Exploring *Shape-Aware* Memory
Through FPGA Prototyping

L.R. Mullin
Department of Computer Science
University at Albany, SUNY
Albany, NY 12222

R.M. Mattheyses
Advanced Computing Lab
GE Global Research
Niskayuna, NY 12309

High performance computing is critically dependent on both the order and speed of data access. In order to maintain performance it may become necessary to reorder data several times during the course of a computation. We propose to integrate reshape and transpose functionality with high speed memories. A formal treatment of reshape and transpose operations is presented in [MR]. The formalism provides a means for performing these operations virtually by tracking their effects on starts, stops, and strides (SSS) for sequential access. Thus some data restructuring can be accomplished without actually moving the data.

The operation of a gate in a simulation of a quantum computation is represented as matrices. The figure below shows the representation of a two input gate applied to a five q-bit system as a $2^5 \times 2^5$ (32x32) density matrix. On the left is the representation for applying the gate to bits 1 and 4. After reshaping into a 10 dimensional hypercube (2^{10} nodes), permuting the index set (generalized transpose) and reshaping back into a square array we have achieved a block diagonalization. It has been shown [M] that the data need not be moved. In fact, through application of the formal definitions of the Psi Calculus[M, MR] we need only track the effect of the transformations on SSS in the index set. We are implementing this tracking of starts, stops and strides in an FPGA which will serve as an interface to memory. This arrangement will allow us to investigate the impact of such *shape-aware* memories on code complexity and run time performance.

The flexibility and speed of the gate array will provide an excellent platform for experimentation. In addition, it will allow us to expand our investigation by adding other formally defined restructuring operations such as rotation, reflection or partitioning of indices along arbitrary axes. The integration of *shape-aware* data transformations, e.g. reshape-transpose, into the fabric of system memories will provide a powerful algebraic basis for the analysis and exploitation of performance improvements.

References

[MR] L. Mullin and J. Raynolds, Optimizing the Fast Fourier Transform over memory hierarchies for embedded digital systems: a fully in-cache algorithm, to appear Proceedings of the the High Performance Embedded Computing Workshop, MIT Lincoln Lab, September 2004.

[M] L. Mullin, "A Uniform Way of Reasoning about Array-Based Computation in Radar", Digital Signal Processing, Elsevier Publishers, accepted for publication, June 2003.

[OD] K.M. Obenland and A.M. Despain, A Parallel Quantum Computer Simulator, Proceedings of High Performance Computing, 1998.

Figure: Example of a two input quantum gate operator being applied to bits 2 and 4 (0 based) using the natural association of states with index values. Location of non-zero elements shown before and after the transformation.
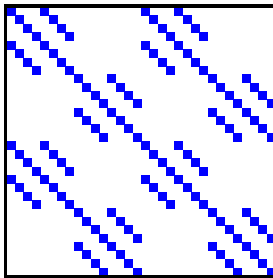
1. Reshape the sparse array (left) into a binary hypercube of dimension 10.
2. Transpose the appropriate axes of the hypercube determined by the bits being gated.
3. Reshape the array back to a 32x32 square with a 4x4 block diagonal structure.