Full Topology of the Laplacian of the Electron Density. IGR report : GR/N04423 P.Popelier, Dept. of Chemistry, UMIST, Aug 2002

1. Background – context

The current work contributes to bridging on aspect of the gap between the information contained in modern *ab initio* wave functions and chemical insight. Bonding and non-bonding electron pairs can be extracted, characterised and visualised by the so-called quantum topological approach, the main theme of this work. This approach has its roots in the theory of "Atoms in Molecules" (AIM), pioneered by the Bader group.

In its original form AIM has two main strands: the topology of the molecular electron density and the topology of its Laplacian. The former characterises chemical bonds and atoms, and the latter provides a novel route to study electron pair localisation. Since 1994 another function, called the Electron Localisation Function(ELF) became popular to study electron pairing. ELF analyses adopted the central AIM idea of using the gradient vector field to partition space into topological basins. More recently other localisation functions such as LOL^2 and the Lennard-Jones function³ were introduced. The surge in activity and popularity⁴ in the area of "quantum chemical topology"(QCT) calls for generalised software that enables – in *general* scalar 3D functions – the localisation of critical points and the integration of property densities over topological basins.

The work we carried out during the current grant delivered the goal of new and general QCT computer code (built into a local version of the world-wide licensed program MORPHY) and substantially increased our knowledge of the full topology of the Laplacian. Whereas before a portion of its critical points(CPs) were studied we are now able to provide the full picture, including all CPs and their connectivity and the shape, volume and population of the Laplacian's basins. For the first time it becomes possible to scrutinise the link between the widely taught Valence Shell Electron Pair Repulsion(VSEPR) model (pioneered by Gillespie and Nyholm) and the Laplacian, a parsimonious 3D function directly derived from modern *ab initio* wave functions.

2. Key Advances

[1] "The full topology of the Laplacian of the electron density. 2: The Umbrella Inversion of

Ammonia", N.O.J. Malcolm and P.L.A. Popelier, J.Phys. Chem.A., 105, 7638-7645 (2001).

The full topology of $L(\mathbf{r})$, which is defined as minus the Laplacian of the electron density, $\nabla^2 \rho$, has recently been explored for the water molecule⁵. In this work we have investigated the changing topology during the "umbrella" inversion in ammonia. The maxima in $L(\mathbf{r})$ are points of local charge concentration, which can be associated with the electron pairs of the VSEPR model. We examine changes in three valence shell charge concentration (VSCC) and three depletion (VSCD) graphs as a function of the angle between the C₃ axis and a hydrogen nucleus. Through the use of planar graphs the transition mechanisms can be easily rationalised. *The previously noted double maxima in* $L(\mathbf{r})$, *corresponding to the lone-pair of nitrogen, found at the transition state for inversion is shown to persist for geometries distorted considerably from planar*. The transitions between structures in the valence shell charge concentration and charge depletion graphs do not occur simultaneously.



Fig.1 The Valence Shell Charge Concentration graph for the transition state in $NH_3(\alpha=90^0)$, associated with VSCC I.

There are eight maxima, (3,-3) (green); nine saddle points, (3,-1) (purple); and three minima, (3,+1)gradient (yellow). The paths connecting the (3, -3) and (3, -1)are critical points shown as superpositions of white spheres. The crossed critical points merge as α is increased beyond 106° , leading to the VSCC II graph, which does not contain the lower non-bonding maximum.

For the first time dynamic changes in the topology of $L(\mathbf{r})$ are reported in detail. The nuclear skeleton of the test case ammonia is controlled by an angle α , which preserves the threefold symmetry. This parameter describes the well-known NH₃ inversion, and distorts the geometry over a considerable range from the planar transition state to a geometry far beyond the equilibrium geometry. More than fifty critical points in $L(\mathbf{r})$ are discerned but based on the network of connecting gradient paths they can be subdivided into graphs. These networks extend over the whole molecule and can be represented by planar graphs. Three VSCC graphs and three VCSD graphs are discerned, each associated with a range of stability of at least $\sim 15^{\circ}$ in α . The transitions between the VSCC graphs do not occur at the same α values as the transition between the VSCD graphs. We have elucidated how lone pairs are presented by $L(\mathbf{r})$ under inversion, thereby providing a valuable complement to studies based on ELF. Below we show some of the typical figures that were obtained in the course of this subproject.

Figure 1 is an example of connectivity diagram of a subset of critical points, as it can be inspected in 3D via typical on-line rendering software. Figure 2 is a summary of all planar graphs found, while Fig.3 represents a transition mechanism corresponding to a change in topological pattern. Figure 4 is the ultimate summary of what happens in ammonia as the molecule is being distorted.

Fig.2 Planar graph representations of the three VSCC graphs (ac) and the three VSCD graphs (d-f) occurring in the inversion of ammonia. The solid circles (vertices) in the VSCC graphs represent the (3,-3) CPs, and the edges the (3,-1) CPs. The open circles (vertices) in the VSCD graphs represent the (3,+3) CPs, and the edges the (3,+1) CPs. Table 1 shows the total number of vertices, edges and faces of each graph. Note that the rest of the plane, outside enclosed faces, also represents a face. The edges have been drawn such that they do not cross, as required for

planar graphs.



Fig. 3 Schematic representation of the mechanism of transition from VSCC I to VSCC II. Filled triangles represent (3,-1) critical points, filled circles (3,-3) and open triangles (3,+1). The triplets in brackets indicate the signs of the eigenvalues.



900	1060	134 ^o		
	VSCC 1	SCC II VSCC III		
	VSCD 1 VSCD	II YSCD III		
90 ⁰	108 ⁰	1280		
	α	-		

р

Fig. 4 Survey of the appearance of the three Valence Shell Charge Concentrations graphs (VSCC I, VSCC II, VSCC III) and the three Valence Shell Charge Depletion graphs (VSCD I, VSCD II, VSCD III) in terms of the angle intervals of the control parameter α .

[2] "Protonation Energies and the Laplacian of the Electron Density, a Critical Examination." P.L.A. Popelier and P.J. Smith, *Phys.Chem.Chem.Phys.*, **3**, 4208-4212 (2001).

The success and limitations of the Laplacian of the electron density in predicting protonation energies is systematically discussed for a selection of molecules, including aldehydes, amines, amides and diglycine. We address four questions related to (a) the preference between competing protonation sites, (b) correlations between the Laplacian and the protonation energy in substituted homologous series and (c) upon conformational change and (d) the Laplacian's uniqueness in using information stored in the unprotonated molecule. This study forms a solid basis to investigate the use of the Laplacian in the rationalisation of mass spectrometric fragmentation patterns of peptides.

With an eye on peptide mass spectrometry we have investigated to what extent the Laplacian of the electron density can be used as an indicator of protonation energy. Although not unique in this capacity, the Laplacian is an example of a function that is able to predict relative trends in protonation energy from the unprotonated molecule only. An essential condition for the success of the Laplacian is that the protonation trends are established in a series where the local environment of the protonation site is not too much perturbed. In particular, alkylation of amines (and phosphines) and non-extreme conformational changes around nitrogen reveal a nearly linear correlation between the Laplacian *cannot* be used over *different* atomic numbers to predict the preferred site of protonation, although it correctly predicts oxygen in amides, and is largely able to point out the correct protonation site in substituted aldehydes. This study has discussed the opportunity and limitation of the Laplacian in the prediction of relative protonation energies, opening the avenue for an informed investigation of its potential impact in the rationalisation of mass spectrometric fragmentation patterns of peptides.

[3] "A topological study of homonuclear multiple bonds between the elements of Group 14"

N.O.J. Malcolm, R.J.Gillespie and P.L.A.Popelier, Dalton Trans., 3333-3341 (2002).

We joined the topical debate on the character of multiple bonding between heavy Group 14 elements. A heated discussion involving Profs Cotton and Schaefer arose in connection with the recent synthesis of "gallynes"(Gp13) and extended later towards Group 14 elements. In the series of REER and R₂EER₂ molecules

(E = C, Si, Ge, and Sn) only the C molecules have classical triple and double bonds. In all the other molecules it becomes increasingly difficult to describe the electrons as either bonding or nonbonding.

The ELF analysis shows that if the Si, Ge and Sn molecules are forced into the higher energy linear and planar conformations then some electron density is forced into the bonding region, increasing the "bond order" and decreasing the bond length correspondingly. However when this geometry relaxes to give the equilibrium geometry electron density moves from the bonding region to the nonbonding, even though these regions cannot be clearly defined. Classical Lewis structures are therefore inadequate for describing the bonding in these molecules. Our main conclusions can be summarised as follows:

- 1. The EE bonds in the REER molecules (R = Si, Ge or Sn) have a bond order less than three but greater than one and the EE bonds in the $R_2 EER_2$ molecules a bond order less than two but greater than one. These bond orders cannot be more precisely determined.
- 2. The electrons that are not attracted into the bonding region are present as nonbonding electrons and they cause these molecules to be non-linear and non-planar respectively.
- 3. The reason for the difference between the carbon molecules and those of the heavier elements is that the electrons are less localised and are less attracted into the bonding region in the molecules of the heavier elements because of the larger size and smaller electronegativities of the atoms of these elements.



Fig.5 The η =0.7 ELF isosurfaces for the REER systems at two levels of theory. Colour legend : Core basin (magenta), valence protonated(blue), valence disynaptic(light green and dark green) and valence monosynaptic(orange and yellow). (a) Si₂H₂ at 6-31G* level (b) Si₂Me₂ at 6-31G* level (c) Ge₂H₂ at cc-pVDZ level and (d) Ge₂Me₂ at cc-pVDZ level. All calculations performed using B3LYP.

[4] "A new Algorithm to Integrate property densities over Quantum Topological Basins",

N.O.J. Malcolm and P.L.A.Popelier, to be submitted to *J. of Comput. Chem.*

The program MORPHY98⁶ has been designed integrate properties over topological basins in the electron density. Integration of the basins in fields such as the Laplacian (or $L(\mathbf{r})$ more precisely) is far more demanding than the atomic basins in ?(\mathbf{r}) because of the more complex nature of the separatrices, which form the boundaries between basins. Here we extend this capability to integration over basins in any 3D function. Although the developed algorithm is more general we focus on $L(\mathbf{r})$ in terms of immediate applications. A great deal of effort was concentrated on obtaining an analytical fit of the separatrix surfaces. Several months of work was directed towards eliminating the "gap problem"⁷, which is due to high ellipticities at the (3,-1) critical point(CP). Clever sampling of so-called initial points to trace gradient paths constituting the separatrices, an event that occurs frequently in $L(\mathbf{r})$. Experiments with Padé approximants replacing the hitherto used Chebyshev polynomials and spline fits instead of Fourier expansions proved equally unfruitful.



Fig. 6 Flow diagram showing the organisation of the modules for basin assignment in MORPHY.

This led us to a drastically different algorithm borrowed from computer graphics¹ called the "octree" algorithm. In this algorithm the usual problems associated of finding a functional form for the separatrices are avoided because no explicit form is ever generated for these boundaries. The current algorithm also benefits from the ability to restart calculations and adaptively update the quadrature. Essentially the object at hand, i.e. the topological basin is described recursively by a set of increasingly smaller cubes that focus on the boundary of the object. A flow diagram with the subroutines as implemented in a pilot version of MORPHY are shown in Fig.6

[5] "Functional Groups expressed as graphs extracted from the Laplacian of the Electron

Density", P.L.A.Popelier, J.Burke and N.O.J. Malcolm, accepted by Intern.J.of Quant.Chem.

The valence charge concentration shell, as determined by the Laplacian is used as a source of quantum topological graphs, called L-graphs. These graphs are extracted from the *ab initio* wave functions of 31

molecules calculated at B3LYP/6-311+G(2d,p)//B3LYP/6-311+G(2d,p) level, covering common functional groups in organic chemistry. L-graphs can be constructed from a largely transferable subgraph, called atomic L-graph. We investigate the topological stability of the L-graphs as a function of the basis set. Reliable and consistent atomic L-graphs are only obtained with basis sets of triple zeta quality or higher. The recurrence of invariant motifs or subgraphs in the L-graphs enables the isolation of sixteen atomic L-graphs.

We have surveyed graphs extracted from the Laplacian of a sizeable number of common organic molecules covering a variety of functional groups. We distinguished sixteen so-called atomic L-graphs, highly transferable building blocks from which all molecules except dioxirane could be constructed. These graphs are classified according to element, coordination and the functional group of which they are part. Atomic L-graphs are juxtaposed by superimposing their terminal (3,-1) critical points, except for two atomic L-graphs that occur in triple bonds, and the construction of the dioxirane L-graph. Here a terminal maximum or (3,-3) critical point is superimposed instead. Topological stability upon basis set variation is reached with very few exceptions when using a triple zeta or quadruple zeta basis set. The use of a double zeta basis set is to be discouraged in the generation of L-graphs. Our systematic classification of L-graphs offers an avenue in the physicalisation of chemical graph theory, possibly the mathematical model of constitutional chemistry proposed by Ugi *et al.*⁸.

In Fig.7 we show the collection of atomic L-graphs recovered from the set of molecules containing the most important functional groups. This library could be used to assemble for example the L-graphs in Fig.8.



Ü Fig.7 Atomic L-graphs. (a) mono-coordinated (i.e. singly bonded) hydrogen, (b) bi-coordinated carbon (in ketene), (c) bi-coordinated carbon (in triple bond), (d) tri-coordinated carbon (in ketene), (e) tri-coordinated carbon (in keto), (f) tricoordinated carbon (in O-C=O), (g) tetracarbon, (h) mono-coordinated coordinated nitrogen, (i) linear bi-coordinated nitrogen, (j) non-linear bi-coordinated nitrogen, (k) "(near) planar" tri-coordinated nitrogen ("VSCC I"), (1) "non-planar" tri-coordinated nitrogen ("VSCC II"), (m) mono-coordinated oxygen, (n) bi-coordinated oxygen (alcohol, carboxyl, enol, ester, ether), (o) bi-coordinated oxygen (oxime), (p) bi-coordinated oxygen (oxirane).



Fig.8 Two examples of molecular L-graphs.

[6] "An improved Algorithm to Locate Critical Points in a 3D scalar Field",

N.O.J. Malcolm and P.L.A.Popelier, J.of Comput. Chem., in press, (2002).

A new algorithm for location of the critical points(CPs) in general scalar fields is described. The new method has been developed as part of an on-going process to exploit the topological analysis of general 3D scalar fields. Part of this process involves the use of topological information to seed the critical point search algorithm. The continuing move away from topological studies of just the electron density requires more general algorithms and the ability to easily 'plug-in' new functions. Another important aspect of the current algorithm is the retention of all possible intermediate information, for example the paths describing the connectivity of critical points, as well as an ability to restart searches, something that becomes increasingly important when analysing larger systems. This new algorithm represents a core part of a new local version of the MORPHY code.

For the first time we present an exhaustive and clear classification of all possible gradient paths (GPs) that topologically connect two rank 3 CPs in an arbitrary scalar 3D field (see Table 1). This insight is then used to improve a CP search algorithm. A new search algorithm for CPs has been included in MORPHY. The source code is more compact yet more powerful than the original code of MORPHY01. New features are the straightforward "plug-in" of new scalar fields, the restart and refinement capability enabling successive and exhaustive searches. Finally the new algorithm exploits topological connectivity of the CPs.

Table 1 Survey of the nine possible types^a of gradient path in any 3D scalar field.

		Dimensionality ^b			
GP type ^c	$(r,s)_{\text{origin}} \rightarrow (r,s)_{\text{terminus}}$	Origin	Terminus	Manifold ^d	Example
[3,2]	$(3,+3) \rightarrow (3,+1)$	3	1	1	Ring line
[3,1]	$(3,+3) \rightarrow (3,-1)$	3	2	2	Interatomic Surface
[3,0]	$(3,+3) \rightarrow (3,-3)$	3	3	3	Atomic basin
[2,2]	$(3,+1) \rightarrow (3,+1)$	2	1	1	
[2,1]	$(3,+1) \rightarrow (3,-1)$	2	2	2	
[2,0]	$(3,+1) \rightarrow (3,-3)$	2	3	2	Ring surface
[1,2]	$(3,-1)\rightarrow(3,+1)$	1	1	1	
[1,1]	$(3,-1) \rightarrow (3,-1)$	1	2	1	GP in ρ in conflict structure
[1,0]	$(3,-1) \rightarrow (3,-3)$	1	3	1	Bond Path

^a Only critical points of rank 3 are considered.

^b For a CP at the origin this is the number of directions in which the CP can "send" GPs, while for a CP at the terminus this is the number of directions in which the CP can "receive" GPs. For the manifold this is dimension of the connecting topological object, i.e. curve, surface or basin, for 1,2,3 dimensions respectively.

^c Each gradient path is characterised by two critical points, an origin and a terminus. The critical points are denoted by their indices, I_{origin} and I_{terminus} respectively and the gradient path as [I_{origin} , I_{terminus}].

^d This is the set of gradient paths that connect two given critical points.

[7] "In search of a physical basis for the VSEPR model in terms of the Laplacian of the electron density", N.O.J. Malcolm and P.L.A.Popelier, to be submitted to *Faraday Discuss.*, invited contribution to Discussion 124 "Quantum Inorganic Chemistry", York, April 2003.

In its original formulation the VSEPR model was built on the concept that valence shell electron pairs behave as if they repel each other and thus keep as far apart as possible. However, in recent years more emphasis has been placed on the space occupied by a valence shell electron pair, called the domain of the electron pair⁹. Bader, Gillespie and McDougall were the first to propose¹⁰ L(**r**) as a physical basis for the VSEPR model. Their proposition was founded on the observation¹¹ that local maxima of L(**r**) faithfully duplicate in number, location, and size the spatially localised electron pairs of the VSEPR model. This work surveyed a selected set of molecules, which we revisit in this article, optimised at B3LYP/6-311+G(2d,p) level. The compounds contain three maxima(SO₂), four maxima(CH₄, SiH₄, NH₃, PH₃, OH₂, SH₂, NF₃, PF₃), five maxima(ClF₃, SF₄, SF₄O), and six maxima(ClF₅). The analysis of Bader *et al.*¹¹ was purely based on the location of so-called critical points in L(**r**) (in this case maxima). Although very valuable this restricted information makes it hard to gauge the size (i.e. volume) of a domain. In parallel with the shift of emphasis in the VSEPR model towards an understanding of the space occupied by a valence shell electron pair we take advantage of the newly developed OCTREE algorithm to visualise L(**r**)-basins and compute their volumes and electronic populations.

We present a visual representation of the topological partitioning of a set of molecules (example Fig.9) in terms of basins of $L(\mathbf{r})$. The electron population of oxygen's core basins (n=1), 1(O), is 0.76, that of the hydrogen core basin, 1(H), is 0.57. The populations of the valence (n=2) basins are 1.83 for the non-bonding ("lone-pair") oxygen basin, 2n(O) and 0.52 for the bonding basin 2b(OH). The region outside this basin extends to infinity and can be viewed as a global basin. Indeed its attractor is global since it is a "sea" of point attractors, at infinity. A large population is associated with this basin, namely 3.19. This phenomenon also occurs in other systems such as CH₄, SiH₄, NH₃ and PH₃.The other puzzling issue is why the basin populations are lower than the corresponding ones in ELF. The results obtained so far will be completely written up and then discussed with Prof. Gillespie, Prof.Bader and Prof.Silvi, as well as during the Faraday Discussion.



Fig.9 Basins of $L(\mathbf{r})$ in water, a full view (left) and an intersection "slab view" (right). Core basins (n=1) are marked in red, non-bonding valence basins (n=2) in green (bright and dark) and bonding valence basins (n=2) in yellow.

Based on the calculations performed on the full set of molecules we conclude that nonbonding or lone pairs have larger domains than bonding pairs in the same valence shell, in accordance with VSEPR. However, bonding pairs in the valence shell of a central atom increases in size with increasing electronegativity of the ligand, contrary to the view of VSEPR. Similarly, we observe that bonding pair domains decrease in size with increasing electronegativity of the central atom, again contrary to a postulate of VSEPR. Finally we confirm that double and triple bond domains are larger than singlebond domains.

Project Summary : Full Topology of the Laplacian of the Electron Density.

Our project featured in the context of the increasingly popular (250 papers per annum and growing⁴) quantum chemical topology (QCT) approach. We embarked on an ambitious project involving the writing of new software with an eye on enhancing our understanding of the most parsimonious and oldest function used to localise electron pairs in a topological manner, namely the Laplacian of ρ , or more precisely L(**r**). Although the impact of ELF is steadily increasing most papers in the literature have so far concentrated on L(**r**). This is why it was important to push the boundaries of our current knowledge of the properties of L(**r**).

The reached milestones are summarised as follows:

- □ We studied dynamic changes in $L(\mathbf{r})$'s topology. A systematic analysis of the umbrella inversion of NH₃ showed three possible VSCC and VSCD graphs and their regime of stability. The transition mechanisms were described in detail. It is now understood how $L(\mathbf{r})$'s topology keeps track of the lone pair in ammonia.
- □ Our ELF study (new objective) on multiple bonding between Group 14 elements showed that EE bonds in REER molecules (R=Si,Ge and Sn) have a bond order less than 3 but greater than 1, and that EE bonds in R₂EER₂ molecules have a bond order less than 2 but greater than 1. Hence we cannot support terms used in the literature such as disilylynes, digermenes or distannenes.
- □ We now have a general, robust and flexible algorithm, called OCTREE (incorporated in a local version of MORPHY). It enables the facile implementation of the integration of any property density over any topological basin.
- \square We have investigated the strengths and weaknesses of L(**r**) in connection with the prediction of local protonation energies. This careful study opens a safe and well-informed avenue to the application of L(**r**) in the rationalisation of mass spectrometric fragmentation patterns of peptides.
- □ For the first time we have obtained an exhaustive and lucid classification of all possible types of gradient paths. This insight is used to improve an algorithm for the localisation of critical points in a general 3D scalar 3D function. Robust and general computer source code was added to MORPHY.
- □ The previous algorithm was extensively applied to extract so-called *L-graphs* from the most common functional groups of organic chemistry. A systematic study on 31 small but representative molecules led to a set of sixteen highly transferable atomic L-graphs, from which the L-graphs of larger molecules can be constructed. Topological stability was exhaustively investigated against basis set variation, leading to the conclusion that triple zeta basis sets are a minimum requirement.
- □ The subsidiary postulates of VSEPR were scrutinised beyond the mere location of critical points, the state of the art before our work. Now it is possible to visualise the topological basins in $L(\mathbf{r})$. This representation is superior to the typical ELF contour plots shown in the literature because for the first time one can now perceive the explicit boundaries of the basins, i.e. the separatrix surfaces. Two postulates of the VSEPR model (size of non-bonding versus bonding domains and size of multiple bond domains) are confirmed but no support can be found for the postulate on electronegativity.

Overall we have achieved the ambitious goal of incorporating a new software tool in MORPHY. Its source code needs to be fine-tuned and prepared for commercial release, but already now it has delivered important applications, which are in various stages of publication. We have furthered the boundaries of QCT and made progress in extracting rigorous chemical insight from modern *ab initio* wavefunctions, one of the main aspirations of QCT. Finally the PRDA spontaneously confided at the end of his grant that he found this project the most stimulating and potentially far-reaching he had come across in his career.

- 1. Foley, J. D.;van Dam A.;Feiner S. K.;Hughes J. F.;Philips R. L., *Introduction to Computer Graphics*, Addison-Wesley, USA, 1990.
- 2. Schmider, H. L.; Becke A. D. J.of Mol.Struct. (Theochem) 2000, 527, 51.
- 3. Gillespie, R. J.;Bayles D.;Platts J.;Heard G. L.;Bader R. F. W. J.Phys.Chem.A 1998, 102, 3407.
- 4. Popelier, P. L. A.; Smith P. J., in "Chemical Modelling : Applications and Theory", Volume 2 Royal Society of Chemistry Specialist Periodical Report, Chapter 8, pp. 391-448., A. Hinchliffe, Ed., 2002.
- 5. Popelier, P. L. A. Coord.Chem.Rev. 2000, 197, 169.
- 6. MORPHY98 a program written by P.L.A. Popelier with a contribution from R.G.A. Bone, UMIST (1998).
- 7. Popelier, P. L. A. Comp. Phys. Commun. 1998, 108, 180.
- 8. Ugi I; Bauer J; Brandt J.; Friedrich J.; Gasteiger J; Jochum C.; Schubert W. Angew. Chem., 1979, 18,111.
- 9. Gillespie, R. J.; Robinson E. A. Angew. Chem. Int. Ed. Engl. 1996, 35, 495.
- 10. Bader, R. F. W.; Gillespie R. J.; MacDougall P. J. J.Am.Chem.Soc. 1988, 110, 7329.
- 11. Bader, R. F. W.; MacDougall P. J.; Lau C. D. H. J.Am.Chem.Soc. 1984, 106, 1594.