Sten Andersson and the Language of Shape: An Interpretation

B. W. Ninham

Department of Applied Mathematics, Research School of Physical Sciences, Australian National University, GPO Box 4, Canberra ACT, Australia 2601


This essay attempts to analyse the later work of Anderson. It is argued that his seminal ideas on the role of hyperbolic geometries and of minimal surfaces in chemistry are in line with and a logical development in the tradition of Faraday and Maxwell.

Dedicated to Professor Sten Andersson on the occasion of his 60th birthday.

It must surely be the least read of books – A Treatise on Statics by George M. Minchin, Professor of Applied Mathematics at the Royal Indian Engineering College, Cooper’s Hill.¹ Not at all a bad book that taught generations of scientific engineers in India. It began with (Article 1): Definition of Force. Force is an action exerted upon a body in order to change its state of rest or of moving uniformly forward in a straight line. This is the definition of force given by Newton (Principia, Book 1, Definition IV). Students beginning physics had to take that definition as dogma, for the authority of Newton was hardly to be questioned. Then again, in Article 3 the good George gives a further definition of matter: Matter is something which exists in space and attests its presence by such observed qualities as extension, resistance and impenetrability. And by way of explanation follows the further remark: A limited portion of matter is called a body, and the quantity of matter contained in a body is called its mass. A very small particle of matter is called a particle.

These definitions, the beginnings of our attempts to extract some sense from the physical world, are tautological. For in such an axiomatic scheme, what on earth is space? Presumably that which is not occupied by matter. At the level of atoms the axioms are absurd. In a molecule where does one atom begin and another end? Langmuir considered water to be a giant molecule. And so it is. In our own times quantum mechanics and relativity represent higher tautologies that attempt to come to grips with these issues. Newton, preoccupied with himself and intent on celebrating the Creator through revealed (scientific) truth, was himself unhappy with his definitions. Indeed in a letter to his friend Bentley, he went so far as to say concerning the idea of action-at-a-distance attributed to his formulae that the notion represented “so great an absurdity that any man who had in philosophical matters any wit whatever could hardly fall into it”.

Any attempt to understand the art of Andersson begins with these things. Dissatisfaction with definitions drove Andersson to the chemistry of his time where the deeper questions could be put to one side. Even though chemists began to accept quantum mechanics which appeared to ‘explain’ the idea of a chemical bond, one could still in main agree with Comte, who in 1830 said that “Every attempt to employ mathematical methods in the study of chemical questions must be considered profoundly irrational and contrary to the spirit of chemistry. If mathematical analysis should ever hold its proper place in chemistry, an aberration that is happily almost impossible, it would occasion a rapid and widespread degeneration of that science”.

Or with Kant, who of the chemistry of his day declared that it was a science but not Science – “eine Wissenschaft aber nicht Wissenschaft”, for the criteria of true science lay in its relation to mathematics.

It is the search, bordering on obsession with the mathematics of minimal surfaces, with the role of curvature, that characterises Andersson’s later work. D’Aracy Thompson in his famous book on Growth and Form inspired some of it.

We can come closer to motivation if we consider one of the first papers of J. Clerk Maxwell, called on “Faraday’s Lines of Force”, given to the Cambridge Philosophical Society on December 11, 1855. As W. D. Niven, the editor of Maxwell’s scientific papers says: Maxwell had previously been attracted by Faraday’s method of expressing electrical laws, and he set himself the task of showing that the ideas which had guided Faraday’s researchers were not inconsistent with the mathematicians’ formulae into which Poisson and others had cast the laws of electricity. His object, he says, is to find a physical analogy which shall help the mind to grasp the results of previous investigations “without being committed to any theory founded in the physical sciences from which that conception is borrowed, so that it is neither drawn aside from the subject in the pursuit of analytical subtleties nor carried beyond the truth by a favourite hypothesis”.

Acta Chemica Scandinavica 45 (1991) 775–780
The laws of electricity are compared with the properties of an incompressible fluid the motion of which is retarded by a force proportional to velocity, and the fluid is supposed to have no inertia. He shows the analogy which the lines of flow of such a fluid will have with the lines of force, to deduce not merely the laws of static electricity in a single medium, but also a method of representing what takes place when the action passes from one dielectric into another. In the latter part of the paper, he proceeds to consider the phenomena of electromagnetism, and shows how the laws discovered by Ampère lead to conclusions identical with those of Faraday. In this paper three expressions are introduced which he identifies with the components of Faraday’s electronic state, though the author admits that he has not been able to frame a physical theory which would give a clear mental picture of the various connections expressed by the equations. In a later paper, on “Physical Lines of Force” he explained that having found the geometrical significance of the electronic state, he now proposed to examine magnetic phenomena from a mechanical point of view. The rest is history.

The search for the role of geometry, of minimal surfaces and of gaussian curvature had for Andersson something like the same roots or conviction that drove Maxwell. Except that no Faraday preceded him, so that the heresy was considered extravagant madness by the orthodox.

If dissatisfaction with the physicists’ ideas of force and energy led him to flee to chemistry, it was natural that acceptance of reality led to a long and distinguished apprenticeship in the orthodox church of inorganic chemistry. Crystallography at least avoided these issues. Atoms were fixed in space, and atoms could be treated as hard spheres, and chemical bonds existed. However vague that notion, spanning electrostatic to covalent bonding and derived from quantum-mechanical perturbation theory applied to a universe of two atoms, one kept the faith, and point-group crystal symmetry was all. The Andersson heresy began to take form with his work on zeolites, where the network of bonds appeared to lie on real minimal surfaces. The obsession resurfaced. Why on earth minimal surfaces? If quantum mechanics works, and it does describe spectra correctly, it loses something in so doing. The cruder Born–Oppenheimer approximation fixes nuclei in space and applies perturbation theory about those points. It does not do so well, but it does build in shape. We know that isomerism is critical in biology. There also emerged dissatisfaction with the definition of phase, e.g. in non-stoichiometric Ti₄O₉, usually described as a mixture of stoichiometric phases. But the full apparatus of solid-state crystallography, the language of dislocations, defects and grain boundaries, is not always adequate.

As developed by S. T. Hyde and Andersson (cf. other papers in this volume) the postulated isometric Bonnet transformation described certain transformations in liquid crystals, as well as the important Martensite transformation in steel. Fast-ion conductors became sensible when viewed from the point of view of minimal surfaces. A better description of things emerged following the parallel work of Von Schnering and Nesper, and colleagues on the same kinds of lines. The lengths of bonds and their curvature do change and are not immutable. Gaussian curvature and long-known obscure theorems in differential geometry loomed large. There was a resurgence of interest on the part of mathematicians in minimal surfaces.

If we like, these primitive ideas were at the level of Faraday’s work, perhaps no less important. From a philosophic viewpoint, crystallography, concerned with the arrangement of points in space and the symmetries derivable therefore, comes closest to the thoughts of the Greeks, at least to the Pythagorean belief in the magic of numbers. Music, astronomy, arithmetic and geometry were the original primaeval sciences. Even up to the Baroque period the building into composition of the Golden section was almost a mandatory requirement for proper harmony. It can be reasonably surmised that Andersson’s well known aversion to music derives from that fact. After the Renaissance the requirement was dropped. Ergo a lack of proper harmony!

It is a modern conceit that the search for final causes, a revealed truth embodied in a few mathematical laws (preferably number-theoretic), should have been abandoned following the proof of Gödel’s theorem. (In essence, that given any set of axioms and a logic to manipulate them, there are within the logic undecidable propositions.) But the search for final causes still underlies science, and it is indisputable that mathematics, a construct of the mind, has been incredibly effective. The marriage between arithmetic and geometry, the bridging of discrete and continuum still represents the most profound and enduring task of philosophy. It is that task that Andersson confronted. What is missing from crystallography is the connection from geometry to force. In the Euclidean desert of physics, forces act (in textbooks) between points, spheres, cylinders and planes. However successful quantum mechanics was numerically, somehow physics appeared to reduce the world to too sterile order, and, as the Born–Oppenheimer approximation shows, misses something. Nowhere in the scheme of things was the notion of shape. And, after all, the recognition and cognition of shape and form are the two earliest tasks presented to the human brain.

Shape and form imply the concept of a surface, the abstraction of discreteness and separateness which is at the core of things. The abstraction has it that an atom can be defined independently of its surroundings, even space.

In most of science that deals with molecular forces, atoms are treated as if point particles and size and shape are invoked as separate concepts. For example, the van der Waals dispersion energy of interaction of two atoms (proportional to \(1/r^6\)) becomes infinite on contact unless some preassigned hard-core contact size is involved. This convention is universal, and its origins lie deep. In fact no distinction can be made between energy, size and shape, which are inextricably linked. The linking concept of self-energy takes on significance whenever an object is con-
sidered to have a finite extent, or is delocalised. For then our abstraction that the object can be considered separately from its surroundings becomes philosophically tenuous, as one part of the object can consider its other parts to belong to the rest of the world. Hence the uncertainty principle. No difficulty occurs if the environment or object is immutable. If the opposite obtains, as indeed it always does at some level, the reaction of the (changed) environment to the object will be different, and the self-energy due to this reaction field will be different. The shift in self-energy due to radiation corrections to energy levels is a central problem of quantum electrodynamics. The Born electrostatic self-energy of an ion is important in electrolyte theory, physical adsorption and in the migration of ions through membranes. The concept is clear, but the details are complicated. We remark only that for a molecule of finite extent its dispersion self-energy (defined as the change in its energy due to its coupling with the electromagnetic field, or equivalently, as the change in quantum-mechanical zero-point energy of the field due to its coupling with the oscillating dipole moment it induces in the molecule) is the same as the binding energy (for a hydrogen atom) but of opposite sign.

The same ideas and formalism permit extension to two or more atoms: The interaction energy, the difference between the complete energy of the coupled system and the sum of the dispersion self-energies of two isolated atoms, reduces to the van der Waals energy at large distance, but now remains finite at zero separation. For like atoms this energy is the same as the binding energy of the molecule that would be formed by them. Again the same concepts, i.e. the recognition that an atom and its surroundings cannot be distinguished, can be used to develop a semiclassical estimate of the Lamb shift in hydrogen. More importantly, we can use the same formulae to explain the differences in binding energy (face-centred cubic versus hexagonal close-packed) of rare-gas crystals. This is not accessible by a description based on the idealisation of an atom represented by a hard or even soft core with two- or three-body interaction potentials. These ideas and formalism to deal with them were developed by Mahanty and the author and summarised in Ref. 2. The observation is important. The entire electromagnetic reaction field must be taken into account to explain these energies. The resolution of this long-standing problem provides an immediate clue: Global (many-body forces) and local properties are linked, and global structure is necessary.

With that in mind, let us now step back further to see what motivated the concept of minimal surfaces. Consider with Nesper and Von Schnering an array of electrostatic point charges arranged in different crystallographic symmetries of, say, NaCl, CsCl etc. We suppose that the particular space-group symmetry is set by packing, i.e. for the moment admit the approximation of a hard-sphere model for the ions to set the lattice parameters. The Ewald sum of the electrostatic energies provides the major contribution to the binding energy. Now consider surfaces of zero electrostatic potential traced out inside the lattice. These can be calculated, and look very much like minimal surfaces. The idea here is that ionic or electronic motion proceeds along the tangential fields of these minimal surfaces. If the crystal zero potentials are truly minimal surfaces, then phase changes can occur easily via the Bonnet transformation. Excited states of the crystal, including anharmonic states, are then included and counted as the (infinity) of different allowed minimal surfaces supposed to form a complete set.

But these surfaces are not minimal, although appearing close to such. The reasons seem clear enough: No electrostatic set of charges can be in an equilibrium state. Presumably if one built in quantum-mechanical zero-point energies and then calculated the new equipotential surfaces, the new zero potential surfaces would be minimal surfaces of the field.

With that idea, we can turn the argument around and say that since the crystal exists as an equilibrium system, it must be permeated by surfaces of zero stress of the entire electromagnetic field. Imposition of this (mechanical) requirement, together with the condition that the Poynting vector \( \mathbf{E} \times \mathbf{H} \) (momentum transport) is zero, presumably will give back quantum mechanics. The argument implies the emergence of quantum mechanics as a consequence of minimal surfaces, a necessary Pythagorean imperative that effects the bridge between geometry and arithmetic, discrete and continuum, the particle and field points of view.

There is much more to say on the impact of Andersson’s thinking which paralleled that of his colleague Larsson on cubic phases, and others on chaotic bicontinuous phases now seen to be ubiquitous in membrane biology and protein chemistry. The shift in thought from a Euclidean desert to a world of hyperbolic geometries, with its linkage of global and local properties, has spurred what is coming to be known as Nouvelle Physique. It will have no less profound an impact, perhaps more, than Hawking’s synthesis in astronomy.

Beyond that we can note the recent developments of Andersson on quasi-crystals, geometric in origin, but motivated again by the drive toward that same bridge between arithmetic (number) and geometry. Advances in number theory, now coming back into fashion with the publishing of Ramanujan’s notebooks, can be expected to lead to new insights. New number-theoretic formulae will be found that elucidate packings of spheres. These are related to theta functions and their transformations, equivalent to Fourier’s theorem through the more fundamental Riemann relation between \( \zeta(s) \) and \( \zeta(1-s) \), but different. Perhaps these new results bear on the whole question of the wave–particle duality and on quasi-crystal symmetry, e.g. eqns. (1) and (2).

\[
\sum_{j=1}^{\infty} \ln \left[ 1 \pm \exp \left( -i\mathbf{x} + y \right) \right] = \sum_{m=1}^{\infty} |\mu(n)| \ln \left[ 1 \pm \exp \left( -n^2 i\mathbf{x} + y \right) \right]
\]

(1)

\[
\sum_{j,m=1}^{\infty} \frac{m x^m}{1-x^m} = \sum_{m=1}^{\infty} |\mu(n)| \sum_{j,m=1}^{\infty} \frac{mn j^2 x^{m j^2}}{1-x^{m j^2}}
\]

(2)
where \( \mu(m) \) is the Möbius function \( \mu(m) = (-1)^r \) if \( m \) is a product of \( r \) distinct primes, zero otherwise, and the coefficient of \( x^n \) in eqn. (2) is \( o(N) \), the sum of the divisors of \( N \). Another we can derive is eqn. (3). This final equation is

\[
\sum_{n=1}^{\infty} \mu(n) \cos ny = \pi \sum_{n=1}^{\infty} \frac{\mu(n)}{2\pi} \left( \sum_{m=0}^{\infty} \delta \left( y - \frac{2m\pi}{n} \right) - \frac{1}{2} \right) \tag{3}
\]

surprising. The structure defined by the convoluted Möbius function is geometrically random (although algebraically well determined). Yet the diffraction pattern of such the structure consists purely of Bragg peaks. This Fourier transform is a sum of delta functions that requires two indices, not one, unlike the fundamental formula of analysis and crystallography, eqn. (4). It is worth noting

\[
\sum_{n=1}^{\infty} \cos ny = \sum_{m=-\infty}^{\infty} \delta \left( y - \frac{\pi}{2} - 2m \right) \tag{4}
\]

that such functions and new theta-function analogues also satisfy Poisson's equation with delta-function sources.

The above paragraphs are an attempt to extract some of the essence of Anderson's thought, which has played, and will continue to play, a key role in what will come to be seen as a great watershed in science. It is a development of which D'Arcy Thompson, who made such eloquent plea that forces should be linked to form, would have liked. It is the beginning of a language of shape that we suspect Pythagoras, Faraday, Maxwell and Leibniz certainly, and maybe even Newton, would have thoroughly approved.

**Appendix**

To make a connection between such apparently obscure new formulæ in number theory, note that Euler's and similar products, the theta functions, elliptic modular functions and Ramanujan's function \( \tau(n) \), formulæ like the Rogers–Ramanujan identities, are intimately related in ways not yet clear to packings of spheres. Some hints at these things can be gleaned from Ref. 3, now long out of print, and in Ref. 4.

The theta functions are fundamental. If we take the conventional definitions, with \( q = e^{2\pi i \tau} \), we have eqns. (A1) and (A2). Their limiting forms as \( \tau \to 0 \) are periodic

\[
\theta_3(z, q) = \sum_{n=-\infty}^{\infty} (-1)^n q^{nz^2} e^{2\pi i n}, \quad \theta_3(z, q) = \theta_3(z + 4\pi i, q) = 1 + 2 \sum_{n=1}^{\infty} q^{nz^2} \cos nz \tag{A1}
\]

\[
\theta_4(z, q) = \theta_4(z + \pi i/2, \tau) = 2 \sum_{n=0}^{\infty} q^{(n+1/2)^2} \cos [(2n+1)z] \tag{A2}
\]

delta functions. Thus we have eqns. (A3)–(A6).

\[
\theta_3(2\pi z, 1) = 2 \sum_{n=0}^{\infty} \cos [(2n+1)2\pi z] = \frac{1}{2} \left( 1 + \theta(z - m) - \frac{1}{2} \delta(z - m) \right) \tag{A3}
\]

\[
\theta_4(2\pi z, 1) = 2 \sum_{n=0}^{\infty} (-1)^n \sin [(2n+1)2\pi z] = \frac{1}{2} \left( 1 - \theta(z - m) - \frac{1}{2} \delta(z - m) \right) \tag{A4}
\]

\[
\theta_3(2\pi z, 1) = \sum_{m=-\infty}^{\infty} e^{2\pi i m} = \frac{1}{2} \sum_{m=-\infty}^{\infty} \theta(z - m) \tag{A5}
\]

\[
\theta_4(2\pi z, 1) = \frac{1}{2} \sum_{m=-\infty}^{\infty} \delta[z - (m - 1/4)] \tag{A6}
\]

These periodic delta functions are the basic source functions for any electrostatic array of charges with the standard crystallographic symmetries. They satisfy the fundamental differential equation (A7), where \( y \) is any of

\[
\frac{\partial^2 y}{4 \delta z^2} + \frac{\partial y}{\delta t} = 0 \tag{A7}
\]

the four theta functions. The electrostatic potential associated with any standard lattice, the Ewald sums, can be written as an integral over products of the four theta functions in appropriate combinations to allow overall charge neutrality. Thus for the CsCl structure, which mimics the infinitely periodic P minimal surface, the electrostatic potential \( V \) is given by the proportionality (A8), where now

\[
W(x, y, z) \propto \int_{\tau=0}^{\infty} t \, dt \, \left( \theta_3 \theta_4 \right) \tag{A8}
\]

the arguments of the theta functions are \( \pi \), \( \pi y \) and \( \pi z \), and \( q \equiv \exp (-t^2) \). Similarly for the NaCl structure we have eqn. (A9), for which, e.g., from eqn. (A7), we get eqn. (A10).

\[
V(x, y, z) = \frac{16}{\pi} \int_{\tau=0}^{\infty} t \, dt \, \theta_3(2\pi x, \tau) \theta_4(2\pi y, \tau) \theta_3(2\pi z, \tau) \tag{A9}
\]

\[
\nabla^2 V = -32\pi \theta_3(2\pi x, 1) \theta_4(2\pi y, 1) \theta_3(2\pi z, 1) = \left\{ \sum_{m=-\infty}^{\infty} \delta[x - m] - \frac{1}{2} \delta[x - m] \right\} \left\{ y \right\} (z) \tag{A10}
\]
which reduces to the charge distribution required. Again a structure of charges with diamond symmetry has the potential of eqn. (A11), with arguments \(2\pi x, \exp(-r^2)\), and so

\[
V(x,y,z) = 16 \int_0^\infty r \, dr \left( \delta_2 \delta_1 \delta_2 + \delta_1 \delta_1 \delta_1 \right) \tag{A11}
\]

on for the IWP, Gyroid and other symmetries.

Potentials for all other Bravais lattices can be written by varying the lattice spacing in the \(x, y\) and \(z\) directions. These are exhaustive.

As already discussed, the infinitely periodic zero-potential surfaces so generated are not minimal surfaces, but close to such surfaces.\(^3\) The mean curvature \(H\) has the simple form \(H = -\nabla \cdot \vec{E}/2\), where \(\vec{E}\) is the unit electric field vector, and the gaussian curvature is a more complicated function related to the stress tensor of the field.

One presumes that the full time-dependent electromagnetic field equations for electrostatic and vector potentials coupled by the gauge condition will also have similar expressions.

A fascinating conjecture of Hyde has it that, because of the conditional convergence of the potential sums, rearrangement of the sums may lead to an infinite set of topologies associated with any given array. This conjecture is open.

The Jacobi theta function transformation is equivalent to the fundamental Riemann relation, eqn. (A12), where \(\Gamma(s) = 2^{s-1} \Gamma(s) \zeta(s) \cos \frac{1}{2}\pi s = \pi^{s} \zeta(1-s) \tag{A12}\)

is the gamma function and \(\zeta(s)\) is given by eqn. (A13). It is connected to the prime-number theorem through the Riemann conjecture that the only non-trivial zeroes of \(\zeta(s)\) lie on the line \(\text{Re}(s) = \frac{1}{2}\).

\[
\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \quad \text{Re}(s) > 1 \tag{A13}
\]

A limiting case of the Jacobi transformation is the fundamental formulae of crystallography, e.g. eqn. (A14). The

\[
\sum_{n=-\infty}^{\infty} e^{2\pi imz} = \sum_{m=-\infty}^{\infty} \delta(z-m) \tag{A14}
\]

infinite product forms of the theta functions, e.g. eqn. (A15), suggest that one may be able to construct a

\[
\theta_3(z,q) = \prod_{n=1}^{\infty} \left( 1 - q^{2n} \right) \prod_{n=1}^{\infty} \left( 1 + 2q^{2n-1} \cos 2z + q^{4n-2} \right) \tag{A15}
\]

wider class of basis functions than those which underly present crystallography, to include, say, quasicrystals. This must be possible, as disordered systems exist. Details will emerge in later publications.

For the moment notice that eqn. (3) of the main text is one such. Another suggestive result is eqn. (A16),

\[
\theta_3(z,q) = \prod_{j,m=1}^{\infty} \left( 1 - |\mu(n)| e^{-2\pi q^2 j} \right) \times \left( 1 + |\mu(n)| e^{-2\pi q^2 e^{iy}} \right) \tag{A16}
\]

where \(\mu(n)\) is the Möbius function. Properties of prime numbers loom large in any such new formulae. A trivial example is this extension of Euler’s product [eqn. (A17)],

\[
\sum_{n=1}^{\infty} \mu(n)/(1-e^{-\eta n}) = \sum_{n=1}^{\infty} \ln (1 - |\mu(n)| e^{-\eta n}) = \frac{1}{2\pi i} \int_c \frac{\Gamma(p)}{y^p} \left( \sum_{n=1}^{\infty} \frac{1}{n^{p+1}} \right) \left( \sum_{m=1}^{\infty} \frac{|\mu(m)|}{m^p} \right) \tag{A17}
\]

where the contour \(c\) in the complex \(p\)-plane of the inverse Mellin transform is sited to allow interchange of orders of integration and summation.

Carrying out the sum one has eqn. (A18), since eqn. (A19) holds. Then since also eqn. (A20) holds, one has eqn. (A21); replacing \(y\) by \(\bar{y}\) and summing over \(j\) one has eqn. (A22), whence eqn. (2) of the main text follows by differentiation.

\[
\sum_{n=1}^{\infty} \frac{|\mu(n)|}{n^p} = \sum_{n=1}^{\infty} \frac{\mu(n)}{n^p} = \frac{1}{\zeta(p)} \tag{A18}
\]

\[
\sum_{n=1}^{\infty} \frac{\mu(n)}{n^p} = \frac{1}{\zeta(p)} \tag{A19}
\]

\[
\sum_{n=1}^{\infty} \ln (1 - |\mu(n)| e^{-\eta n}) = \ln \prod_{n=1}^{\infty} (1 - e^{-\eta n}) \tag{A20}
\]

These considerations, some firm, some conjecture, are in the spirit of Anderson’s work itself.

One thing alone is certain. Anderson’s influence, drawing on an interaction based on rock-solid long-term careful experiment mixed with a breathtaking capacity to speculate, has opened up an extraordinary new field.

There appears much to do.
References


Received November 28, 1990.