CLAIMS

1. A neural inference chip comprising:

a plurality of network nodes;

a network on chip interconnecting the plurality of network nodes, the network comprising at least one pair of directional paths, the paths of each pair having opposite directions and a common end, wherein the plurality of network nodes is organized in a grid of two or more dimensions with at least one row and at least one column, each row and each column having a pair of directional paths, wherein each network node comprises a router interconnecting a row network and a column of that network node,

wherein the network is configured to accept data at any of the plurality of nodes,

the network is configured to propagate data along a first of the pair of directional paths from a source node to the common end of the pair of directional paths and along a second of the pair of directional paths from the common end of the pair of directional paths to one or more destination nodes.

2. The neural inference chip of claim 1, wherein the network is configured to deliver data to the one or more destination nodes on a second one of the pair of directional paths, according to a network address.

3. The neural inference chip of any one of the preceding claims, wherein the two or more dimensions are ordered such that for each dimension, data propagates along each of its pair of directional paths followed by a subsequent dimension.

4. The neural inference chip of claim 3, wherein each network node comprises a router interconnecting the pair of directional paths of each dimension with the pair of directional paths of the subsequent dimension, the router adapted to deliver data from the second of the pair of directional paths to the first of the pair of directional paths of the subsequent dimension.

5. The neural inference chip of claim 4, further comprising the router at the common end of each pair of directional paths, the router configured to receive data along a first of the pair of directional paths and send data along a second of the pair of directional paths.

6. The neural inference chip of any one of the preceding claims, wherein the common end of each pair is configurable at runtime.

7. The neural inference chip of any one of the preceding claims, wherein the network is partitioned into a plurality of subpartitions, each subpartition operating as an independent network on a subset of the network nodes.

8. The neural inference chip of any one of the preceding claims, wherein the network is configured to accept one packet of data per node per clock cycle, and to propagate data to an adjacent node along each directional path per clock cycle.

9. The neural inference chip of claim 8, wherein the network is configured to propagate data between adjacent network nodes at every clock cycle without stopping or stalling.

10. The neural inference chip of claim 1, wherein the network is configured to deliver data by broadcast to all nodes on the second directional path.

11. The neural inference chip of any one of the preceding claims, wherein the network is configured to deliver data by multicast to a subset of nodes on the second directional path.

12. The neural inference chip of claim 1, wherein the network is configured to deliver data by unicast to a single node on the second directional path.

13. The neural inference chip of any one of the preceding claims, further comprising an external interface adapted to send and receive data from sources and destinations other than the network nodes.

14. The neural inference chip of claim 13, wherein the node routers or common end routers are configured to route data bound to the external interface.

15. A method comprising:

accepting data at a plurality of network nodes, the plurality of network nodes being interconnected by a network on chip, the network comprising at least one pair of directional paths, the paths of each pair having opposite directions and a common end, wherein the plurality of network nodes is organized in a grid of two or more dimensions with at least one row and at least one column, each row and each column having a pair of directional paths, wherein each network node comprises a router interconnecting a row network and a column of that network node;

propagating data along a first of the pair of directional paths from a source node to the common end of the pair of directional paths and along a second of the pair of directional paths from the common end of the pair of directional paths to one or more destination nodes.

16. The method of claim 15, further comprising:

propagating multiple adjacent pieces of data over the network on chip wherein a plurality of source nodes inject data onto the network in a given cycle.

17. The method of claim 15, further comprising:

propagating multiple adjacent pieces of data over the network on chip, wherein a plurality of source nodes inject data onto the network such that the data is interleaved on the network.

18. The method of claim 15, further comprising:

propagating multiple adjacent pieces of data over the network on chip, wherein a plurality of source nodes inject data onto different networks of a first dimension such that the data is later interleaved on a network of a second dimension.

19. The method of claim 15, wherein the injection of data onto the network from one or more source nodes proceeds according to a predetermined schedule.

20. The method of claim 19, wherein the predetermined schedule ensures that data collisions do not occur on the one or more networks.

21. A computer program stored on a computer readable medium and loadable into the internal memory of a digital computer, comprising software code portions, when said program is run on a computer, for performing the method of any one of claims 15 to 20.