan computed the size of our galaxy, and Edwin Hubble the distance to the nearest other galaxy (the Andromeda galaxy). Other standard candles are specific types of supernovae (SN Ia) that take us out to the edges of the Universe at 14 billion light years.

Deep time and deep space took hundreds of years to get the true scale, but deep data has happened in my lifetime. It's less of a search for truth, and more of a thrill ride through the megabytes, gigabytes, and terabytes, each being 1,000 times as much data as the previous. Scientists in the past worked hard for each measurement – each 4-byte number – but now the data is gushing out so fast they need machine-learning and citizen science just to look through it. You can see the media in Figure 14-1: in 1972 I lugged thousands and thousands of those cards on the London Underground and it was really heavy and wasn't even a million bytes. Now the Micro SD card is so tiny and yet carries trillions of bytes.

The IBM 7094 on which I learned programming had 150 kilobytes of "core" storage, made by hand with tiny magnetic rings and fine copper wire. The "large scale integration" of the 1970s etched layers on a silicon wafer to make transistors. This kicked off Moore's law, where the data per gram doubled every two years, along with the processing capability to make it useful. In between were the spinning disks and tapes, all made from domains of magnetic iron that could have a field induced, going one way or another for the binary 0 or the binary 1. Now we are so profligate with data. The tiny terabyte MicroSD card can store a million photos, ten million pages of documents, 20,000 hours of music, a month of streaming movies for a household. It is incredible to me. The Rubin telescope, that I have worked on for several years, will deliver 20 terabytes *every night*. If only human wisdom has increased as much as data storage capacity. If only!

I remember moving into a different office at Caltech in 2010, it had a row of file cabinets on each wall, taking up most of the space in the room, with a narrow path in the middle to get to the desk. Once they were removed, the room was so much larger. I notice now, in movies and TV before 2010, how offices had so much space allocated to the storage of paper files. But now it is stored in remote data centres, available easily and searchable. And fluid in a way: documents and spreadsheets can be combined, summarised and joined in ways that in the past would have taken a long time and a lot of effort.



14-1: A computer card from my youth carries 30 bytes per gram; and a 2TB Micro SD card from 2025 carries 1,000,000,000,000 bytes per gram.

Unification is a Satisfying Joy

Sometimes we learn and then there is a surprise, a realisation that two things are connected, or a unification: what we thought was two things is actually one thing. When people talk about Isaac Newton "inventing" gravity, they mean that Newton was the first to understand the great reach of the gravitational force: in the short range it makes the apple fall from the tree, vertically down towards the centre of the Earth; but it is also the force that controls the Moon's movement around the Earth, and all the planets around the Sun. This was the ah-ha realisation that allowed Newton to compute the motion of the planets with such precision. One of the greatest achievements was understanding the orbits of comets; Edmund Halley realised that three previous comets were actually the same body, and predicted it would return in 1758, which brought popular prestige to Newton's theory.

The greatest Scottish scientist, James Clerk Maxwell, was fascinated by colour as a child, and indeed all of physics. He knew the detailed measurements of electricity from the work of Faraday, Franklin, Volta, Ampere, and others; but he was also a brilliant mathematician, and was able to condense the laws and theories already made into a single set of equations describing the variation of electric and magnetic fields. The symmetry of the new equations led to the idea that electricity and magnetism were different expressions of the same underlying system of "electromagnetism". Maxwell realised that his new equations had a wave-like solution, and found that the speed of the waves is related to previously measured properties of electricity and of magnetism. But when he computed that speed, he recognised that it was the speed of light – leading to another unification, of electromagnetism and light. Maxwell had realised his boyhood ambition in a world-changing way: not just light and colour, but all the other: radio, Xray, microwave, infrared radiation. A truly magisterial unification!

Celestial Mechanics was a huge success in the 1700's and 1800's. Mathematicians such as Euler, Laplace and Legendre developed perturbation theory and understood the role of resonance and were able to predict the motions of the planets with great accuracy. As discussed before, Neptune was the first "mathematical" discovery of a planet. But there was a last discrepancy to be cleared up between observation and Newton's theory: the motion of Mercury. Many observers tried to follow the success of the Neptune discovery by hypothesising "Vulcan", a planet inside the orbit of Mercury, and many false observations were made and debunked. But Einstein provided the real answer, through another Unification: this time between gravity and geometry.

It starts with the so-called Luminiferous Ether, a substance supposed to fill all space, allowing the propagation of light, gravity, and magnetism. Numerous experiments attempted to isolate the Ether, but nothing was found. The Michelson-Morley experiment in 1887 considered the speed of the motion of Earth through the Ether, and hypothesised that the speed of light would be greater in one direction that the other. However they did not find the expected difference, they found that the speed of light is the same in all directions, independent of the motion of the observer through the Ether. It was Einstein who grasped the bull by the horns, and reformulated space, time, and gravity starting with the axiom that the speed of light is always constant.

Einstein was able to break the logjam and dispose of the ether for ever. In an astonishingly daring move, he posited that the vacuum speed of light is everywhere constant, and derived laws of motion from there. In the old world, a bullet shot at speed u from a train moving at speed v has velocity relative to the ground of u+v. No longer true in the Einstein world of special relativity. Further to the theory, information cannot travel faster than light, so the gravitational influence of a body cannot be felt instantaneously. In the further theory – general relativity – Einstein rewrote gravity as a geometrical effect: a mass distorts the geometry, so a test particle moves on a curved trajectory in that geometry. The curved trajectory is not because of a gravitational force, it's because space-time itself is distorted.

Another great and magical unification was recognised by Emmy Noether, a German mathematician. Her theory, published in 1918, unifies symmetry and conservation laws. The laws of motion are the same everywhere in the Universe – which is the same as saying that momentum is conserved. The laws of motion are the same yesterday as today – the same as saying that energy is conserved. The laws of motion are the same even if I rotate my equipment – the same as saying angular momentum is conserved.

What Have We Learned

The most obvious thing, from my point of view, is that computers plus their applications have created much of our world, over the last 50 years. It was luck for me to stumble early into what would become a huge industry, blossoming into spreadsheets, desktop publishing, the internet, and so many facets of our lives. Back in 1972 when I took the box of cards to Imperial College, computers were for just that: computing. For calculating planetary motion, simple simulations, rocketry, technical things. Enjoying maths and physics combined perfectly with computers – as I went through life, I found there was always a job for me, not just writing software, but also understanding what the science team wants to achieve from that software.

In 2001, my boyhood fascination with astronomy was allowed out, and I tried to catch up my knowledge to sound like I knew what I was talking about. It was a time when both the US and UK governments put lots of funding into information technology, and the astronomical community started building the "virtual observatory": a set of standard interoperable services so that astronomical data can be easily exchanged. Astronomy has really changed in the last decades from a useless degree to a degree that gets a job with the big tech companies; instead of telescopes and galaxy dynamics, the graduate student and postdoc learns about software, data mining, and machine learning, some really valuable skills.

I have a friend here in Edinburgh who made five babies in her 20's and now is a great grandmother with over 30 descendants. All of them live locally and provide a lively social structure. Whereas I started off with very few blood relatives, and I have moved constantly – London to California, to Oxford, back to California, then to Scotland. As you get older, the roots and family become more and more important. I started as part of a small nuclear family of four with no relatives within a thousand miles, and never met any of my four grandparents. Then all the moving makes it difficult to keep up with friends. So I have always wondered what it must be like to be surrounded by a ready-made community, instead of having to constantly make the effort to meet new people. But there are still the friends from when we were young. No matter how long we are apart it's as if nothing has changed when we meet up. With age the desire for roots increases: I see stories of these "digital nomads" flitting from country to country and wonder how they will cope when they finally settle down at 60 with nobody local from their numerous past lives.

My first marriage didn't work out for me. It seemed like all through the 1990's I was yearning for home and family life, while Barbara was yearning for a successful career. Sometimes I would wake in the night and think is this what life will be forevermore? When we are young everything is possible, the horizons stretch to infinity, ever-expanding, and we are free to shape and reshape ourselves. Change comes from the inside, and if you don't like things, its easy to change city, change job, change lover, and it will be better. But when you are older, the paths are restricted, and you can see, or think you can see, all the way to the end. And if you don't like that future, then you get anxious. I think they call it a midlife crisis. Anyway, in 1998 I looked out at my future and didn't see anything changing. Stay at Caltech, stay with Barbara. And yet I saw other ways the world could be. Her family was far away, and I liked them very much, but they weren't local, so there was no family without great travel.

I decided that what I would like is a new start, and to be a father – which finally happened at the age of 46. When you know in your heart that you have to do something, when it's a continual ache, its best to grasp the nettle and just do it. So that was that. I found dating an entirely different matter from what it was in my 20s. Women in their 30s want something different, they want somebody reliable and employed, rather than the rock-climber who looks like a film star. I met Jessica Goeller in 1999, smart, happy, and self-sufficient, and we married in 2001, and are still happily together. Now our children have grown and left home.

Twice in my life I have avoided death by a whisker. The first was a motorbike crash when I was 22, life saved by a big helmet wrapping all round my face. Three weeks in hospital, eating through a straw for six weeks, steel pin in the leg. While in the hospital, people come visit and bring cards and flowers, but it's not bad because of a continuous course of Demerol – floating above the pain. It's getting home when it hurts, no more painkillers, and physical therapy forcing the damaged limb to work again. The second avoidance was cancer at the age of 60, very nasty and malignant, but luckily the bladder is one of the organs you can live without. As of this writing, six years cancer free (knock on wood), but with the endless possibility of leaking from the plastic pee-bag stuck to my belly. Being near death really makes you think of who you are, how people would respond to your absence, and what you promise to do if only you can survive.

Social media

Computers and social media have been rather a mixed blessing, haven't they? On the one hand, I can keep up with people I knew decades ago in a way that is much closer than writing a Christmas card once a year. I can call up any music I want to hear even if I never bought the record. I can find a recipe for pears with stilton, I can find out the price of jet fuel, I can get a street view of anywhere in the world, all so easy. But there is also a loss of serendipity, you can't go to the record shop and see who is hanging out there; you have a choice of so much media that your choices follow a familiar furrow and its difficult to see the new and exciting. Travel is diminished by multinational brands displacing local colour. Having vast choice leads to flicking though dozens of movies on Netflix, none of them quite right, rather than seeing whatever is on at the local cinema. The young have a vast choice of possible romantic dates, which of course leads to the inability to commit, not like the old days when you met somebody perfect and that was that.

Science high points

There have been a few signal times in my science life. The first and greatest, I think, was watching the moon landings as a child of 11. Timed for US prime-time, the landing and first footsteps were at four in the morning in London, and I insisted on staying up to see it all! Perhaps less well known – because it was a robot not a human – was the Voyager mission that visited Jupiter, Saturn, Uranus, and Neptune. I remember reading about the "Grand Tour" of the solar system in a magazine when I was a child in the 1960s, an orbit expert at JPL who grasped that the so-called gravitational slingshot was the key to visiting distant planets. Without it, it would be impossible to carry enough fuel and get to the required speed for the spacecraft to actually get to Uranus and Neptune in a decade – before the people driving the project have all retired. I watched the project go from imagination to implementation to launch to fabulous images. I was in my 30's, with a crowd at Caltech, when I watched the images come back from the Neptune encounter, all of us in awe of what humans could accomplish when they work together in a rich, free, diverse, and meritocratic society.

I worked on parallel computing in the 1990s, and at one point was in the "Delta" team. In a collaboration with Intel, we built what was for a while the world's fastest computer, a collection of 512 individual processors, which could work on a single simulation because of the sophisticated communication between them. There was a delicious display of coloured lights on the front of the Delta, which gave a real insight to the communication patterns. It was a real highlight to be part of that inspiring project.

One of the simulations that ran on the Delta and its descendants was the hypersonic flow around the Huygens spacecraft that landed on Titan, the moon of Saturn. Another highlight of my science life was to see that mission complete flawlessly in 2004, returning pictures from this utterly alien world, where water is as hard as rock and there are lakes and rivers of liquid methane.

Most recently, in 2016, was the first detection of gravitational waves. The hundreds of scientists in the know had kept quiet for six months after the actual detection, a frenzy of checks, writing the paper, and for me, preparing the data release so that the claim could be independently verified and peer reviewed. Something that is common now, but quite unusual at the time, was the release of not just the data, but also scientific software – as Jupyter notebooks – that would make it easy for others to verify the analysis themselves. There was a press release at Caltech, the world's press in attendance for a "significant announcement". The open data system was to be switched on at precisely the same time as the press conference and release of the announcement paper. By the end of that day, I was wreathed in smiles from ear to ear, because my website HAD NOT CRASHED in spite of thousands of requests and many many gigabytes.

Last Word

I know I have had a privileged life. And this account of my science life inevitably emphasises the successes over the failures, like those who post on Instagram the very best parts of their holiday without the bad bits. As the Ruskin quote says on the title page: "This is the best of me; for the rest, I ate, and drank, and slept, loved and hated, like another". There has been physical trauma, a serious road accident and cancer, but the marvels of modern medicine have pulled me through both times. I have not suffered as many have from emotional trauma: my parents were good people who did their best for my brother and I; there have been no terrible people dragging me down. I have always been envious of those with grandparents and close cousins, having never experienced either.

I am of the age (boomer) when the post World War II consensus provided extended peace to Europe and the US, and there were plentiful resources to extract (fossil fuel) so that a decent democracy could be funded. The decades of my youth were the glory times of the world, with advancements of civil rights and magnificent science like the moon landings. I was exposed to computers as a teenager ("Computing as a Sport") and discovered I loved their logic. I was lucky to go to a school with computer access at exactly the time I needed it, at a time when very few people saw the potential of the computer. I was lucky to have a mathematical brain and thus be able to attend one of the best Universities (Cambridge). I have always been active, I cannot understand those who can lie on a sofa all day, or a lounger beside the pool. I try, but then leap up to attend to an obligation or an idea.

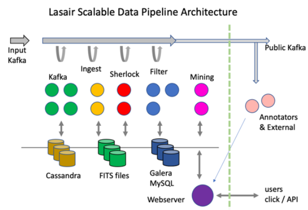
And now we have the future facing us, the hopes for our children and for all the children. Global warming is bearing down, and people's lives are worse than they were 50 years ago. People in difficult times turn to an authoritarian, which usually makes their life worse than if they had gone for democracy. We are not ready for the great heat and fire and flood, we are not prepared for a heatwave to kill a million, or drought causing long-term harvest failure, we are not prepared for the extra-powerful storms and sea-level rise. We are not prepared for mass migration, or the border wars, or the water wars. Instead of working together with science and education and self-sacrifice, we elect authoritarians who make the problems worse.

Whatever happens to the world in the future, at least we landed on the Monn, we detected gravitational waves, we built the Webb telescope, we thought about the deep mathematics of $\exp \pi\sqrt{163}$, and we discovered that gold is created in the collisions of neutron stars. Really hoping that humans can somehow get it together and hold on to our beautiful world, natural and artificial.

15. References

Note: Most of the papers below are available from
`https://roywilliams.github.io/`

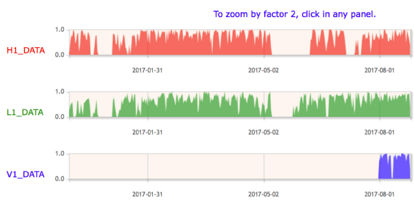
A systematically-selected sample of luminous, long-duration, ambiguous nuclear transients
P. Wiseman, R. D. Williams, et al
MNRAS 537 (2) 2024–2045 .



Enabling Science from the Rubin Alert Stream with Lasair
Roy D. Williams, Gareth P. Francis, Andy Lawrence, Terence M. Sloan, Stephen J. Smartt, Ken W. Smith, David R. Young
Click for RAS Techniques and Instruments, 3(1) 362–371 or arxiv 2404.08315.

NEural Engine for Discovering Luminous Events (NEEDLE): identifying rare transient candidates in real time from host galaxy images, X. Sheng et. al.
MNRAS 531(2) 2474–2492

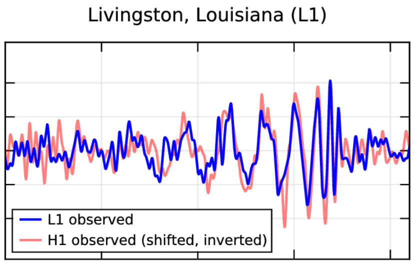
GW190425: Pan-STARRS and ATLAS coverage of the skymap and limits on optical emission associated with FRB190425, S. J. Smartt et. al.
MNRAS 528, 2299–2307 (2024).



Open data from the first and second observing runs of Advanced LIGO and Advanced Virgo
Abbott et. al. (LIGO collaboration)
Nature Scientific Data

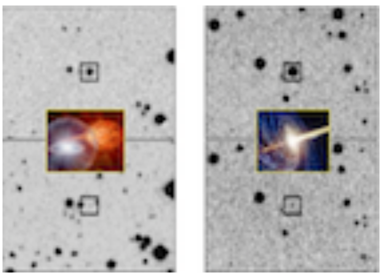
Lasair: The Transient Alert Broker for LSST:UK
Ken Smith, Roy D. Williams, D. R. Young, S. J. Smartt, A. Lawrence, D. Morris, S. Voutsinas, and M. Nicholl
Research Notes of the AAS, Volume 3, Number 1

GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral
LIGO Scientific Consortium
Phys. Rev. Lett. 119, 161101

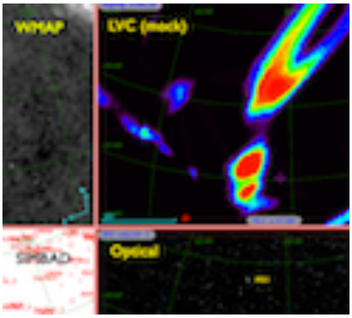


Observation of Gravitational Waves from a Binary Black Hole Merger
LIGO Scientific Consortium
Phys. Rev. Lett. 116, 061102

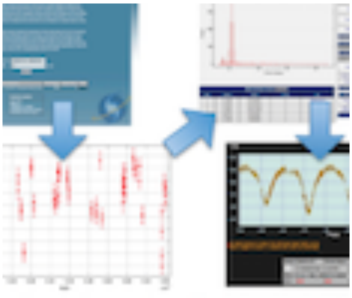
Responding to the Event Deluge
Roy D. Williams, Scott D. Barthelmy, Robert B. Denny, Matthew J. Graham, John Swinbank
Proceedings of SPIE Observatory Operations, Amsterdam, 2012



Astronomy with Cutting-Edge ICT: From Transients in the Sky to Data over the Continents (India-US)
Ashish Mahabal, Ajit Kembhavi, Roy Williams and Sharmad Navelkar.



Gravitational Waves and Time-Domain Astronomy
Joan Centrella, Samaya Nissanke and Roy Williams
Proceedings of the International Astronomical Union / Volume 7 / Symposium S285 / September 2011, pp 191-198



Using the VO to Study the Time Domain
Rob Seaman, Roy Williams, Matthew Graham and Tara Murphy
Proceedings of the International Astronomical Union / Volume 7 / Symposium S285 / September 2011, pp 221 - 226

Skyalert.org website: authoring, dissemination, annotating, mining, and messaging of immediate astronomical events. 2009 - 2013.

Sky Event Reporting Metadata (VOEvent)
R. Seaman, R.D. Williams, A. Allan, S. Barthelmy, J. Bloom, M. Graham, F. Hessman, S. Marka, A. Rots, C. Stoughton, T. Vestrand, R. White, P. Wozniak,
A Recommendation of the International Virtual Observatory Alliance, 2006.

First Results from the Catalina Real-time Transient Survey
A.J. Drake, S.G. Djorgovski, A. Mahabal, E. Beshore, S. Larson, M.J. Graham, R. Williams, E. Christensen, M. Catelan, A. Boattini, A. Gibbs, R. Hill, R. Kowalski, Astrophysical J. 696 (2009) 870.

National Virtual Observatory Summer School 2008
Roy Williams, Faculty Chair
Organized the content and teaching for a two-week intensive course on the Virtual Observatory, September 2008, Santa Fe, New Mexico.

VIM: Visual Integration and Mining
Integrate data from multiple positions and datasets.

Skyalert: Real-time Astronomy for You and Your Robots
R. D. Williams, S. G. Djorgovski, A. J. Drake, M. J. Graham, A. Mahabal.
Proc. ADASS XVIII (2008), PASP, eds. D. Bohlender, P. Dowler and D. Durand

Describing Data and Data Collections in the VO
Kent, B.R.; Hanisch, R.J.; Williams, R.D.
in The National *Virtual Observatory: Tools and Techniques for Astronomical Research,* eds. Graham M.J., Fitzpatrick M.F., McGlynn T.A. ASP Conference Proceedings, vol. 382.

Data-Intensive Computing in the 21st Century
Ian Gorton, Paul Greenfield, Alex Szalay, Roy Williams
Computer 05/2008; DOI:10.1109/MC.2008.122

Transient Event Reporting and Response with VOEvent
Seaman, R.; Allan, A.; Williams, R.
Astronomical Data Analysis Software and Systems ASP Conference Series, Vol. 394, Eds R.W. Argyle, P.S. Bunclark, and J.R. Lewis., p.213 (2007)

The IVOA in 2007: Assessment and Future Roadmap
Roy Williams and the IVOA Technical Coordination Group
International Virtual Observatory Alliance,

Google Sky mashup
Roy Williams and Andrew Drake
Real-time display of events from multiple streams. 2007.

Astronomical network event and observation notification
White R.R., Allan A., Barthelmy S., Bloom J., Graham M.J., Hessman F.V., Marka S., Rots A., Scholberg K., Seaman R., Stoughton C., Vestrand W.T., Williams R., Wozniak P.R.,
Astronomische Nachrichten, 327 (2006) 775.

Object detection in multi-epoch data
G. Jogesh Babu, Ashish Mahabal, S. G. Djorgovski, R. Williams
Statistical Methodology, Volume 5, Issue 4, p. 299-306.

The IVOA in 2006: Assessment and Future Roadmap
Roy Williams and the IVOA Technical Coordination Group
International Virtual Observatory Alliance,

HotGrid: Graduated Access to Grid-Based Science Gateways
Williams, R. Steenberg, C. Bunn, J. Lect. Notes Computer Sci. 3470 (2005) 78.

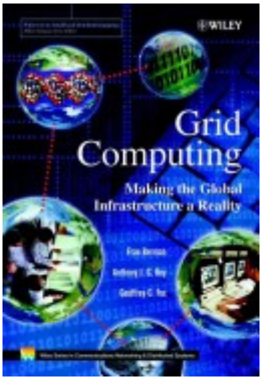
The IVOA in 2005: Assessment and Future Roadmap
Roy Williams and the IVOA Technical Coordination Group, International Virtual Observatory Alliance,

Virtual Observatory: From Concept to Implementation
S. G. Djorgovski and R. Williams
In: Science with Wavelengths on Human Scales, ASPCS, vol. 3xx. ASPCS

VOTable Format Definition
Francois Ochsenbein, Roy Williams, Clive Davenhall, Daniel Durand, Pierre Fernique, David Giaretta, Robert Hanisch, Tom McGlynn, Alex Szalay, Mark B. Taylor, Andreas Wicenec,
A Recommendation of the International Virtual Observatory Alliance.

An IVOA Standard for Unified Content Descriptors
S. Derriere, N. Gray, R. Mann, A. Preite Martinez, J. McDowell, T. Mc Glynn, F. Ochsenbein, P. Osuna, G. Rixon, R.D. Williams.
A Recommendation of the International Virtual Observatory Alliance.

Multi-Wavelength Image Space: Another Grid-Enabled Science
Roy Williams, Bruce Berriman, Ewa Deelman, John Good, Joseph Jacob, Carl Kesselman, Carol Lonsdale, Seb Oliver, Tom Prince
Concurrency & Computation, vol 15 (2003), pp 539-549.



Grids and the Virtual Observatory
Roy Williams
in Grid Computing: Making The Global Infrastructure a Reality by Fran Berman, Anthony J.G. Hey, and Geoffrey Fox, Wiley, 2003, pp 837-858.

Scientific Data Mining Integration and Visualization
Bob Mann, Roy Williams, Malcolm Atkinson, Ken Brodlie, Amos Storkey, Chris Williams, report of a workshop held in Edinburgh, 24-25 October 2002.

XML and Web Services for Astronomers
Roy Williams and Robert Brunner
tutorial at ADASS 2002, Baltimore, October 2002.

Network data analysis server (NDAS) prototype development
Szabi Marka, Benoit Mours, Roy Williams
Classical and Quantum Gravity 19 (7): 1537-1540 (2002)

Agent based data management in digital libraries
Y.Y. Yang, O.F. Rana, D. W. Walker, R. D. Williams
Parallel Computing 28 (5): 773-792 (2002)

Analysis of Shear Layers in a Fluid with Temperature-Dependent Viscosity
D. J. Estep, S. M. Verduyn Lunel, and R. D. Williams
J. Comp. Physics, vol 173 (2001), pp 17-60.

Web access to supercomputing
G. Aloisio, M. Cafaro, C. Kesselman, R. Williams
Computational Science and Engineering 3 (6): 66-72 (2001)

A Virtual Data Grid for LIGO
Ewa Deelman, Carl Kesselman, Roy Williams, Albert Lazzarini, Thomas A. Prince, Joe Romano and Bruce Allen
Lecture Notes in Computer Science, vol 2110 (2001) pp 3-12

Estimating the Error of Numerical Solutions of Systems of Nonlinear Reaction-Diffusion Equations
D. J. Estep, M. G. Larsen, and R. D. Williams
Memoirs Amer. Math. Soc., 146 (2000) number 696, pp 1-109.

Analysis of Shear Layers in a Fluid with Temperature-Dependent Viscosity
D. J. Estep, S. M. Verduyn Lunel, and R. D. Williams
J. Comp Phys. 173 (2001) pp 17-60

A Test Suite for High-Performance Parallel Java
Jochem Hauser, Thorsten Ludewig, Roy Williams, Ralf Winkelmann, Torsten Gollnick, Jean Muylaert
Advances in Engineering Software 31 (8-9): 687-696 (2000)

Grid computing on the web using the globus toolkit
G. Aloisio G, M. Cafaro, P. Falabella, R. Williams
Lecture Notes in Computer Science 1823: 32-40 (2000)

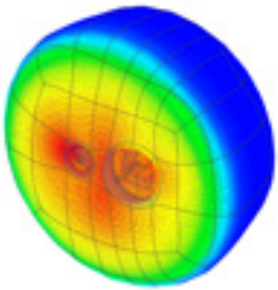
Approaches to Federation of Astronomical Data
Roy Williams
in Virtual Observatories of the Future, eds. R. J. Brunner, S. G. Djorgovski, and

A. S. Szalay
Astronomical Society of the Pacific, Conference Series 225. November 2000.

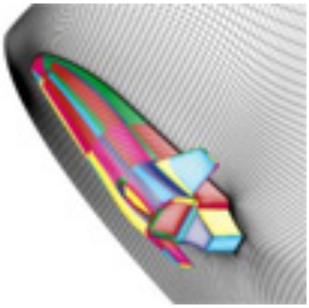
Special issue: Interfaces to Scientific Digital Archives,
ed. Roy Williams
Future Generation Computer Systems 16 (1): VII-VIII (1999)

The Digital Puglia Project: An active digital library of remote sensing data
G. Aloisio, M. Cafaro, R. Williams
Lecture Notes in Computer Science 1593: 563-572 (1999)

An XML Architecture for High-Performance Web-Based Analysis of Remote-Sensing Archives
Giovanni Aloisio, Giovanni Milillo, and Roy Williams
Future Generation Computing Systems 16 (1): 91-100 NOV 1999.



Strategies for Parallel and Numerical Scalability of CFD Codes
Ralf Winkelmann, Jochem Hauser and Roy Williams
Computational Methods in Applied Mechanical Engineering. 174 (3-4): 433-456 (1999).



A Pure Java Parallel Flow Solver
Jochem Hauser, Thorsten Ludewig, Torsten Gollnick, Ralf Winkelmann, Roy Williams, Jean Muylaert, and Martin Spel
AIAA paper 99-0549, proceedings of 37th Aerospace Sciences Meeting, Reno NV, Jan 1999.

Report of the European Union - United States Workshop on Large Scientific Databases
Roy Williams, Paul Messina, Fabrizio Gagliardi, John Darlington, and Giovanni Aloisio
October 1999

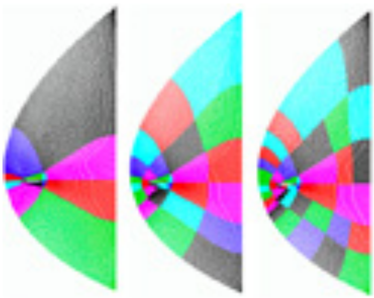
XSIL: Extensible Scientific Interchange Language
Kent Blackburn, Albert Lazzarini, Tom Prince and Roy Williams
Lect. Notes. Comp. Sci. 1593 (1999) 513

Interfaces to Scientific Data Archives, Report of the NSF Workshop
Roy Williams, Julian Bunn, Reagan Moore and James Pool,
May 1998.

Paraflow: A dataflow distributed data-computing system
Roy Williams and Bruce Sears
Lecture Notes in Computer Science 1401: 556-565 (1998)

Metacomputing - From ideas to real implementations
Wolfgang Nagel and Roy Williams
Parallel Computing 24 (12-13): 1709-1711 (1998)

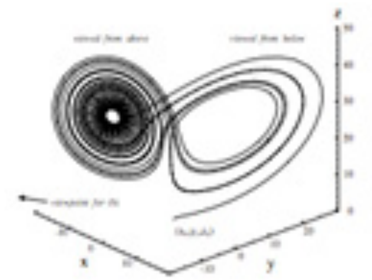
A Distributed Web-based Metacomputing Environment
G. Aloisio, M. Cafaro, P. Messina, R. D. Williams,
in HPCN97, Vienna, April 1997.



A Tangled Web Strategy for Numerical and Parallel Scalability in Aerospace Simulation
J. Haeuser, R. D. Williams and R. Winkelmann
Proceedings of CFD97, University of Victoria, Canada, 25-27 May 1997.

Dreaming of Hypermaps
Roy Williams
Caltech Engineering and Science magazine,

Error Estimation for Numerical Differential Equations
D. J. Estep, S. M. Verduyn Lunel and R. D. Williams
IEEE Antennas and Propagation 38 (2): 71-76 (1996).



Accurate Parallel Integration of Large Sparse Systems of Differential Equations
D. J. Estep and R. D. Williams
Mathematical Models and Methods in Applied Sciences 6 (4): 535-568 (1996).

Efficient Convergence Acceleration for Parallel CFD Codes
R.D. Williams, J. Hauser and R. Winkelmann,
Proceedings of the Parallel CFD '96 conference, Capri, Italy
July 1996.

"Infinite" Information on the Internet
Roy Williams
Caltech Engineering and Science magazine

A Newton-GMRES Method for the Parallel Navier-Stokes Equations
J. Hauser, R.D. Williams, H.-G. Paap, M. Spel, J. Muylaert and R. Winkelmann
Proceedings of the Parallel CFD '95 conference, Pasadena, California
July 1995.

Parallel Computing Works!
G. C. Fox, R. D. Williams and P. C. Messina,
Morgan-Kaufmann, San Francisco, 1994

Strategies for Parallelizing a Navier-Stokes Code on the Intel Touchstone Machines
J. Häuser and R. D. Williams,
Int. J. Numerical Methods in Fluids, 15, 51 (1992)

Voxel Databases: A Paradigm for Parallelism with Spatial Structure
R. D. Williams,
Concurrency, 4 (1992) 619.

Adaptive Parallel Meshes with Complex Geometry
R. D. Williams
in Numerical Grid Generation for Computational Fluid Dynamics, ed. A.-S. Arcilla et. al., Elsevier 1991.

Performance of Dynamic Load Balancing Algorithms for Unstructured Mesh Calculations
R. D. Williams,
Concurrency, 3 (1991) 457.

Adaptive Parallel Meshes with Complex Geometry
C. F. Baillie, R. D. Williams, S. M. Catterall and D. A. Johnston
Nucl. Phys B348, 543 (1991)

Performance of Dynamic Load Balancing Algorithms for Unstructured Mesh Calculations
R. D. Williams,
Concurrency, 3, 457 (1991)

Benchmarking Advanced Architecture Computers
P. Messina, C. F. Baillie, E. W. Felten, P. G. Hipes, D. W. Walker, R. D. Williams, et al,
Concurrency, 2, 195 (1990)

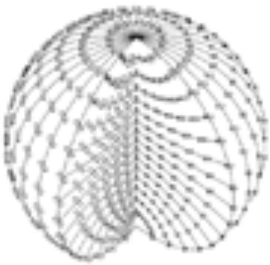
Crumpling Transition in Dynamically Triangulated Random Surfaces
C. F. Baillie, R. D. Williams, D. A. Johnston
Nuclear Physics B 335 (2): 469-501 (1990)

Computational Aspects of Simulating Dynamically Triangulated Random Surfaces
C. F. Baillie, R. D. Williams, D. A. Johnston
Computer Physics Communications 58 (1-2): 105-117 (1990)

Distributed Irregular Finite Elements
R. D. Williams
A software package for distributed parallel computing with unstructured triangular and tetrahedral meshes.

Supersonic Flow in Parallel with an Unstructured Mesh
R. D. Williams
Concurrency, 1, 51 (1989)

Free-Lagrange Hydrodynamics in Parallel
R. D. Williams
Parallel Computing 7, 439 (1988)



Two Transitions in Tangentially Anchored Nematic Droplets
R. D. Williams
J. Phys. A19, 3211 (1986)

Critical Spin Dynamics of Europium Oxide
S. W. Lovesey and R. D. Williams
J. Phys. C19, L253 (1986)

Scattering Response of a Phonon Damped Harmonic Oscillator
R. D. Williams, S. W. Lovesey and W. Renz,
Z. Physik B64, 129 (1986)

Neutron Scattering from a Substitutional Mass Defect
R. D. Williams and S. W. Lovesey
Z. Physik B62, 413 (1986)

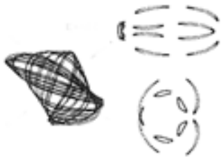
Sub-Saturation Phases of Nuclear Matter
R. D. Williams and S. E. Koonin,
Nucl. Phys. A345, 844 (1984)

Bound-State Decay of Rhenium-187
R. D. Williams, W. A. Fowler and S. E. Koonin,
Astrophys. J. 281, 363 (1984)

Semiclassical Quantization in Many Dimensions
R. D. Williams,
Ph.D. Thesis, California Institute of Technology, 1983

The Coupling of Phonons to a Helium Atom Adsorbed on Graphite
R. D. Williams, M. W. Cole and S. E. Koonin,
Phys. Rev. B28, 1076 (1983)

Atomic Final-State Interactions in Tritium Decay
R. D. Williams and S. E. Koonin,
Phys. Rev. C27, 1815 (1983)



Semiclassical Quantization of the Shell Model
R. D. Williams and S. E. Koonin,
Nucl. Phys. A391, 79 (1982)

Direct Capture Cross Sections at Low Energy
R. D. Williams and S. E. Koonin,
Phys. Rev. C23, 2773 (1981)