Converse BlueGene Emulator

Gengbin Zheng

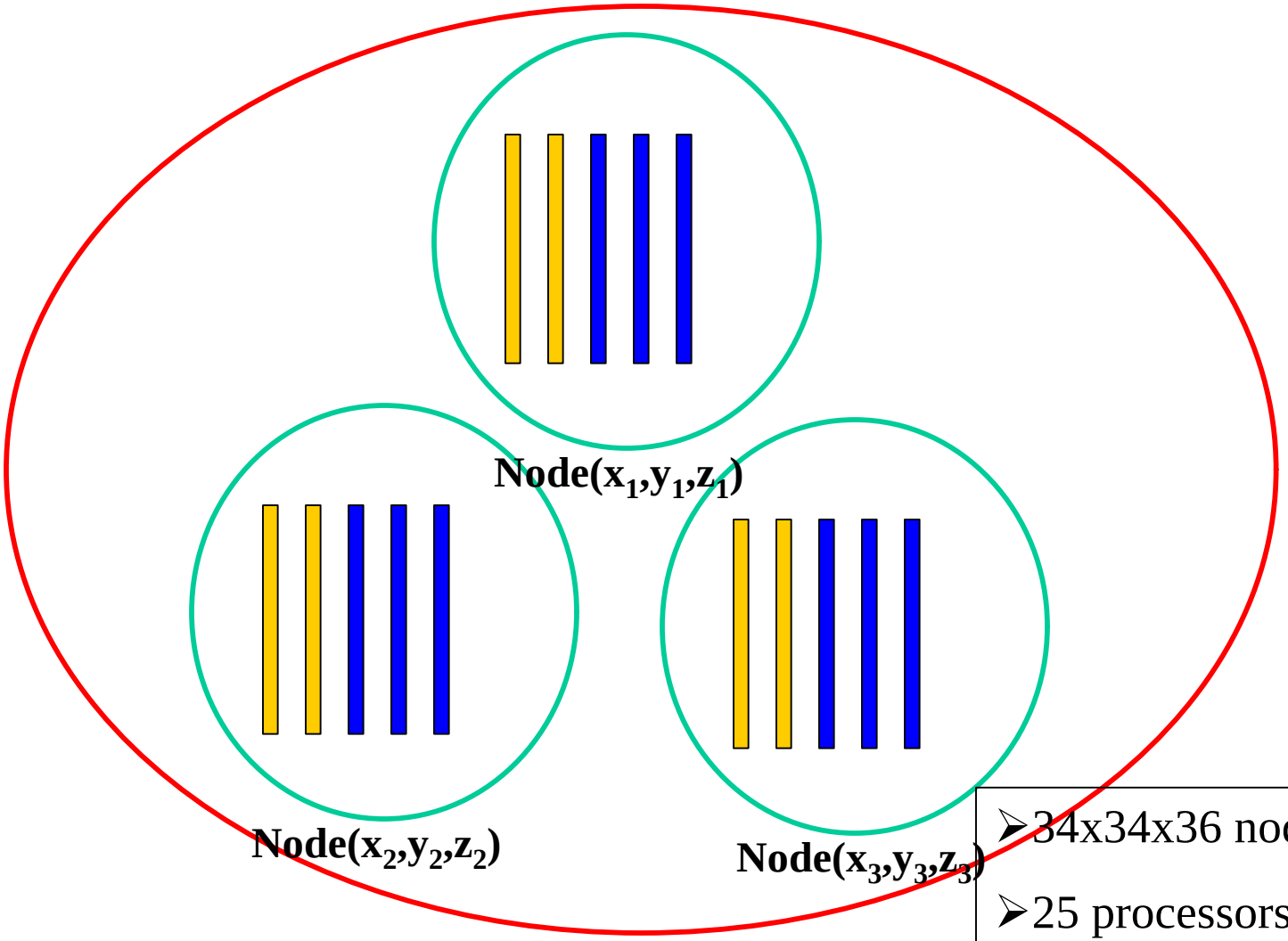
Parallel Programming Lab

2/27/2001

Objective

- Completely rewritten the previous Charm++ Blue Gene emulator;
- Bluegene emulator for architecture studying (PetaFLOPS computers);
- Performance estimation (with proper time stamping)
- Provide API for building Charm++ on top of it.

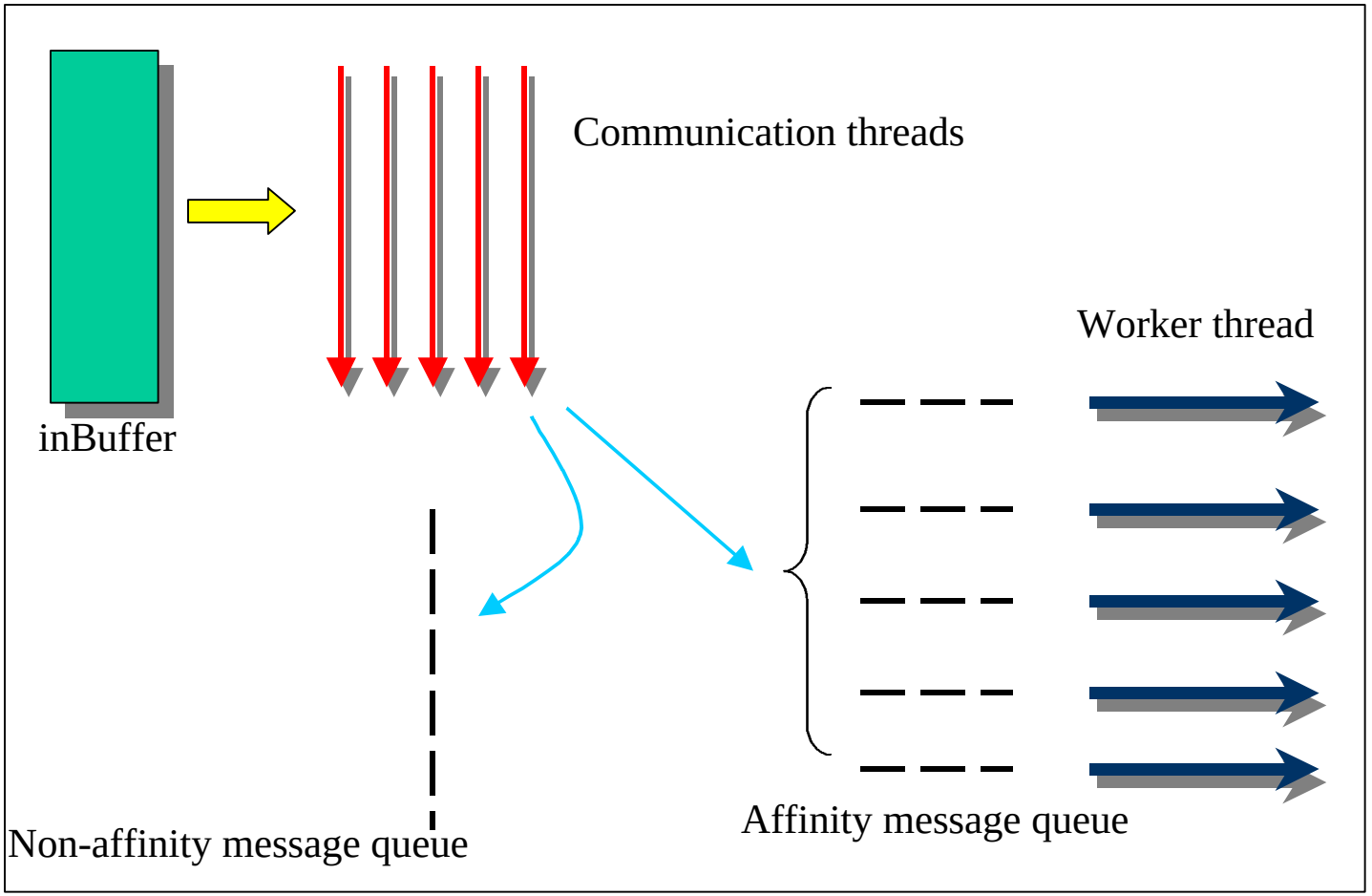
Big picture - emulator



Emulator Processor

- 34x34x36 nodes
- 25 processors per node
- 8 threads per processor

Bluegene Emulator



Node Structure

Communication Threads

- Communication threads get messages from inbuffer
  - If small work, execute the task itself.
  - If affinity message, put to the thread's local queue;
  - If non-affinity message, put to the node queue;

Worker threads

- Worker threads examine messages from two queues: affinity queue and non-affinity queue;
- Compare the receive-time of two messages and pick the one that comes first and execute it;

Low-level API

- Class ***NodeInfo:***

  id, x, y, z, udata, commThQ, workThQ
- Class ***ThreadInfo:*** (thread private variable)

  id, type, myNode, currTime
- Class ***BgMessage***:

  node, threadID, handlerID, type, sendTime, recvTime, data
- getFullBuffer()
- checkReady()
- addBgNodeMessage()
- addBgThreadMessage()
- sendPacket()

User's API

- BgGetXYZ()
- BgGetSize(), BgSetSize()
- BgGetNumWorkThread(), BgSetNumWorkThread()
- BgGetNumCommThread(), BgSetNumCommThread()
- BgRegisterHandler()
- BgGetNodeData(), BgSetNodeData()
- BgGetThreadID(), BgGetGlobalThreadID()
- BgGetTime()
- BgSendPacket(), etc
- BgShutdown()
- BgEmulatorInit(), BgNodeStart()

Bluegene application example - Ring

```
void BgEmulatorInit(int argc, char **argv)
{
  if (argc < 6)     CmiAbort("Usage: <ring> <x> <y> <z> <numCommTh> <numWorkTh>\n");
  BgSetSize(atoi(argv[1]), atoi(argv[2]), atoi(argv[3]));
  BgSetNumCommThread(atoi(argv[4]));        BgSetNumWorkThread(atoi(argv[5]));
  passRingID = BgRegisterHandler(passRing);
}

void BgNodeStart(int argc, char **argv)
{
  int x,y,z;  int nx, ny, nz;  int data=888;
  BgGetXYZ(&x, &y, &z);           nextxyz(x, y, z, &nx, &ny, &nz);
  if (x == 0 && y==0 && z==0)
      BgSendPacket(nx, ny, nz, passRingID, LARGE_WORK, sizeof(int), (char *)&data);
}

void passRing(char *msg)
{
  int x, y, z;   int nx, ny, nz;   int data = *(int *)msg;
  BgGetXYZ(&x, &y, &z);            nextxyz(x, y, z, &nx, &ny, &nz);
  if (x==0 && y==0 && z==0)     if (++iter == MAXITER) BgShutdown();
  BgSendPacket(nx, ny, nz, passRingID, LARGE_WORK, sizeof(int), (char *)&data);
}
```

Performance

- Pingpong
  - Close to Converse pingpong;
    - 81-103 us v.s. 92 us RTT
  - Charm++ pingpong
    - 116 us RTT
  - Charm++ Bluegene pingpong
    - 134-175 us RTT

Charm++ on top of Emulator

- BlueGene thread represents Charm++ node;
- Name conflict:
  - Cpv, Ctv
  - MsgSend, etc
  - CkMyPe(), CkNumPes(), etc