inconsistent pinning behavior due to missing SMT info on AMD Zen

The mdrun native hardware detection does recognize hardware thread order on the AMD Zen uarch processors, but does not detect the SMT to correctly assign the hardware threads to cores. As a result (besides the reporting being incorrect), only at most half of the cores are used when a run is launched with #threads<#hwthreads/2. On Intel with HT in such cases the default stride is switched to 2 to spread threads across cores when the total thread count is <=#cores.

- Native detection

Hardware topology: Basic

Sockets, cores, and logical processors:

Socket 0: [ 0 ] [ 16 ] [ 1 ] [ 17 ] [ 2 ] [ 18 ] [ 3 ] [ 19 ] [ 4 ] [ 20 ] [ 5 ] [ 21 ] [ 6 ] [ 22 ] [ 7 ] [ 23 ] [ 8 ] [ 24 ] [ 9 ] [ 25 ] [ 10 ] [ 26 ] [ 11 ] [ 27 ] [ 12 ] [ 28 ] [ 13 ] [ 29 ] [ 14 ] [ 30 ] [ 15 ] [ 31 ]

- Detection with hwloc:

Hardware topology: Full, with devices

Sockets, cores, and logical processors:

Socket 0: [ 0 ] [ 16 ] [ 17 ] [ 2 ] [ 18 ] [ 3 ] [ 19 ] [ 4 ] [ 20 ] [ 5 ] [ 21 ] [ 6 ] [ 22 ] [ 7 ] [ 23 ] [ 8 ] [ 24 ] [ 9 ] [ 25 ] [ 10 ] [ 26 ] [ 11 ] [ 27 ] [ 12 ] [ 28 ] [ 13 ] [ 29 ] [ 14 ] [ 30 ] [ 15 ] [ 31 ]

Associated revisions

Revision b9c04931 - 03/01/2018 10:34 PM - Berk Hess
Detect AMD SMT topology

On AMD Zen the cpuinfo code detected hyperthreading but put all threads on different cores in the topology. Now the correct topology is detected using extended APIC.
Also disabled topology detection for non-AMD, non-x2APIC x86.

Fixes #2388
Change-Id: I194f3e09e669c20d1d62355a36be062e6c264e

History

#1 - 01/19/2018 05:20 PM - Szilárd Páll
Having briefly looked at CpuInfo::detect(), I'm not sure whether this is a technical limitation of cpuid on AMD, but if it is not possible/hard to correct, I

02/19/2020
suggest we make the assumption that if the logical processor indexing suggests that SMT is on, we switch to stride 2 as we do on Intel. IIUC, this
should be safe and as long as sibling + index indicates #cores stride, the kernel has to be configured in a very strange manner for the the assumption
to not be correct.

#2 - 01/19/2018 09:20 PM - Mark Abraham
Does 2016 have an issue?

#3 - 01/20/2018 02:08 AM - Szilárd Páll
- Description updated

Mark Abraham wrote:

Does 2016 have an issue?

I don't think the code has changed, so it should. IIUC there is an assumption made hat when there is no detailed topology info, stride should always
be 1.

#4 - 01/20/2018 02:09 AM - Szilárd Páll
- File _test-native_1x16_thrp01_gmx16.log added

#5 - 01/20/2018 02:10 AM - Szilárd Páll
PS: behavior confirmed with '16.

#6 - 01/20/2018 02:10 AM - Szilárd Páll
- Affected version - extra info set to 2016.4

#7 - 01/31/2018 02:49 PM - Erik Lindahl
- Tracker changed from Bug to Feature
- Affected version - extra info deleted (2016.4)
- Affected version deleted (2018)

Well, it's not technically a bug since the hardware info module properly detects that we can't see it, and correctly specifies that only basic topology
information is available. It would be nice to have it for this type of hardware too, but that's a new feature I'm not sure how easy it is to implement
(depends on whether it can be extracted from cpuid).

Overall, isn't the best solution to simply recommend people to use hwloc? I'm skeptic to start assuming we have SMT in cases where it has not been
properly detected, because such assumptions have historically come back and bitten us in hard ways.

#8 - 02/06/2018 10:41 AM - Szilárd Páll
- Tracker changed from Feature to Bug
- Subject changed from SMT info AMD Zen lacking with native hardware detection to inconsistent pinning behavior due to missing SMT info on AMD Zen
- Affected version set to 2018

This is not a feature request. The described issue leads to inconsistent behavior -- both between the hwloc and no-hwloc build on the same machine
and between different x86 platforms (in that in this single case out of four use-cases a different set of threads will be used by default). There is no
practical difference between Intel HT and AMD SMT, so we should not implement entirely different default behavior unless something makes it
impossible to do the same detection we do on Intel -- that's why I asked whether the Intel-specific cpuid can be extended.

Suggesting people to use hwloc won't solve the inconsistency either (though removing the Intel-only topology detection and related assumptions
would perhaps improve things a bit).

#9 - 02/23/2018 03:24 PM - Gerrit Code Review Bot
Gerrit received a related patchset '1' for Issue #2388.
Uploader: Berk Hess (hess@kth.se)
Change-Id: gromacs~release-2018~I194f3e09e669c20d1d62355a36be062a6cce264e
Gerrit URL: https://gerrit.gromacs.org/7622

#10 - 02/23/2018 03:37 PM - Mark Abraham
- Target version set to 2018.1
Note: Even with the patch to add more AMD SMT detection, there **will** be cases where we do not perfectly detect SMT on some architectures, in particular new ones or if APIC support is disabled in BIOS.

It seems important that we also turn off thread pinning when not using all logical cores on a system, unless full topology information is available.

Erik Lindahl wrote:

> It seems important that we also turn off thread pinning when not using all logical cores on a system, unless full topology information is available.

We do have pinning in all cases when we do not use all hardware threads. I was referring to the use-cases of Intel+HT vs AMD+SMT with only half of the hardware threads used.

The cpuinfo code returned from `detectX86LogicalProcessors()` a vector of logical processors that didn't reflect the hardware. That sets the support level to `LogicalProcessorInfo` (ie the highest). So it is correct for the pinning code to act. Because there were 32 logical processors found in `2018-no-hwloc-nb-cpu.log` (from the attached tarball), `m`drun decided to use only 24 threads, and to pin. That's correct behaviour given that the detection was inaccurate.

Gerrit received a related DRAFT patchset '1' for Issue #2388. Uploader: Berk Hess (hess@kth.se)
Change-Id: gromacs~master~I194f3e09e669c20d1d62355a36be062e6c6e264e
Gerrit URL: https://gerrit.gromacs.org/7635

Applied in changeset b9c04931de7709424b2931720b5bb6952c005780.

Files

<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>_test-native_1x16_thrp01.log</td>
<td>22.7 KB</td>
<td>01/19/2018</td>
<td>Szilárd Páll</td>
</tr>
<tr>
<td>_test-hwloc_1x16_thrp01.log</td>
<td>23.5 KB</td>
<td>01/19/2018</td>
<td>Szilárd Páll</td>
</tr>
<tr>
<td>_test-native_1x16_thrp01_gmx16.log</td>
<td>22.1 KB</td>
<td>01/20/2018</td>
<td>Szilárd Páll</td>
</tr>
<tr>
<td>amd-cpuinfo-fix.tgz</td>
<td>12.3 KB</td>
<td>02/27/2018</td>
<td>Mark Abraham</td>
</tr>
</tbody>
</table>