Department of Computer Science

Advanced Computer Architecture

“Ball Computer”

Or

How I Learned to Stop Worrying and Love Computing with 3D Wireless Grids

Amir Mansoor Kamali, Richard Hind, Christopher Crispin-Bailey, Jim Austin
{amir, rhind, chrisb}@cs.york.ac.uk, jim.austin@york.ac.uk

THE UNIVERSITY *of York*

Introduction

In this project we are investigating if wireless devices can replace wireline networks to serve as a connection network for a massively parallel computer.

The radio devices are getting faster, cheaper and smaller. They are even built on silicon occupying a fraction of a square centimetre. Despite a significant drop in their power consumption, radio devices still consume more energy compared to wire connections. The data rates of wireless devices are still lower than their wired counterparts; although this gap is much narrower than past decade.

By using wireless devices a significant reduction of system complexity and cost is anticipated due to elimination of wirings. High flexibility and scalability are also among the benefits of such a system. The lower bandwidth for data links and higher power consumption are still big challenges which may limit the wide usage of such a scheme. In this project we want to find a more precise picture of the benefits and restrictions of a 3D wireless massively parallel machine.

A novel concept machine called the *Ball Computer* is presented which is basically a multi-channel 3D wireless grid which serves as a connection network for a parallel machine. As it is the first step, the machine is implemented only in a simulation environment. A novel method to share the available radio bandwidth is invented. To test the performance of the system a couple of task models are built which are simulated traffic load generators. Moreover, a set of simulation environment and visualisation tools is implemented.

Wireline vs. Wireless

Wireless still has some big problems. Among the biggest: bandwidth and energy consumption.

Bandwidth

The gap between wireline and wireless technologies is narrowing.

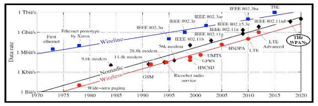

Figure 1: Comparison between wireline and wireless technologies over last decades as reported by Thomas Kürner[1].

A survey on on-chip data communication rates during past decade (Figure 2) shows a sharp increase:

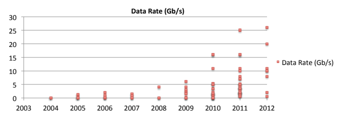

Figure 2: On-chip short range millimetre wave radio communication over last decade.

Energy Consumption

Wirelines need as low energy per bit as 1 pJ/bit while the same number for wireless transfer over 10mm is at least in the range of 10s of pJ/bit. But the figures (e.g. figure 3) show the energy per bit is falling for wireless transactions.

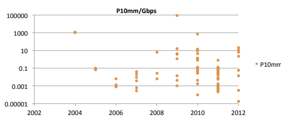

Figure 3: A very rough estimation of energy needed for data transfer of rate 1Gbps over a 10mm link

Power Delivery

There are still big challenges on this field and an effective technology is yet to be found.

Wireless solutions:

a) Capacitive coupling: Occupies very small area but it is suitable only for very short ranges in nanometre scale.
b) Inductive coupling: Is capable of transfer power to longer distances but both methods do not have high enough performance.
c) Radio waves.

Optic connections can also be a solution.

A combination of RF for longer distances and optical links for shorter distances seems to be a good choice for an efficient power delivery.

Area

Over recent decade the size of on-chip transceivers is sharply fallen and it is anticipated to keep falling over coming years.

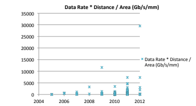

Figure 4: Today on-chip radios can send data on faster rates and over longer distances and at the same time occupy smaller areas on silicon.

Most researches are on implementing 60GHz radios on silicon. Some has started testing higher bands including 120GHz. This can increase the data rate and at the same time decrease the size of the devices as the size of on-chip antennas can be smaller.

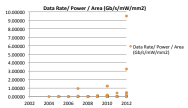

Figure 5: The power consumption is also fallen so that the higher data rates are now transmitted using less energy and less silicon area.

Simplicity, Cost and Scalability

A wireless network does not need wiring costs. The complexity of the system is also reduced.

A wireless network is highly scalable. In contrast, in a wired network there is a limited capacity in adding new nodes to a shared bus. Using switches makes it easier to let the bandwidth stay unaffected by new nodes but the problem with switch capacity is still there.

In a wireless scheme nodes can easily join the network just by tuning to a proper frequency channel. There will be no worries on the shortage of neither bandwidth nor I/O links.

Ball Computer Architecture

We propose a 3D wireless network of processing elements that can perform as a massively parallel computer. The machine is informally called the ***Ball Computer***. Nodes are in their most compact form. Each node has up to 6 neighbours in a 2D plain and 12 neighbours in a 3D grid.

Figure 6 shows the position of a node between its neighbours in both 2D and 3D networks.

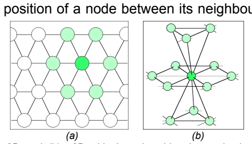

Figure 6: Nodes in (a) a 2D; and, (b) a 3D grids. In real world nodes are in physical contact with all of their immediate neighbours.

Each node is made of a small electronic board inside a small plastic ball (1 or 2 cm in diameter). Each node has a processor and a set of radio transceivers.

At the current design the number of transceivers is 8. The radios can be tuned to any channels but it is proved that in order to have no packet collision it is important to assign the channels based on a network partitioning algorithm.

That algorithm dictates that for a 3D grid of any size each node needs 8 radio links shared with its neighbours. Moreover, any increase in the size of the network does not affect the total number of channels.

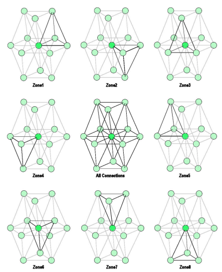

Figure 7: An illustration of a node with its 12 neighbours in a 3D environment.

Network Partitioning

Figure 7 shows how a node in a 3D grid of a ball computer interacts with its neighbours. The central node has eight radio links each of which is shared with 3 neighbours.

Each pyramid-shape subset of the network is called a zone. This is the outcome of a novel network partitioning algorithm.

To guarantee total elimination of packet collision, it is important to assign proper channels to each radio.

The algorithm is trying to satisfy the following criteria:

a) To eliminate packet collision;
b) To maximise the connectivity between nodes;
c) To provide redundant links between neighbours; and
d) To keep the number of channels as low as possible.

Network partitioning has been investigated mobile phone network design as well as wireless internet access networks.

Our network partitioning has two sub-algorithms as follow:

Zoning

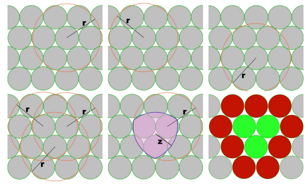

Figure 9: A zone forms out of three overlapping radio ranges. Nodes outside a zone cannot use the same channel.

The network is partitioned into overlapping subsets (zones). Inside a zone all nodes can hear from others and no packet collision can happen.

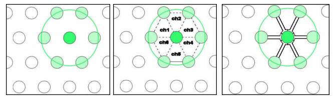

Figure 10: A central node and its zones. Compare the wireless connections with their wireline equivalent.

In figures 10 and 11 a 2D multi-channel wireless grid is compared with its wireline equivalent.

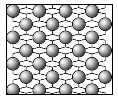

Figure 11: full wireline network equivalent to a 2D ball computer. Each node is connected through 6 separate small networks to its direct neighbours.

Channel Assignment

Channel assignment can be compared with map colouring problem. In channel assignment a zone not only cannot share its colour with that of its neighbours but also it cannot share its colour with that of the neighbours of its neighbours.

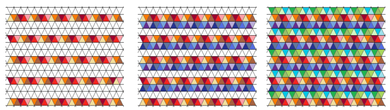

Figure 12: Choosing channels for zones. Each colour represents a channel. Close zones cannot have the same colour/channel.

The number of colours does not increase when the size of the networks increases. For a network of any size the total number of channels is 86.

Task Modelling

Task models are artificial traffic load generators which mimic a certain task or class of tasks without dealing with task-specific details to get rid of unimportant details of real-world tasks. Two main task models were studied in this project:

a) A task model which makes least possible dependencies between tasks on different nodes. Each task is almost independent of that of other nodes.
b) A task model inspired (but not restricted to) FFT algorithm which makes a vast and tightly connected network of tasks. The finish time of nodes are directly dependent on their neighbours.

Simulation Results

The network partitioning algorithm produced perfect results and led to complete elimination of packet collision. Table 1 lists how many times packet collision occurs in a 2D simulated network when 1, 2 and 6 distinct frequencies are used for each node.

No. of Frequencies	No. of Packet collisions
1	1731342
2	665271
6	0

Table 1: Effect of multi-channel communication on packet collision in a 2D network.

The effect of using multiple channels is studied. The task time and the overall performance of the parallel machine are tested. The results can be seen in figures 13 and 14. The number of nodes was 800 in this test.

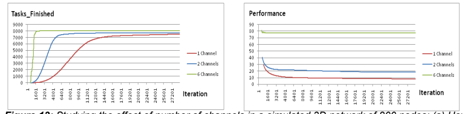

Figure 13: Studying the effect of number of channels in a simulated 3D network of 800 nodes: (a) How quickly tasks finish; (b) The performance of the whole system

The performance of an FFT task model is shown in figure 14. The performance is measured over different data sizes and different intervals between data packets. It can be seen that for larger data sizes the ball computer performs much better.

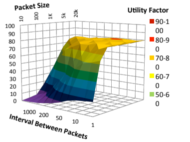

Figure 14: The performance of an FFT task model in a network of 800 nodes.

Figures 15 and 16 show how the idea of multi-tasking increases the performance of the network.

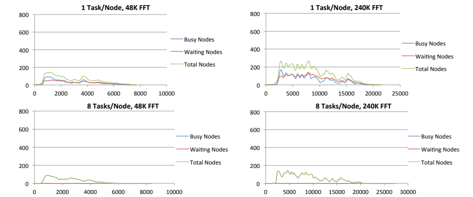

Figure 15: The effect of multi-tasking on improvement of the performance of an FFT task model.

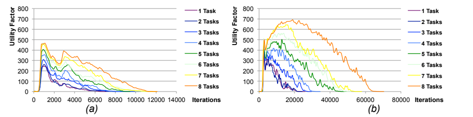

Figure 16: Utility factor in a set of multi task FFT with a) 48K bytes, b) 240K Bytes.

Visualisation Results

The performance of the simulated network can be studied by measuring a number of traffic and computation factors. As an example, figure 17 shows that there are a few nodes which are suffering from congestion of their I/O queues. Architects and software designers can use these data to fine-tune the network parameters or develop more efficient programs for this topology.

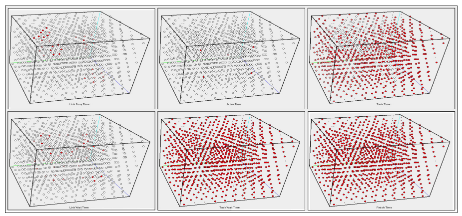

Figure 17: The behaviour of the ball computer can be studied. Bottlenecks and hot spots need special attention.

1: https://mentor.ieee.org/802.15/dcn/12/15-12-0320-01-0thz-what-s-next-wireless-communication-beyond-60-ghz-tutorial-ig-thz.pdf