

Chapter 16

Visualization as Fine Art

16.1 Introduction

The best way to learn graphic design is to do a lot of looking. Because of this, part of an education in scientific visualization is a course in fine arts. Instead of looking at paintings and sculpture in a museum or the slide show of a fine arts course, one needs to look at a lot of exemplars in scientific visualization.

Merely looking is insufficient. We need to look **ACTIVELY** rather than passively. That is to say, to learn from looking at graphs, we need to ask ourselves as we look:

- What information is this graph trying to convey?
- What aspects of the design enlighten?
- What aspects of the design are confusing, or require the reader to work hard to decode the message?
- Would a different format of graph have been more informative?
- What could be changed to improve this graph?

The second qualifier is that a scientific graph is only as good as the science. Many Nobel Prizes have been won by papers that did not contain a single chart or diagram. Conversely, pretty visuals can't make a bad idea influential. However, they may disguise the smell of a stinky concept for a time.

The third qualifier is that we have to look at visualization with the eye of a commercial illustrator. Most professional artists work at a mixture of advertising drawings and book illustration. The purpose of the former is sell toothpaste; the latter is to sell ideas. Scientific visualization is about selling ideas, too.

We do not wish to imitate the dishonesty of much advertising. However, every parent has had the experience of being presented with a drawing of bright splashes of yellow, red, brown and black by a small child.

“That’s a very nice, dog, Emma.”

“It’s not a dog, Daddy, it’s a person!” (Indignantly, followed by a snort.)

Metaphorically speaking, we want to design our scientific drawings so that one can tell the dog from the person.

16.2 Gilbert: *De Magnete*(1600), Illustration of Experimental Method



Figure 16.1: A woodcut from William Gilbert's treatise on magnetism (1600). It shows the manufacture of a magnetized bar. If the bar is heated and then placed on the anvil so that its ends are oriented towards magnetic north (*septentrio*) and south (*auster*), then after the bar cools, it is magnetized. This suggested to Gilbert (correctly!) that the earth itself was a magnet.

Fig. 16.1 is taken from a Latin treatise by William Gilbert (1544-1603), personal physician to Queen Elizabeth I and King James I. This illustration shows technique rather than concept: by heating an iron bar and orienting it parallel to the earth's magnetic field, one can magnetize the bar.

A modern book would probably use a photograph instead of a woodcut. Still, Gilbert's diagram requires choices that face modern illustrators, too.

The first issue is that visualizations should always be stripped to the minimum; extraneous graphical elements merely distract without enlightening. In Gilbert's picture, the essentials are the iron bar and the directions north and south. The table, windows, and other furnishings and the earnest man pounding away with a hammer are all irrelevant to the science *per se*.

However, it is very important to always remember the background and experience of the readership. In Gilbert's time, when there were no professional scientists, his readers were much less comfortable with abstract, minimalist graphics. Gilbert's picture that shows an actual man banging a hammer on an anvil was probably quite appropriate for *his* audience.

In our time, we can get away with more abstraction than was possible in 1600. However, the scientific literature is riddled with graphs which use esoteric terminology or graphical elements which are perfectly comprehensible to the author and her drinking buddies, but as cryptic as Etruscan to the other 99% of her readers.

Another issue is that Gilbert's figure is a pseudo-photograph: When are photos better than drawings and vice-versa? Alley (1996) writes "The major advantage of photographs is realism. For instance, once readers see a photograph of the eruption of Mount St. Helens, they know that the volcano has, in fact, erupted. ... This advantage of realism is also a disadvantage. A photograph not only shows a subject's true textures and tones, but also its true shadows and scratches. These extraneous details can easily confuse the reader." (pg. 153.)

What Alley means is that a photograph of an experimental apparatus will inevitably pick up the rest of the room. A schematic drawing, in contrast, can be limited to just the essentials. Furthermore, the parts can be moved around (as in a so-called "exploded view"), shown in different sizes, and manipulated as needed for clarity.

Alternatively, one could crop a photograph by cutting away the unessentials. This is very common in non-scientific graphics. For example, the children's guides from Doris-Kindersely publishing contain pictures of many objects that are cropped to the irregular outline of the object itself. Unfortunately, most scientific authors and journals are not this skillful, so scientific readers expect rectangular photos. If one crops a photo to an irregular, one needs to explain what has been cut away, and its irrelevance, or the scientific reader may become suspicious. Even if this is done, there is still the problem of showing all the "shadows and scratches" in Alley's words.

In short, photographs are for faith, drawings for concepts.

One qualifier: Gilbert's illustration is a depiction of *method* and *apparatus*. When the photo is of the phenomenon, as in an astronomical image, a whole set of additional issues arise.

16.3 Cosmology Models of the Seventeenth Century

Fig. 16.2 is not merely a graphical representation of a theory, but is a comparison between six theories. Since it is extremely difficult to compare six verbal descriptions, the graphical comparison — all the key ideas on a single page — is very valuable.

The illustration is clearly labelled with the name of each system. This is a sound practice that too many authors forget, three-and-a-half centuries later.

The sun and the earth are labelled as icons — the earth as a disk with the continents, the sun as a human face with rays. The solar face is false to the physics. However, such representations were common in seventeenth century art and thus familiar to the readers.

The planets are labelled with the astrological signs for the planets; Venus, associated with the Roman Goddess of Love, is marked with a stylized mirror. Although most modern readers find them mystifying, these labels-by-icon were entirely appropriate for the readership of the maker's time.

One theme inherent in this graph is therefore: label everything, and label it using symbols appropriate to your readers. If you need to use symbols or abbreviations that may not be obvious to some of your readers, explain them clearly in the caption!

If the reader enjoys decoding, he can buy a book of cryptogram puzzles at the supermarket. Don't turn your graphs into an exercise in cryptography! The annoyed or disgusted reader may be reviewing your next grant proposal.

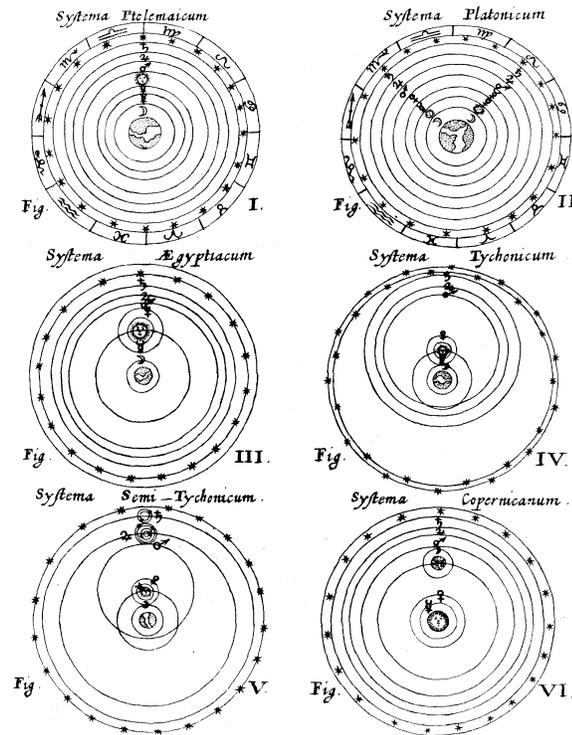


Figure 16.2: An early seventeenth century comparison of six models of the solar system. The first five are geocentric models with the rest of the universe revolving around the earth; the scheme in the lower right is the Copernican system with the sun at the center.

16.4 Nodal Patterns of the Vibrations of a Thin Metal Plate

Ernst Chladni (1756–1827) was educated as a lawyer, groomed to take over his father's legal practice. He turned to science instead. Because he was also a very good amateur musician, he became especially interested in acoustics. *Discovery of the Theory of Pitch* was the first broad treatise on acoustics.

He mapped the normal modes of a thin metal plate by sprinkling sand on top of the plate and then drawing a violin bow across different points on the edges of the plate. This excited various eigenfunctions of the plate. As the metal jiggled up and down, the sand was expelled from the moving parts of the plate. The sand collected in the nodal lines where the eigenfunction has no motion. Chladni then drew the patterns he observed to classify the structure of the oscillations.

Fig. 16.3 shows a sample of his drawings.

The first theme is that to make a great scientific visualization, do great science. The eigenfunctions of vibrating plates and membranes of various shapes is still a subject of intensive research two centuries after Chladni's drawings were first published. The title of a paper by the late, great Mark Kac, "Can You Hear the Shape of a Drum?" was answered only in the early 1990's; it has been proven that there are plates of different shapes that have exactly the same eigenfrequencies. Joyce McLaughlin has made a great reputation in mathematics by developing a general theory for the inverse theory of how different shapes of membrane or plate are associated with different patterns of frequencies and nodal curves.

The second theme is that sometimes geometry = physics, as here. The nodal pat-

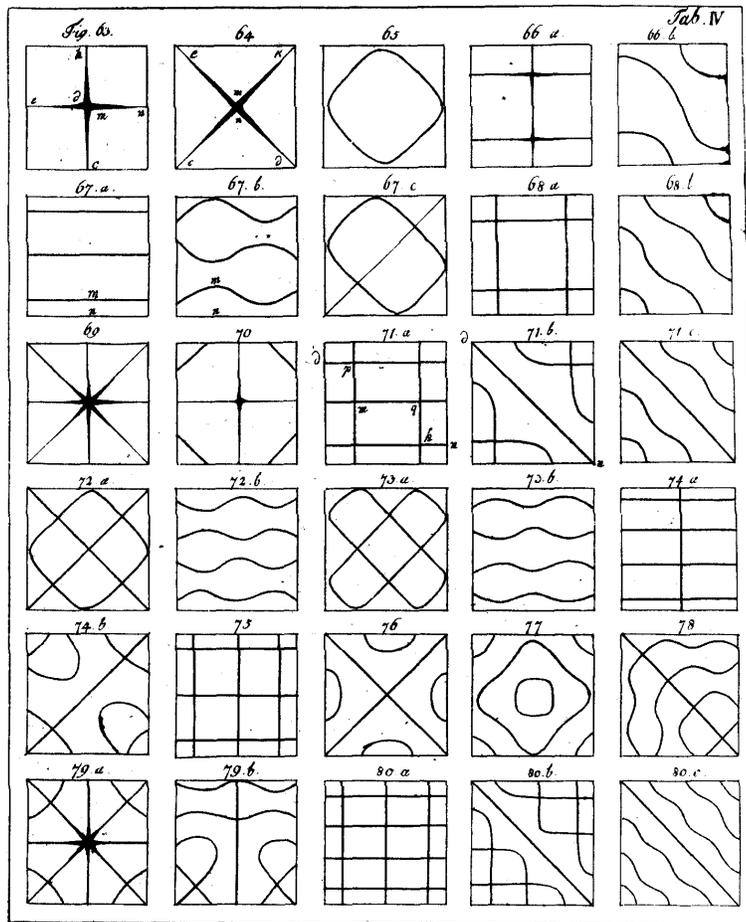


Figure 16.3: Nodal patterns, observed by sprinkling sand on an oscillating metal plate, from Chladni's 1802 book on acoustics. Each graph illustrates the zero-curves of a different eigenfunction.

terns and frequencies change in subtle ways as the shape of the vibrating membrane is altered. It is difficult to imagine a verbal description that could do justice to Chladni's complicated patterns.

The nodal patterns are very clear and each eigenmode is numerically labelled. However, the graph is incomplete as presented. In acoustics, the sound is the center of everything. The sound of a mode is its eigenfrequency, which is not given in the figure. It follows that this illustration needs to be accompanied by a caption or table which lists the frequencies or tones of each eigenfunction.

16.5 Darwin's Flowers

Fig. 16.4 illustrates a major contribution of Charles Darwin's which was unconnected with his more famous work on evolution. The mystery is that the common cowslip is a flower that occurs in two forms, one long-styled, and the other short-styled. (The "style" is part of a plant's reproductive organ.) Darwin at first believed that the two physical forms were the male and female of the species. Later, he realized that actually both varieties were hermaphrodites, that is, possessed both male and female organs. One drawback of hermaphroditism is that it is easy for a plant to reproduce with itself, creating natural clones that defeat the whole purpose of sexual reproduction. Darwin showed that the peculiar geometry of these plants ensured that the long-styled variety could only be fertilized by a short-styled plant, and similarly a short-styled plant could only reproduce with a long-necked plant. The geometry prevented self-reproduction.

The left figure shows what the discussion is all about: the cowswill has two forms. A photographic reproduction would hide the interior organs, so Darwin opts for a cut-away drawing to illustrate the all-important interior. The petals are extraneous to the discussion of sex organs. Still, it was not imprudent to include them because the petals remind a botanist which species is under discussion — a pictorial label at the top of the figure to reinforce the verbal label at the bottom. Also, the petals show the position of the styles, which are hidden from direct view, to the stem and petals that are directly visible.

The right diagram is conceptual. It sums up in a single visual image the essence of Darwin's theory about cowsill reproduction. Notice the profusion of labels. At the bottom, the long-styled and short-styled forms are labelled. But Darwin does not stop there; he labels each of the arrows, too. Lest we have any doubts, the direction of the arrows show how the pollen moves in reproduction — or tries to move — for what Darwin calls "illegitimate unions".

The only criticism is that "legitimate" and "illegitimate" are odd-sounding terms, reminiscent of lawyer-talk more than precise scientific lingo. "Impossible" and "possible" might have been better.

Still, these are good figures, and remind us that Darwin was more than just the "Theory of Evolution" guy. But there is one final reason why these figures are good: they are excellent scientific visualizations because Darwin was an excellent scientist!

16.6 Hertzian's Discovery of Electromagnetic Waves

Fig. 16.5 shows a milestone: the experimental verification of electromagnetic waves. Hertz could have made measurements only along one or two rays radiating away from two-ball antenna, and plotted the results as a line graph. However, he was trying to confirm James Clerk Maxwell's theory, which made quite specific predicts about the

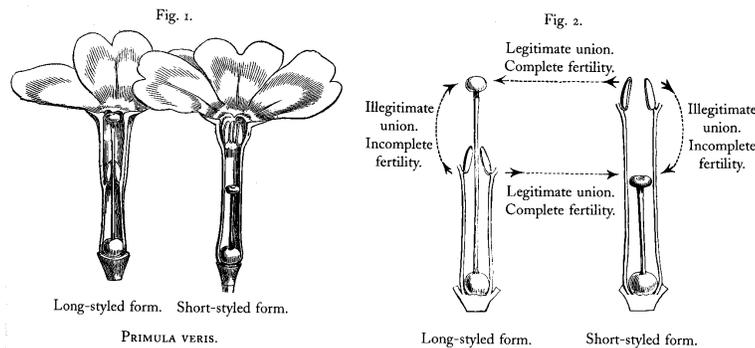


Figure 16.4: From Charles Darwin's *The Different Forms of Flowers* (1862). The left figure shows the two forms of the cowslip, *Primula veris*. The right is a conceptual diagram: it illustrates Darwin's thesis that the geometry prevents self-pollination, forcing the cowslip to always reproduce sexually.

form of the wave. And so he made sufficient measurements to make a contour plot, which made his experiment much more convincing.

Note that vertical axis running through the middle of the plot, carefully labelled. The units are not arbitrary metric or English units, but rather the tic marks are placed at integral or half-integral multiples of λ so the scale of the waves is measured in terms of physically relevant length scales.

Maxwell's theory predicts that the waves would be polarized. To what is otherwise a standard contour plot, Hertz added arrows to mark the directions of polarity.

The graph is not flawless. Negative contours could be dashed or hachure lines added to distinguish between the troughs and peaks at a given time. The wave oscillates, of course, so that every portion of space has both positive and negative values. However, there are phase relationships such that one lobe is positive whenever the major feature on the other side of the antenna is negative and vice versa; Hertz has not attempted to indicate these, nor has he labelled any of the individual contours. Nevertheless, the graph is sufficiently clear and detailed that the confirmation of Maxwell's theory is plainly visible.

16.7 Byerly (1893): Integrating Text and Graphics

Fig. 16.6 is an illustration from a treatise on Fourier series and spherical harmonics by Byerly which was first published in 1893. The graphical problem is that Gibbs' Phenomenon occurs in a very narrow layer. Byerly therefore found it expedient to draw a graph with a very high aspect ratio to represent the narrowness of this layer of high error discovered by Gibbs. On a page by itself, the graph would look rather peculiar, surrounded by so much white space. The printer solved this problem by shifting the graph to the left and filling the rest with text.

Unfortunately, it is rather difficult to imitate this 19th century piece of sound design with late 20th century computer technology. The usual \LaTeX style and class files, for example, force all figures to fill the entire width of the page. Some publishers, such as Kluwer, supply their own style files which allow captions to be set to the side of the figure, but not the main text. Allowing text to flow freely around the side of a figure is rather tricky although one can find user-contributed files on Web \LaTeX sites which purport to allow this.

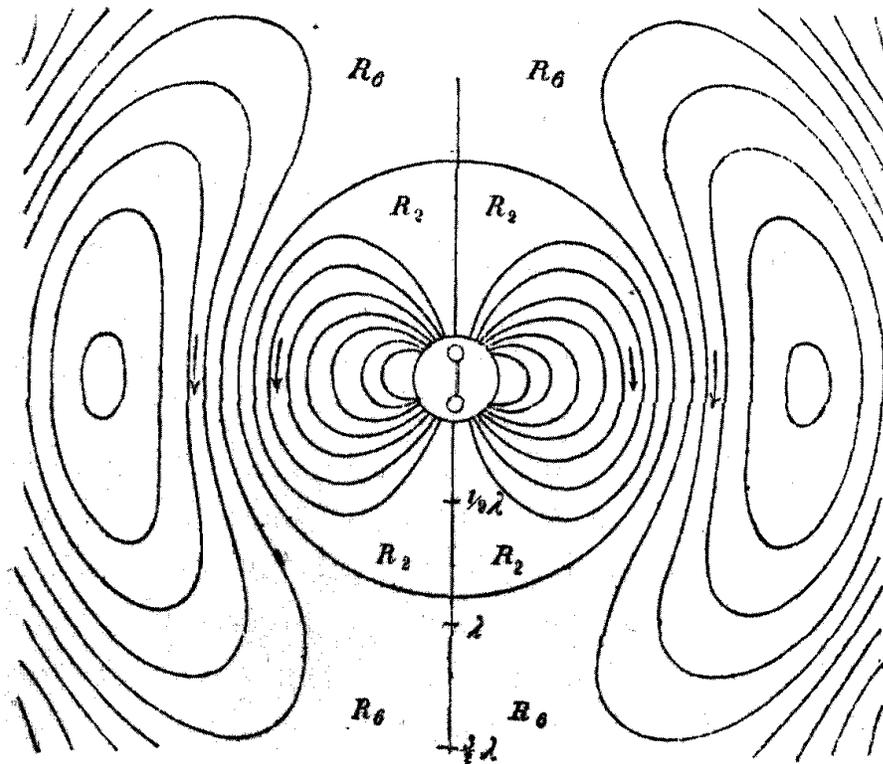


Figure 16.5: Heinrich Hertz' discovery of electromagnetic waves. The two circles at the center are metal balls which radiate waves when a spark leaps between them. The curves are isolines of the electric field as measured by Hertz by repeating the experiment over and over and moving his detector to different parts of the room. The arrows denote the polarities of the waves.

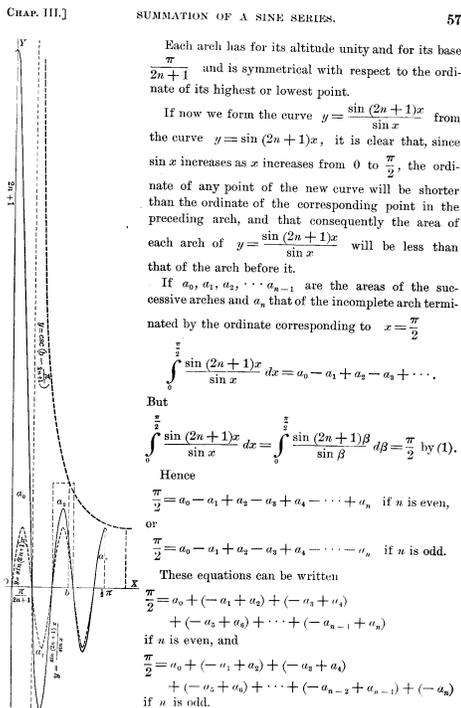


Figure 16.6: Left: an illustration of Gibbs' Phenomenon. Right: the text. Because the graph is very tall and narrow, the printer set it so as to fill the entire height of the page.

Scientific visualization has not been monotonic progress from the old pre-computer days. Some tasks that were easy for old-fashioned metal type are difficult with modern technology.

16.8 The Difference Between Two-Dimensional and One-Dimensional Waves

Fig. 16.7 is taken from a 1954 two-volume treatise of some 2000 pages by Philip Morse and Herman Feshbach, two physics professors who taught a legendary applied mathematics course at M. I. T., based on their text. Its goal is to illustrate the dimension-induced differences between the solutions of the ordinary, nondispersive linear wave equation

$$\nabla^2 u = u_{tt} \tag{16.1}$$

where ∇^2 is the Laplace operator and the subscript t denotes time differentiation. In one space dimension, the wave front remains very sharp at all times. In two space dimensions, however, there is a ripply wake behind the rounded leading edge of the spreading wave disturbance.

To translate this physics into pictures, Morse and Feshbach do several good things. First, the behavior of the wave equation in a given number of dimensions is illustrated by showing five snapshots of the solution. Each of the five graphs is consistently ordered (initial time at the top, time increasing downward) in a graph of identical format.

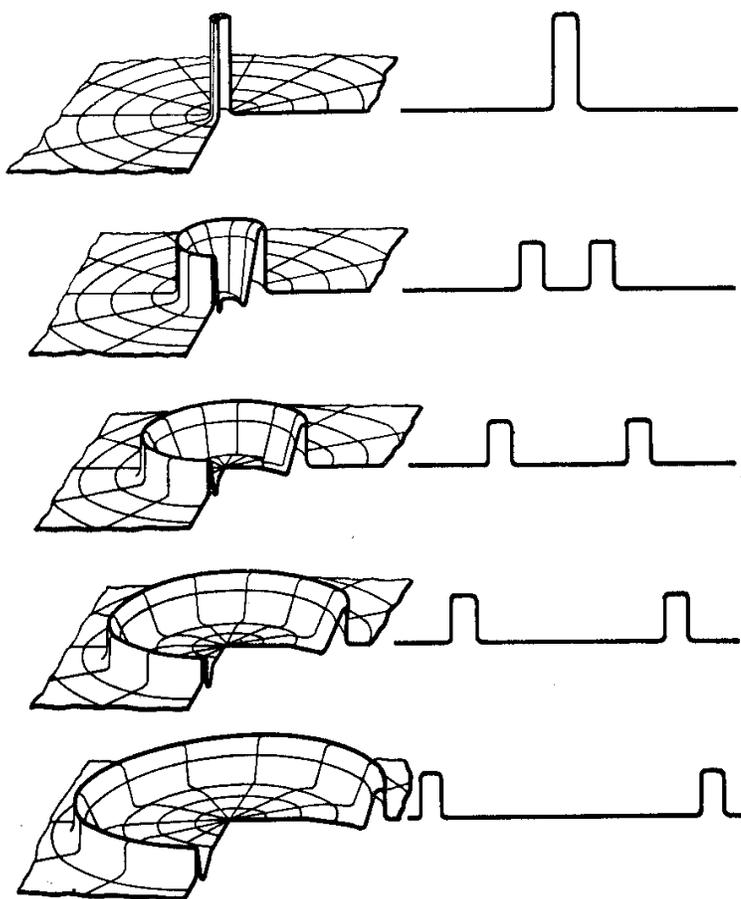


Figure 16.7: Left side: mesh plots illustrating the outward radial spreading of solutions to the two-dimensional wave equation. Right side: the same but in one space dimension where the leading and trailing edges remain sharp and there is no oscillatory wake.

Second, because the precise numerical values of the solution are irrelevant to the concept — for a linear equation, one can multiply the solution by an arbitrary constant without changing the shapes — the axes are not labelled.

Third, the graph type is different for different dimensions. An ordinary line graph is employed in one space dimension, but a surface mesh plot is used in two dimensions.

Lastly, because the wake is small and would be completely hidden in a standard mesh plot, Morse and Feshbach cut away one-quarter of the mesh.

Even so, the graph is not completely successful. The problem is that the wake is so small compared to the leading edge that it is almost invisible in the cut-away mesh plot. A graph can hardly be a triumph if its most important graphical feature is invisible.

This failure reminds us of another theme: good visualization is hard, even if you're as clever and careful as Morse and Feshbach.

16.9 The Importance of Color Schemes

Pseudocolor images are simple in concept: fill each little square of the image with a color denoting the value of the plotted quantity in that square. Unfortunately, creating a really useful pseudocolor image is tricky because the human eye has only a limited

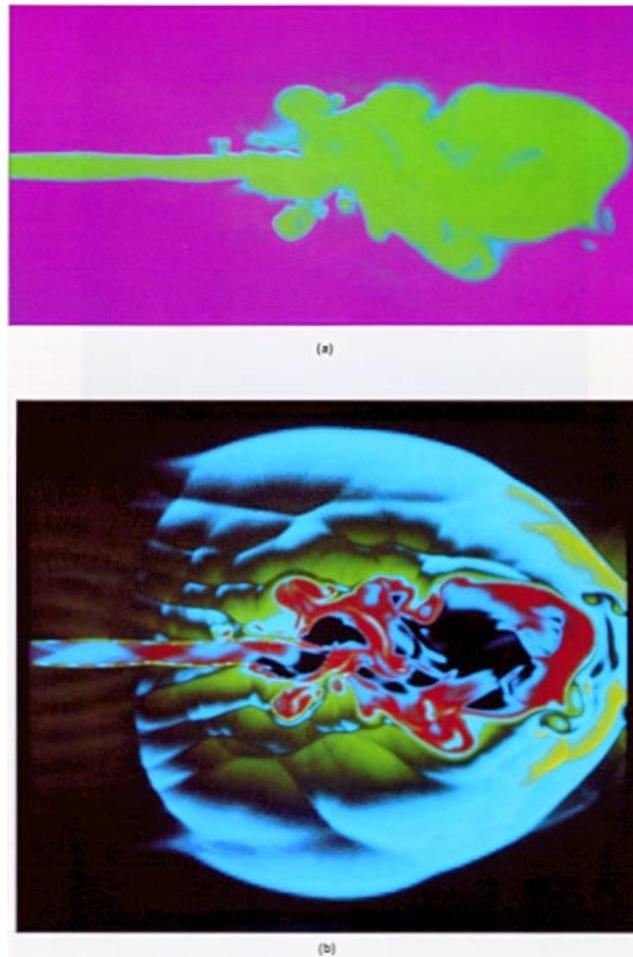


Figure 16.8: Two pseudocolor or “image” plots of the same flow, a turbulent jet streaming into quiescent fluid. The top figure employs a standard color mapping in which values are associated with color through a linear relationship. The visualization is poor because most of the small scale features are invisible. Donna Cox, an arts professor who has been associated with the supercomputer center at the University of Illinois (NCSA) for many years, devised a nonlinear color mapping which makes an amazing variety of additional features visible.

ability to distinguish fine shades of color. A little ingenuity in the color-coding scheme can greatly improve the result.

Fig. 16.8 shows two images of the same turbulent jet. The top image is very bland and boring; one can identify the core of the jet and the surrounding quiescent fluid, but there is little detail. The bottom image, which employs a complicated, nonlinear color-coding, brings out a whole host of small details which are missing from the image at the top.

Why are the two images so different? The answer is that most of the fine structure consists of variations in density of only a few percent of the mean density of the gas. In the color scheme at the top, these small variations translate into only slight, almost invisible variations of hue. The nonlinear color map has magnified these small variations so that these are easily visible.

One issue is: Are such small variations important? Only the goals of the study, not graphic design *per se*, can provide an answer. For this investigation, the small features are indeed important. The nonlinear color mapping shows that the jet is flanked by lots of intersecting shock waves. Although small at the time shown, these are serious challenges to the numerical algorithm used to compute this jet flow. Furthermore, these small-scale features may be the seeds of much larger features that will develop from them as time increases. Invisibility of these features, while still small, implies that the early stages of this evolution from little-to-big are veiled from the modeller.

Furthermore, the jet is turbulent. In the words of Lewis F. Richardson: "Big whirls have little whirls that feed and scratch and bite 'em, and little whirls have lesser whirls and so *ad infinitum*." This parody of a couple of lines by Jonathan Swift is an apt description of a turbulent cascade; a turbulence modeller has to care about the role of the very small features as much as the very large because they are intimately connected. Although large scale features do form, new eddies of very small scale are constantly being born in turbulence, so that the coupling of disparate length scales continues throughout the entire life of the hydrodynamic turbulence.

Fig. 16.9 is another example of the effectiveness of a nonlinear color map. The range of the plotted variable is split into ten pieces. Each of the ten parts of the interval was given its own color range, always ending in black. As a result, the pseudocolor image of the jet has ten contours of black plotted. In between, there are various shades of color. This makes it fairly easy to identify very fine features in this turbulent jet.

This pseudocolor image could really be called a "color-filled contour plot". However, the colors are not shaded, but are blended continuously within each sector. Thus, the green sector shades smoothly from black to a very light green. This is less garish than using blocks of solid color.

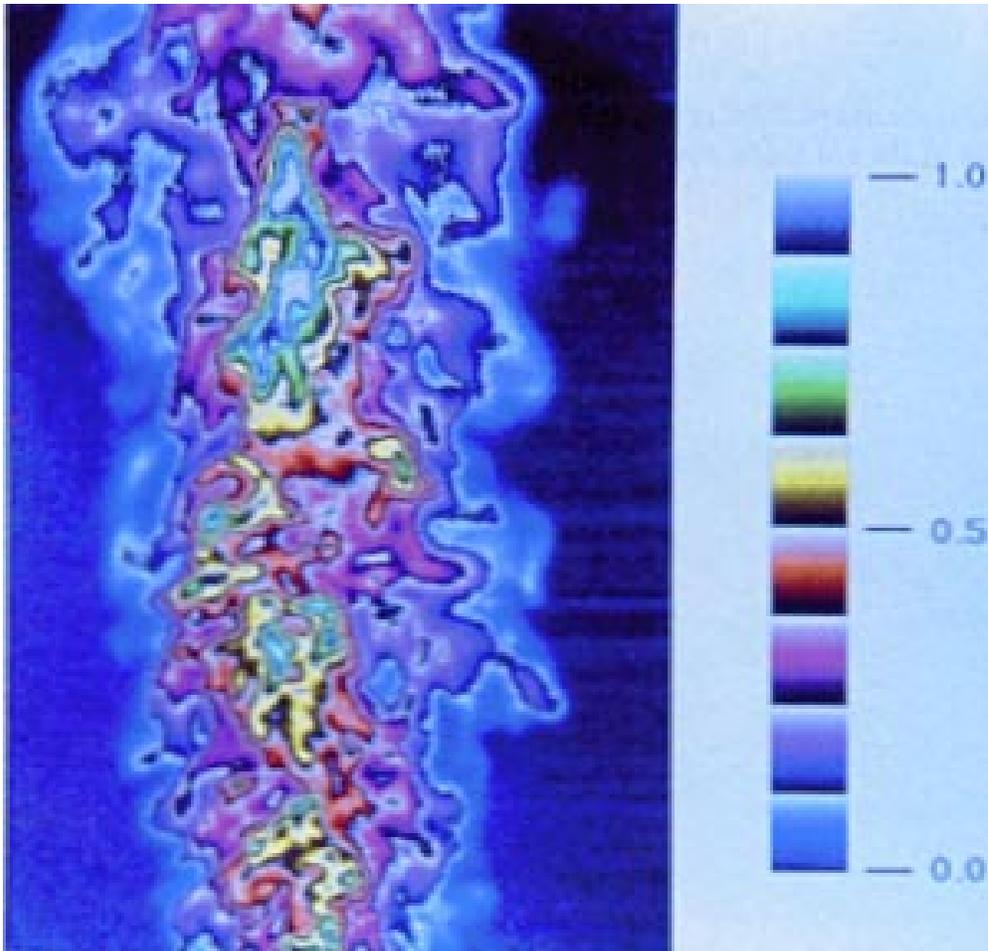


Figure 16.9: Pseudocolor image of a turbulent jet. The color-coding is shown in the colorbar. From an article by Long, Lyons and Lam, reprinted in Nielsen, Shriver and Rosenblum(1990)

16.10 Repetition: Turbulence in the Sargasso Sea

In Fig. 16.10, the top plot is of temperature variability, dubbed “activity”, in the Sargasso Sea. A region of “high activity” region was targeted for transects by the oceanographic research vessel. This region is delineated by the white contours that are superimposed on the lower three images. The point of the experiment is to understand the correlation between regions of high activity — probably patches of turbulence — and other physical variables. The white contours are very valuable to the reader because they make it much easier to identify the turbulent regions on the lower plots.

The artifice of repeating a previous result on a later graph is often used with line graphs; this set of panels shows that the same can be done with pseudocolor plots. The same idea could be applied to a black-and-white contour plot by using a thick line for the repeated contours, for example.

Matlab has a “hold on” command that allows the superposition of graphs, making it easy to combine a contour line from one figure with a contour plot of a completely different quantity. (One does have to use the Matlab “axis” command to be sure that both figures use the same horizontal and vertical scales; “axis off” will suppress overprinting of the axis labels.) The artifice of combining multiple plots in a single figure is so common that most graphics libraries, like Matlab, provide subroutines to make this superposition easy.

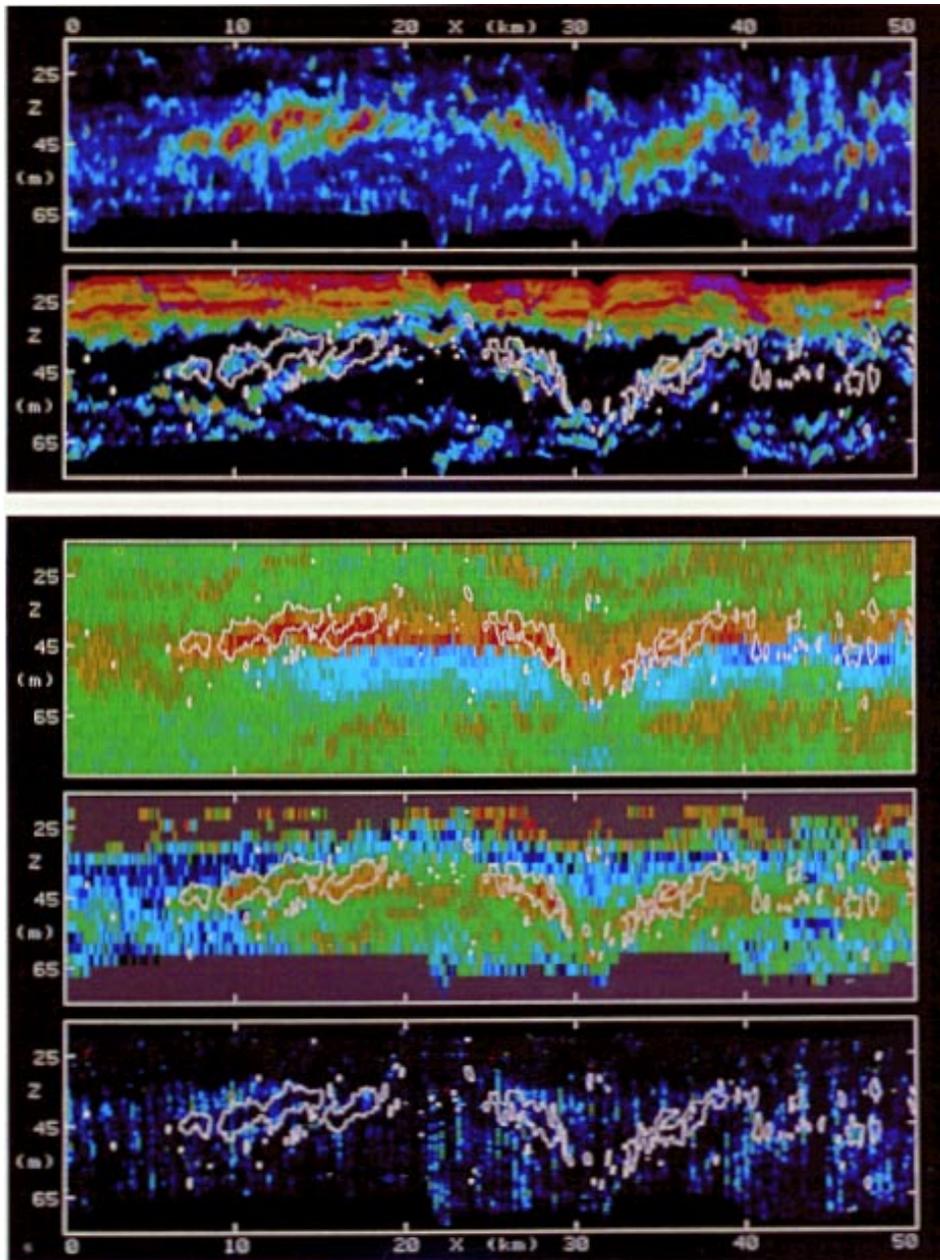


Figure 16.10: Oceanographic measurements in the Sargasso Sea as functions of depth and longitude, depicted by pseudocolor images. The panels show temperature variability, vertical temperature gradient, shear and Richardson number. The white contours which are superimposed on the lower three plots mark a region of high temperature variability (top image). Figure originally from an article by L. J. Rosenblum, reprinted in Nielson, Shriver & Rosenblum(1990), pg. 247

16.11 Philander (1990): Equatorial Currents

The equatorial current system is complicated because it involves flow in all three dimensions coupled with both surface (North Equatorial Current, etc.) and subsurface (Equatorial Undercurrent) features.

Fig. 16.11 is a very effective resolution of these difficulties. Currents are depicted in three different ways. First, the east-west surface horizontal currents are shown by thin, horizontal arrows with a smooth curve drawn through the tips of the arrows for greater clarity. The subsurface east-west currents are depicted by a combination of contour lines (“isotachs”, as lines of constant velocity are called) together with shading to distinguish eastward from westward flow. (Because the contour lines have to be crowded into a small space, shading is probably better than dashing negative contours for this application.)

The “meridional” circulation, which to a geophysicist means the vertical and latitudinal velocities (closely connected by the physics), is marked by heavy solid arrows. To avoid obscuring the isotachs of east-west flow, the meridional arrows are marked on a different slice through the ocean.

One enormous help to Philander is that velocities are *approximately* independent of longitude. Consequently, it is sufficient to show one cross-section at fixed longitude for the east-west flow and another for the meridional flow, and the reader can then apply those images to all cross-sections.

Well, *almost* all. Near the coasts, the east-west flow has to turn towards or away from the equator. A few additional heavy arrows schematically denote this turning near the coasts.

To orient the reader, the major continental land masses are delineated by thin curves on the surface and clearly labelled. To make it easier to distinguish land from water, thin vertical lines have been run downward along the sides of South America, but only for the land.

A complex schematic illustrates why Matlab is not enough. A contour plot can be made through a blind process: after computing a two-dimensional array of values of a function, one can issue a one-line command and see contours appear on the screen.

A schematic, however, has to be DESIGNED. The scientist has to make conscious decisions about what needs to be included and how to mix different graphical representations in a single diagram (contours, arrows-with-curve-at-tips, heavy solid arrows, shading, in this example) . Even once the design has been finalized, Matlab is inadequate. The schematic has to be DRAWN rather than computer-plotted.¹

¹Of course, with sufficient ingenuity one can combine multiple Matlab routines to duplicate Philander’s diagram. This merely replaces individual brushstrokes with individual subroutines, and is legitimately labelled “drawing” rather than plotting, albeit drawing by computer rather than by the pen of a draftsman.

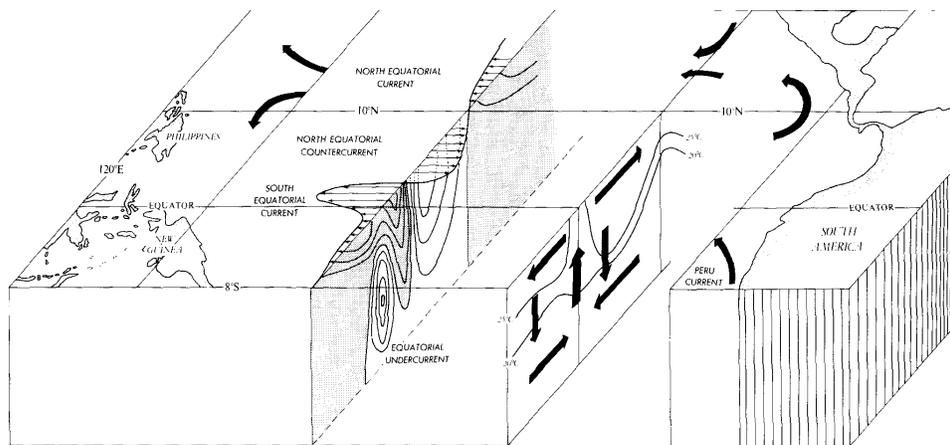


Figure 16.11: A schematic diagram of the circulation in the tropical Pacific Ocean.