ClearSpeed™

LINPACK
Getting started guide

Purpose

This brief guide is intended to provide basic guidance for achieving a demonstrable accelerated LINPACK using hardware and software supplied by ClearSpeed.

The reader is expected to have some Linux technical experience and whilst familiarity with LINPACK is beneficial for further optimization, it is not essential for the basic goal of demonstrating accelerated LINPACK.

Prerequisites

- ClearSpeed MPICH tarball[1],
- ClearSpeed HPL (high performance LINPACK) tarball[1],
- ClearSpeed base software package V2.50,
- ClearSpeed Advance Accelerator card,
- supported operating system.

1. Downloaded from http://support.clearspeed.com/downloads/applications/linpack/

Table of contents

1 Building and installing the components

1.1 ClearSpeed Advance card(s) and base package

Please refer to the installation guides for the Advance X620 (or e620) card and the base software (http://support.clearspeed.com/downloads/).

Please ensure that the card diagnostics, a part of the base software, has successfully completed its tests before proceeding with any LINPACK testing or benchmarking.

Note: *Both MPICH and LINPACK packages can be installed as either regular user or super user.*

1.2 MPICH

ClearSpeed provide a copy of MPICH[1] that has been modified from the default to permit messages up to one gigabyte in size (the default is much smaller than this).

1. Untar `ClearSpeed.mpich.tar.gz`.
2. Move to the `mpich-1.2.5` directory.
3. Configure MPICH with `./configure` and any required prefix options for installation.
4. Build MPICH with

   ```
   make ; make install
   ```

   The ClearSpeed modifications provide a warning that the message size has been increased with respect to the reference source code.

Example:

Use the following command to configure MPICH so that it will install to directory `/filesystem/shared/`:

```
./configure --prefix=/filesystems/shared/
```

1. Available at http://www-unix.mcs.anl.gov/mpich/mpich1/

1.3 HPL

ClearSpeed provide a copy of HPL[1] that has been modified to overlap MPI communication with panel updates using threads rather than the default. Additionally, the HPL source has been instrumented for performance visualization for those customers with access to the ClearSpeed Visual Profiler.

1. Untar `ClearSpeed.hpl.tar.gz`.
2. Move to the `hpl` directory.
3. Modify the file `Make.Clearspeed.no.profiler.support` to point the `MPdir` variable to the installed MPICH directory (see *Note 3* in *Section 1.2: MPICH*).
4. Modify the file `Make.Clearspeed.no.profiler.support` to point the `TOPdir` variable to the installation directory (`/.../hpl`).
5. Using `make arch=Clearspeed.no.profiler.support` build LINPACK.

1. Available at http://www.netlib.org/benchmark/hpl/

2 Running LINPACK and choosing parameters

Tuning LINPACK for a specific cluster is dependent on the characteristics of the system itself. The following list details the parameters that are the most significant and is based on the assumption that the MPICH installation provided with HPL has been used to launch LINPACK rather than a bespoke solution like those delivered with a high-speed interconnect.

To measure the performance differential between the base cluster without using the Advance card, run the benchmark before the card installation, or by setting `CS_HOST_BLAS_ASSIST_PERCENTAGE` to the value `100` (see *Section 2.2: Launch scripts*).

2.1 HPL.dat parameter file

The parameter guide is specific to a version of ClearSpeed's software library (CSXL) and this guide assumes the package version stated in the *Prerequisites* section of this document. This list in is addition to the LINPACK tuning guide(1) and FAQ(2).

- NB: set to 1152 or a multiple of 288.
- N: set to use most of the cluster or machine memory.
- L1 and U: set to no-transposed (1).
- Depth: set to 1 for clusters (more than one machine).

1. Available at http://www.netlib.org/benchmark/hpl/tuning.html/
2. Available at http://www.netlib.org/benchmark/hpl/faqs.html/

2.2 Launch scripts

Launching LINPACK with one process (P=Q=1) on the current node is simple.

1. Source the ClearSpeed `bashrc` file in `/opt/clearspeed/CSX600_m512_le/bin/`
2. Set up the host `BLAS` library with the `CS_HOST_BLAS` environment variable.
3. Calibrate and set any BLAS host assistance according to the CSXL user guide.
4. Use `./xhpl` to launch LINPACK.

For cluster systems a simple script launched with MPICH can run the above process.

Example:

```
#!/bin/sh
cd /usr/local/hpl/bin/ClearSpeed.no.profiler.support/
source /opt/clearspeed/csx600_m512_le/bin/bashrc
export CS_HOST_BLAS=/lib64/blas.co
export CS_BLAS_HOST_ASSIST_PERCENTAGE=11
./xhpl $@
```

Assume that the above is in a file named `launch.sh`. In the above example the host assist fraction is calibrated to 11%. To measure the performance of the base system without acceleration, set this value to 100% or remove the Advance cards from the system.

3 Running LINPACK

Prerequisites:

- MPICH built as above with a launch tool `/usr/local/bin/mpirun`
- MPICH machines file with one name per Advance card (hence machines with multiple cards have multiple entries in the machines file)
- LINPACK built as above with the executable `xhpl`
- LINPACK parameters set as above in `HPL.dat`
- LINPACK launch script as above in `$PWD/launch.sh`

With the total number of cards in the system being *K*, launch LINPACK with the following command:

```
/usr/local/bin/mpirun -np K -machinefile machines.HPL ./launch.sh
```

If the cluster used has a head node to launch the application, and it does not itself have an Advance card, add the `-nolocal` option to `mpirun`:

```
/usr/local/bin/mpirun -np K -nolocal -machinefile machines.HPL ./launch.sh
```

Appendix A Note on larger clusters

Clusters of four or more nodes, especially with fast host processors (for example, Intel 5160 or faster) should use a high performance interconnect (better than Gigabit Ethernet) for MPI, such as Infiniband. Scaling of performance with or without ClearSpeed accelerator cards will degrade rapidly as the number of nodes is increased with Ethernet.

Additionally, users of MPI stacks for high speed interconnect, for example MVAPICH, should take into consideration the configuration options when building the MPI stack. This release of LINPACK is intended to aggregate multiple x86 cores and ClearSpeed accelerator cards into one process. It is possible to configure a build, of say MVAPICH, such that only one x86 core is available per process. We recommend either re-configuring and rebuilding the MPI stack, or modifying HPL to reset the process affinity masks. The latter is achieved by adding the following to `main()` in the file `testing/ptest/HPL_pddriver.c`.

```
...
  int sched_result=0;
  unsigned long sched_mask=0xffffffff;
...

  MPI_Init( &ARGC, &ARGV );

sched_result=sched_setaffinity(getpid(), sizeof(sched_mask),
                    &sched_mask);
```

This should make all cores available to run a multi-threaded host BLAS.

Revision history

Date	Revision	Changes
September 2010	1.B	Updated company information
September 2008	1.A	Amendment to copyright statement.
July 2007	1.0	Initial version.

Table 1. Document revision history

ClearSpeed Technology Ltd
130 Aztec West
Park Avenue
Bristol BS32 4UB
United Kingdom

Tel: +44 (0)1454 629 623
Fax: +44 (0)1454 629 624

Email: info@clearspeed.com

Web: http://www.clearspeed.com

Support: http://support.clearspeed.com

Acknowledgments:

LINPACK: This product includes software developed at the University of Tennessee, Knoxville, Innovative Computing Laboratories