Superpositional Quantum Network Topologies

Christopher Altman,[1,2] Jarosław Pykacz,[3,4] and Romàn R. Zapatrin[5,6]

We introduce superposition-based quantum networks composed of (i) the classical perceptron model of multilayered, feedforward neural networks and (ii) the algebraic model of evolving reticular quantum structures as described in quantum gravity. The main feature of this model is moving from particular neural topologies to a quantum metastructure which embodies many differing topological patterns. Using quantum parallelism, training is possible on superpositions of different network topologies. As a result, not only classical transition functions, but also topology becomes a subject of training. The main feature of our model is that particular neural networks, with different topologies, are *quantum states*. We consider high-dimensional dissipative quantum structures as candidates for implementation of the model.

KEY WORDS: Neural networks; quantum topology.

1. INTRODUCTION

Quantum learning networks have been suggested to offer new domains for quantum algorithm design (Behrman *et al.*, 1999; Chrisley, 1993; Ventura and Martinez, 1997). Machine-learning–inspired architectures are self-organizing, robust, and ideal for such tasks as pattern recognition and associative processing. Contemporary quantum network models convey their advantage using superposed quantum states on a fixed topological background. We suggest superposed quantum *topologies* as a novel approach to quantum neural networks, and present a model of learning and evolving superposed quantum network topologies, or SQNTs. The mathematical basis of this model is predicated upon two existing formalisms:

[1] Quantum Information Science and Technology Project, ATIP, Tokyo, Japan.
[2] Universiteit van Amsterdam, The Netherlands.
[3] Instytut Matematyki, Uniwersytet Gdański, Wita Stwosza, Gdańsk, Poland.
[4] Center Leo Apostel of the Vrije Universiteit Brussels (VUB), Krijgskundestraat, Brussel.
[5] Friedmann Lab. for Theoretical Physics, SPb UEF, Griboyedova, St. Petersburg, Russia.
[6] To whom correspondence should be addressed at Friedmann Lab. for Theoretical Physics, SPb UEF, Griboyedova 30–32, 191023, St. Petersburg, Russia; e-mail: zapatrin@rusmuseum.ru.

- Neural networks: the classical learning model of pattern recognition (Dorogov, 2001; Rumelhart *et al.*, 1986).
- Algebraic quantum foam model: the formalism which describes superposed and continuously evolving discrete structures (Raptis and Zapatrin, 2001; Zapatrin, 1998).

In order to show how our model can be implemented we consider multidimensional quantum systems admitting highly degenerate states as described in (Savvidy, 2000) and the Bose–Einstein condensate as outlined in Vitiello (1995).

   An overview of the model is arranged as follows. We begin with a classical perceptron model, namely, of a multilayered feedforward, weakly connected neural network. Then we extend this model, making it quantum by admitting the existence of superpositions of differing topological structures of neural networks using Rota algebraic formalism (Zapatrin, 1998) as applied to describe the evolution of reticular patterns of quantum space-times. This means that

1. we admit that the topological structure of a perceiving entity can change, and
2. we admit that these changes may be continuous.

At first sight these two requirements look contradictory: how can discrete structures evolve continuously? These requirements are reconciled in quantum mechanics. Along these lines we present the model of superposition-based quantum network topologies.

2. INTERLUDE ON RETICULAR QUANTUM SPACE-TIME FORMALISM

   Before coming to SQNT models in more detail and for the sake of self-consistency, we recall the necessary constructions from quantum mechanics used to describe reticular quantum space-time. The main feature of the quantum mechanical description of a physical system is that we pass from configuration space to a complex linear space, called the STATE SPACE of the system. We shall consider systems with finite configuration spaces, therefore their state spaces will be finite-dimensional Hilbert spaces $\mathcal{H} = \mathbf{C}^N$. Suppose we have a system whose configuration space is $\{1, \ldots, n\}$. According to quantum mechanics, it can be in pure (that is, dispersion free) superposed state such as

$$|\psi\rangle = \cos \alpha |1\rangle + e^{i\phi} \sin \alpha |2\rangle \tag{1}$$

Quantum observables, that is, measuring apparata are described by self-adjoint operators in the state space of the system in question. The values of the observables

are the eigenvalues of the operators; they are always real. Note that mutually commuting (thus having the same eigenvectors) observables are associated with the same measuring apparatus—they are interconverted just by relabelling the pointer's values. To specify an observable K, we consider a decomposition of the unit operator by orthogonal projectors and associate a *number* with each projector. In Dirac notation this reads:

$$K = \sum_i k_i |i\rangle\langle i| \tag{2}$$

2.1. Spatialization

The *spatialization procedure* was developed in Raptis and Zapatrin (2001) for the purpose of describing space-time foam. Its primary feature is to associate discrete structures, rather than numerical values (2), with subspaces of the state space. More specifically, in standard quantum mechanics the state space is defined as a Hilbert space, that is, a complex linear space with an inner product, $\mathbf{H} \times \mathbf{H} \to \mathbf{C}$. Each subspace of $\mathbf{H}$ can be associated only with its dimension as an integer. We associate with each subspace $\mathcal{C}$ a disjoint graph, whose number of vertices equals its dimension, $\dim \mathcal{C}$.

The spatialization procedure links some of the vertices, thus associating the subspace $\mathcal{C}$ with richer structure than a cardinal number. The only requirement is for $\mathcal{H}$ to be endowed with an associative product structure, rather than an inner product. In some cases this structure already exists, for instance, when $\mathcal{H}$ is a tensor product of two copies of the same state space $\mathbf{H}$, that is $\mathcal{H} = \mathbf{H} \otimes \mathbf{H}$. The associative product is defined on factorable vectors as

$$\phi \otimes \phi' \cdot \psi \otimes \psi' = \langle \phi', \psi \rangle \cdot \phi \otimes \psi'$$

and then extended by linearity (this is nothing but a usual matrix product). The spatialization procedure is briefly outlined as follows: with any directed acyclic transitive graph G with N vertices its ROTA ALGEBRA is associated $\mathcal{A}$, whose elements are $N \times N$ matrices of the following form. They have zero entries a_{ik} when the vertices i, k are not connected in the graph G. For instance, consider the graphs

$$\tag{3}$$
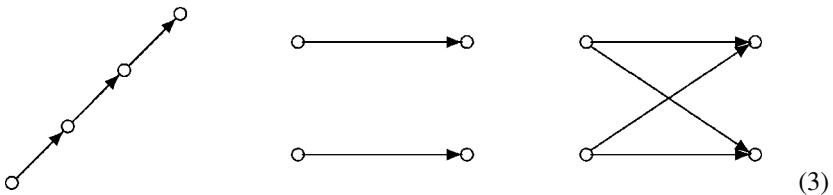
For these examples, the appropriate Rota algebras take the following form

$$
\begin{pmatrix} * & * & * & * \\ 0 & * & * & * \\ 0 & 0 & * & * \\ 0 & 0 & 0 & * \end{pmatrix}
\begin{pmatrix} * & 0 & * & 0 \\ 0 & * & 0 & * \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix}
\begin{pmatrix} * & 0 & * & * \\ 0 & * & * & * \\ 0 & 0 & * & 0 \\ 0 & 0 & 0 & * \end{pmatrix}
\tag{4}
$$

For a detailed account of Rota algebras and the spatialization formalism supplied with examples the reader is referred to Appendices A, B and Raptis and Zapatrin (2001) and Zapatrin (1998).

2.2. Neural Metastructures

The main feature of the proposal put forward in this paper is to shift from a particular structure of neural network as perceptron to something different: a metasystem whose *states* are perceptrons. This is accomplished as follows. Superposed quantum neural networks are themselves linear spaces rather than neural networks. With each subspace of the state space of SQNT—which is its quantum property, see (2) above—a particular neural network configuration is associated using the above mentioned spatialization procedure. In Section 3 we suggest to use the quantum features of SQNTs to enhance the performance of the main task in neural networks, namely, their training.

Realistic physical candidates to implement SQNTs have already been proposed. The first is based upon the Bose–Einstein condensate, on which a quantum brain model has been suggested by G. Vitiello (Vitiello, 1995). A second model has been suggested by Savvidy (2000). Both have the crucial property needed for our purposes, namely they have many individual degrees of freedom and admit superpositions. We emphasize that no ab initio association of states with graphs is needed, as graphs are automatically produced as a consequence of the spatialization procedure. SQNT implementation can also be modelled classically as demonstrated in Werbos (2002), a broad class of quantum algorithms can be simulated on classical systems.

3. NEURAL NETWORKS AND ROTA ALGEBRAS

One of the basic tasks of neural networks is to function as perceptrons, that is to recognize signals for which we have no structural theory, for instance, to recognize visual patterns. In this section, we review the basic principles on which perceptrons are based, their learning and training in classical setting and show how Rota algebras introduced above emerge in their description. We shall deal

with multilayered feedforward NNs, such as
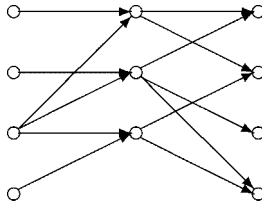
that is, their nodes can be arranged in layers so that (i) no nodes in a given layer communicate and (ii) the signals propagate only consecutively via layers. The state of a given neural network is the set of values, numerical or vectorial, assigned to its nodes. A time-step propagation changes the perceptron's state so that the value of a node i at time $t + 1$ depends on its own value and the values of nodes adjacent to i at time t.

Besides their multilayered, high-dimensional structure the filtering and recognizing power of neural networks is based on the nonlinearity of their transition functions. In some cases, in particular while dealing with stability or elasticity issues, we may use linear approximations. This is exactly the case we explore in this paper: optimal topological configurations are assumed to be robust under small permutations (Dorogov, 2001).

3.1. Training

Initially one starts with a set of patterns for which the classification is known. Usually by means of heuristic methods, the topological structure of the network is chosen and then trained via input of known patterns and subsequent adjustment of network parameters (transition functions). Output signals are correlated with patterns from different classes to be well-separated with respect to certain criterion. The most popular method to adjust transition functions is error backpropagation (see, for instance, Rumelhart *et al.*, 1986).

3.2. Performance

Signal propagation in the linear approximation can be viewed as a matrix multiplication, which reduces to a number of arithmetic operations. The more links there are between neurons, the more computational resources are consumed by the process of pattern recognition. In order for a neural network to be faster, we should seek for sparser configurations.

Therefore, the criterion for "good matching" should also take performance into consideration. As an example we consider the approach of *fast neural networks*, which are weakly connected multilayered feedforward neural networks

(Dorogov, 2001). The idea of these networks grows from the FFT—fast Fourier transform. When finding optimal configurations of such networks, we restrict the allowed links between neurons and force networks to be sparse. Note that in our setting this requirement is followed automatically as we are confined by the dimensionality of the state space $\mathcal{H}$.

4. SUPERPOSITION-BASED TRAINING

The basic proposal of this paper is to replace training of particular neural network configurations with training of superposed ones. This can be achieved in the following steps:

(i) SQNT is treated as a system described by quantum mechanical formalism, in our special case represented by a matrix algebra $\mathcal{A}$. Note that the term "quantum" here is a mere indication of the rules of behavior of the object on which the SQNT is implemented, whatever be its "real" nature.

(ii) Input signals are vectors from a representation space of the algebra $\mathcal{A}$, as are output signals. Connection weights are represented by the matrix elements, and can further be represented by matrices of linear operators consequent of the spatialization procedure.

(iii) Learning is a matching algorithm which selects a subset $\mathcal{C}$ of $\mathcal{A}$ for which the criterion of matching between input and output signals is optimal. At this step we obtain the optimal configuration on the metalevel.

(iv) In order to pass to a particular neural network structure, we consider a subalgebra (rather than a subset) of $\mathcal{A}$ which is the nearest to the subset $\mathcal{C}$.

Now let us dwell on the above issues. For the first step, an appropriate physical system for SQNT to be implemented should be found. The main requirement which it has to satisfy is to possess sufficiently many degrees of freedom with controlled access to them. The initial candidates for these purposes are entangled quantum registers on which quantum computers are based. An alternative model can be considered using the Bose–Einstein condensate (Vitiello, 1995).

So, we start with a quantum system S with sufficiently many degrees of freedom. To implement item (ii), we must draw the distinction between the classes of states treated as states of the input register and between the states responsible for the operations of multiplication. This step does not affect the overall performance of the future perceptron, as it has to be done only once (this is analogous to building a computer). The work of the SQNT is to evolve according to a prescribed evolution, starting from the initial state (=a pattern to recognize) to the final one (=its classification identifier).

An algorithm seeking for best matching (iii) can be realized as follows. We feed in a pattern from a given sample set (=given prescribed state) and then let it evolve. The evolution is set in such a way that it realizes the multiplication

operation. To be more precise, we consider any available evolution U_t and then label the states of SQNT in such a way that for the unit time δt the resulting action of $U_{\delta t}$ would be multiplication of appropriate elements. This is possible due to the arbitrary nature of labeling of eigenstates of operators (see Section 1).

The condition for learning to be successful is that the input vector from a sample set should be unchanged. In physical terms, this means patterns from the sample set should be eigenvectors for the energy operator responsible for the evolution. Why do we require it? The point is that we would like our output signals to be well separated, that is, they should be stable with respect to small variations of the input vector. This, in turn, requires the states of the overall register (*input*, *output*) to be orthogonal. If they are eigenstates of the generator of the evolution, this holds automatically. We denote by $\mathcal{C}$ the appropriate eigenspace and call it Calculation Subspace.

The last item (iv) is a kind of "classical correspondence principle" in quantum mechanics. SQNT operates as it is—namely, it undergoes quantum evolution— but if we would like to represent (perhaps with some losses) its work in terms of neural networks, we proceed in the following way. We deal with the calculation subspace $\mathcal{C}$ of our state space, which we obtained as a result of the best matching algorithm, and we have the evolution operator in our disposal which is interpreted as multiplication.

The spatialization procedure can be applicable to $\mathcal{C}$ only if it is closed with respect to the multiplication, while this requirement may not hold for eigenspaces of the energy operator. That is why, following the lines of quantum mechanics, we consider the nearest *subalgebra* to $\mathcal{C}$ and immediately interpret it as a neural network. This suggestion is in full accordance with quantum measurement theory, where the wave function of the system immediately "collapses" to an eigenstate of the appropriate operator.

5. CONCLUDING REMARKS

Presently only a handful of quantum algorithms exist, and these are confined to a limited set of specialized applications. Quantum learning architectures offer the potential to expand this domain to a much broader class of functionality. We have outlined the fundamentals of a quantum fast-training pattern recognition model—*superpositional quantum network topologies*, or SQNT—which provides a rich source for the development of a novel class of quantum algorithms. The distinguishing feature of our model lies in its ability to utilize coherent superposition of unique topological configurations of neural networks.

We suggest two ways to implement our model in physical media based on high-dimensional dissipative quantum systems (Savvidy, 2000; Vitiello, 1995). Quantum simulation methods offer an immediate candidate for study of the model: nonclassical dynamics can be efficiently simulated on ensembles of states of

classical Turing machines ruled by second-order differential equations, see Werbos (2002) for more detail. It should be mentioned that although macroscopic physical systems are composed of quantum components, quantum phenomena in the macroscopic realm are usually invisible due to averaging over a large number of degrees of freedom. Nevertheless, the existence of macroscopic systems exhibiting quantum properties cannot be neglected in many subdomains of the life sciences, especially in molecular biology (Aerts *et al.*, 1993). For example, A. S. Davydov in his book "Biology and Quantum Mechanics" (Davydov 1982) studied collective soliton excitations in large protein molecules and applied them to such biological phenomena as membrane transport, nerve impulse conduction, and muscle contraction. Extending beyond relevance to the current model, we suggest this could be a potential explanation of the surprisingly high adaptation capabilities of living organisms. Classical models of learning and response to continuously changing environments require immense computational resources. We interpret this as a signal that perhaps quantum models could be more efficient for this purpose.

APPENDIX A: ROTA ALGEBRAS OF TRANSITIVE GRAPHS

Rota Algebras in Dirac Notation

Let X be a reflexive transitive graph. For brevity a pair of nodes i, j of X connected by an arrow $i \rightarrow j$ is said to be Tending. Consider the linear space Ω whose basis $|i\rangle\langle j|$ is labelled by tending pairs $i \rightarrow j$ of nodes of X.

$$\Omega(X) = \left\{ \sum_{i, j \in X} |i\rangle\langle j| \quad \text{such that} \quad i \rightarrow j \right\} \tag{5}$$

In the sequel, when no confusion occurs, we omit the notation of the graph X in parentheses and simply write

$$\Omega = \Omega(X)$$

Define the product on Ω by setting it on its basic elements:

$$|i\rangle\langle j||k\rangle\langle l| = \begin{cases} |i\rangle\langle l|, & \text{if} \quad j = k \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Note that $|i\rangle\langle l|$ in (6) is always well-defined since the graph X is assumed to be transitive, that is why the existence of darts $i \rightarrow j$ and $j \rightarrow k$ always enables the existence of $i \rightarrow k$. The space Ω with the product (6) is called the Rota Algebra of the topological space $(X, \rightarrow)$. These algebras were first introduced in Rota (1968) in the context of combinatorial theory.

The Matrix Representation of Rota Algebras

Given the Rota algebra of a transitive graph X, its standard matrix representation is obtained by choosing the basis of Ω consisting of the elements of the form $|i\rangle\langle k| = \mathbf{e}_{ik}$, with ik ranging over all tending pairs $i \to k$ of nodes of X. The matrices $\mathbf{e}_{ik}$ (called matrix units) are defined as follows:

$$\mathbf{e}_{ik}(m, n) = \begin{cases} 1, & m = i \quad \text{and} \quad n = k \text{ (provided } i \to k) \\ 0, & \text{otherwise} \end{cases} \tag{7}$$

where $\mathbf{e}_{ik}(m, n)$ stands for the element in the mth row and the nth column of the matrix $\mathbf{e}_{ik}$. We can also extend the ranging to *all* pairs of elements of X by putting $\mathbf{e}_{ik} \equiv 0$ for $i \to k$. Then the product (6) reads:

$$\mathbf{e}_{ik}e_{i'k'} = \delta_{ki'}\mathbf{e}_{ik'} \tag{8}$$

To specify a Rota algebra in the standard matrix representation we fix the template matrix replacing the unit entries in the incidence matrix I_{ik} of the graph X:

$$I_{ik} = \begin{cases} 1, & i \to k \\ 0, & \text{otherwise} \end{cases}$$

by wildcards $*$ ranging independently over all numbers. For instance, the algebra associated with the two-node graph $\cdot \to \cdot$ has the the following template matrix

$$\Omega(\cdot \to \cdot) = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} = \left\{ \begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \mid a, b, c \in \mathbf{C} \right\}$$

So, we see that any transitive graph can be described in terms of a finite-dimensional algebra (for further details we refer to Zapatrin, 1998).

APPENDIX B: SPATIALIZATION PROCEDURE

Here we describe the spatialization procedure which associates a graph with an arbitrary finite-dimensional algebra.

The Emergence of Nodes

Let us start with a given finite-dimensional associative (and noncommutative, in general) algebra Ω. According to standard conceptions and methods of modern algebraic geometry, as well as the general algebraic approach to physics, we introduce the points of Ω as its irreducible representations (IRs). So, the first step of the specialization procedure is creating (or finding) points of Ω (which will become nodes of the future graph):

$$\{\text{points}\} = \{\text{IRs}\} \tag{9}$$

Standard Set With Nonstandard Topology

When the first spatialization step in a standard way (9) is done we may wish to proceed by connecting the set of nodes by arrows. This problem is mathematically equivalent to equipping the set of points of the algebra by a topology. There are standard recipes for this step as well like, say, the Zariski topology on the prime spectrum of Ω. Unfortunately, on finite-dimensional algebras this topology is always discrete, which leaves us no chance to fit the above requirement of being non-Hausdorff (i.e., not T_2). In terms of graphs that means that the standard recipes can not help us to create arrows. So, we are compelled to find another topology. For these purposes the Rota topology is suggested (first it was introduced in Zapatrin, 1998).

Let Ω be a finite-dimensional algebra. Denote by X the set of points of Ω, each of which we shall associate with a prime ideal in Ω. Consider two points (representations of Ω) $i, j \in X$ and denote by ker i, ker j their kernels. Both of them, being kernels of representations, are two-sided ideals in Ω, in particular, subsets of Ω, hence both of the following expressions make sense:

$$\ker i \cap \ker j \subset \Omega \quad \text{and} \quad \ker i \cdot \ker j \subset \Omega$$

the latter denoting the product of subsets of Ω : ker $i \cdot$ ker $j = \{a \in \Omega | \exists u \in$ ker $i, v \in$ ker $j : uv = a\}$. Since ker i, ker j are ideals, we always have the inclusion ker $i \cdot$ ker $j \subseteq$ ker $i \cap$ ker j which may be strict or not. Define the relation $\propto$ on X as follows:

$$i \propto j \quad \text{if and only if} \quad \ker i \; \ker j \neq \ker i \cap \ker j \tag{10}$$

Then the Rota Topology is the weakest one in which $i \propto j$ implies the convergence $i \to j$ of the point i to the point j. Explicitly, the necessary and sufficient conditions for i to converge to j in the Rota topology reads:

$$i \to j \quad \text{if and only if} \quad \exists k_0, \dots, k_t, \dots, k_n | k_0 = i, \quad k_n = j; \; k_{t-1} \propto k_t \tag{11}$$

This operation is called the transitive closure of the relation $\propto$. Note that, in general, the Rota topology can be defined on any set of ideals.

It was proved by Stanley (1986) that in that particular case when Ω is the Rota algebra of a reflexive transitive graph, then its spatialization X endowed with the Rota topology (11) is homeomorphic to the initial topological space. However, in general if we have two reflexive transitive graphs and an arrow-preserving mapping between them, their Rota algebras may *not* be homomorphic. Recently, "good" classes of reflexive transitive graphs topological spaces were discovered for which the transition to Rota algebras is functorial (Breslav and Zapatrin, 2000; Zapatrin, 2001). In our approach, it supports the existence of the classical limit.

ACKNOWLEDGMENTS

REFERENCES

Aerts, D., Van Bogaert, B., Grib, A. A., and Zapatrin, R. R. (1993). Quantum structures in macroscopical reality. *International Journal of Theoretical Physics* **32**, 489–498.

Behrman, E. C., Steck, J. E., and Skinner, S. R. (1999). A spatial quantum neural computer. In *Proceedings of the International Joint Conference on Neural Networks*, Washington, DC.

Breslav, R. B. and Zapatrin, R. R. (2000). Differential structure of Greechie Logics. *International Journal of Theoretical Physics* **39**, 1027; quant-ph/9902065.

Chrisley, R. L. (1993). Quantum computation and quantum learning. In *Workshop on Brain, Mind and Physics*, Prague.

Davydov, A. S. (1982). *Biology and Quantum Mechanics*, Pergamon Press, Oxford.

Dorogov, A. (2001). Structural synthesis of modular weakly connected neural networks. I. Methodology of structural synthesis of modular neural networks. *Cybernetics and Systems Analysis* **37**, 175–181.

Raptis, I. and Zapatrin, R. R. (2001). Algebraic description of spacetime foam. *Classical and Quantum Gravity* **20**, 4187–4205.

Rota, G.-C. (1968). On the Foundation of Combinatorial Theory, I. The Theory of Möbius Functions. *Zeitschrift für Wahrscheinlichkeitstheorie* **2**, 340.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by backpropagating errors. *Nature* **323**, 533–536.

Savvidy, G. K. (2000). *The system with exponentially degenerate vacuum state*. LANL technical report cond-mat/0003220.

Stanley, R. P. (1986). *Enumerative Combinatorics*, Wadsworth and Brooks, Monterey, CA.

Ventura, D. and Martinez, T. (1997). An artificial neuron with quantum mechanical properties. In *Proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithms*, G. D. Smith, N. C. Steele, and R. F. Albrech, eds., Norwich, England, 1–4 April, 1997, pp. 482–485.

Vitiello, G. (1995). Dissipation and memory capacity in the Quantum Brain Model. *International Journal of Modern Physics B* **9**, 973–989.

Werbos, P. J. (2002). Proof of partial equivalence of classical and quantum dynamics in Bosonic Systems. *International Journal of Bifurcation and Chaos* **12**, 2031–2049.

Zapatrin, R. R. (1998). Finitary algebraic superspace. *International Journal of Theoretical Physics* **37**, 799–815.

Zapatrin, R. R. (2001). Incidence algebras of simplicial complexes. *Pure Mathematics and Its Applications* **11**, 105–118; eprint math.CO/0001065.