Creating Hybrid Codes with Cray Reveal

Heidi Poxon
Technical Lead
Programming Environment
Cray Inc.

March 2016
Legal Disclaimer

Information in this document is provided in connection with Cray Inc. products. No license, express or implied, to any intellectual property rights is granted by this document.

Cray Inc. may make changes to specifications and product descriptions at any time, without notice.

All products, dates and figures specified are preliminary based on current expectations, and are subject to change without notice.

Cray hardware and software products may contain design defects or errors known as errata, which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Cray uses codenames internally to identify products that are in development and not yet publically announced for release. Customers and other third parties are not authorized by Cray Inc. to use codenames in advertising, promotion or marketing and any use of Cray Inc. internal codenames is at the sole risk of the user.

Performance tests and ratings are measured using specific systems and/or components and reflect the approximate performance of Cray Inc. products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance.

The following are trademarks of Cray Inc. and are registered in the United States and other countries: CRAY and design, SONEXION, URIKA, and YARCDATA. The following are trademarks of Cray Inc.: ACE, APPRENTICE2, CHAPEL, CLUSTER CONNECT, CRAYPAT, CRAYPORT, ECOPHLEX, LIBSCI, NODEKARE, THREADSTORM. The following system family marks, and associated model number marks, are trademarks of Cray Inc.: CS, CX, XC, XE, XK, XMT, and XT. The registered trademark LINUX is used pursuant to a sublicense from LMI, the exclusive licensee of Linus Torvalds, owner of the mark on a worldwide basis. Other trademarks used in this document are the property of their respective owners.

Copyright 2016 Cray Inc.
When to Move to a Hybrid Programming Model

- **When code is network bound**
  - Increased MPI collective and point-to-point wait times

- **When MPI starts leveling off**
  - Too much memory used, even if on-node shared communication is available
  - As the number of MPI ranks increases, more off-node communication can result, creating a network injection issue

- **When contention of shared resources increases**
Approach to Adding Parallelism

1. Identify key high-level loops
   - Determine where to add additional levels of parallelism

2. Perform parallel analysis and scoping
   - Split loop work among threads

3. Add OpenMP layer of parallelism
   - Insert OpenMP directives

4. Analyze performance for further optimization, specifically vectorization of innermost loops
   - We want a performance-portable application at the end
WARNING!!!

- Nothing comes for free, nothing is automatic
  - Hybridization of an application is difficult
  - Efficient code requires interaction with the compiler to generate
    - High level OpenMP structures
    - Low level vectorization of major computational areas

- Performance is also dependent upon the location of the data
  - CPU: NUMA, first-touch
  - Accelerator: resident or data-sloshing

- Software such as Cray's Hybrid Programming Environment provides tools to help, but cannot replace the developer's inside knowledge
The Problem – How Do I Parallelize This Loop?

- How do I know this is a good loop to parallelize?
- What prevents me from parallelizing this loop?
- Can I get help building a directive?

subroutine sweepz
    do j = 1, js
        do i = 1,isz
            radius = zxc(i+mypez*isz)
            theta = zyc(j+mpey*js)
            do m = 1,npez
                do k = 1,ks
                    n = k + ks*(m-1) + 6
                    r(n) = recv3(1,j,k,i,m)
                    p(n) = recv3(2,j,k,i,m)
                    u(n) = recv3(5,j,k,i,m)
                    v(n) = recv3(3,j,k,i,m)
                    w(n) = recv3(4,j,k,i,m)
                enddo
            enddo
        enddo
    enddo
... 
call ppmlr
    do k = 1,kmax
        xa(n) = zza(k)
        dx(n) = zdz(k)
        xa0(n) = zza(k)
        dx0(n) = zdz(k)
        e(n) = p(n)/(r(n)*gamm)+0.5*(u(n)**2+v(n)**2+w(n)**2)
    enddo
    call ppmlr
    return
end

subroutine ppmlr
    call boundary
    call flatten
    call paraset(nmin-4,nmax+5,para,dx,xa)
    call parabola(nmin-4,nmax+4,para,p,dp,p6,p1,flat)
    call parabola(nmin-4,nmax+4,para,r,dr,r6,r1,flat)
    call parabola(nmin-4,nmax+4,para,u,du,u6,ul,flat)
    call states(pl,ul,r1,p6,u6,dp,dr,plft,ulft,rlft,prgh,urgh,rrgh)
    call riemann(nmin-3,nmax+4,gam,prgh,urgh,rrgh,&plift,ulift,rlift pmid umid)
    call evolve(umid, pmid) \ contains more calls
    call remap \ contains more calls
    call volume(nmin,nmax,ngeom,radius,xa,dx,dvol)
    call remap \ contains more calls
return
end
Simplifying the Task with Reveal

- Navigate to relevant loops to parallelize
- Identify parallelization and scoping issues
- Get feedback on issues down the call chain (shared reductions, etc.)
- Optionally insert parallel directives into source
- Validate scoping correctness on existing directives
Hybridization Step 1: Loop Work Estimates

Gather loop statistics using CCE and the Cray performance tools to determine which loops have the most work

- Helps identify high-level serial loops to parallelize
  - Based on runtime analysis, approximates how much work exists within a loop

- Provides the following statistics
  - Min, max and average trip counts
  - Inclusive time spent in loops
  - Number of times a loop was executed
perftools-lite-loops

- CrayPat-lite loop work estimates
- Must be used with Cray compiler
- Load before building and running program to get loop work estimates sent to stdout and to .ap2 file for use with Reveal

Automates loop work experiment by:
- modifying the compile and link steps to include CCE’s –h profile_generate option
- instrumenting the program for tracing (pat_build –w)

–h profile_generate reduces compiler optimization levels
- After experiment is complete, unload perftools-lite-loops to prevent further program instrumentation.
Collecting Loop Work Estimates

- Load PrgEnv-cray module (must use CCE)
- Load perftools-base module if not already loaded
- Load perftools-lite-loops module

- Build and run application

- Loop work estimates will be available for Reveal in file with .ap2 extension and in text format in file with .rpt extension

- Unload perftools-lite-loops module
### Table 2: Loop Stats by Function (from `-hprofile_generate`)

<table>
<thead>
<tr>
<th>Loop Incl Time</th>
<th>Loop Hit</th>
<th>Loop Trips Avg</th>
<th>Loop Trips Min</th>
<th>Loop Trips Max</th>
<th>Function=../LOOP[.] PE=HIDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.995914</td>
<td>100</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepy_LOOP.1.1i.33</code></td>
</tr>
<tr>
<td>8.995604</td>
<td>2500</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepy_LOOP.2.1i.34</code></td>
</tr>
<tr>
<td>8.894750</td>
<td>50</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepz_LOOP.05.1i.49</code></td>
</tr>
<tr>
<td>8.894637</td>
<td>1250</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepz_LOOP.06.1i.50</code></td>
</tr>
<tr>
<td>4.420629</td>
<td>50</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepx2_LOOP.1.1i.29</code></td>
</tr>
<tr>
<td>4.420536</td>
<td>1250</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepx2_LOOP.2.1i.30</code></td>
</tr>
<tr>
<td>4.387534</td>
<td>50</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepx1_LOOP.1.1i.29</code></td>
</tr>
<tr>
<td>4.387457</td>
<td>1250</td>
<td>25</td>
<td>0</td>
<td>25</td>
<td><code>sweepx1_LOOP.2.1i.30</code></td>
</tr>
<tr>
<td>2.523214</td>
<td>187500</td>
<td>107</td>
<td>0</td>
<td>107</td>
<td><code>riemann_LOOP.2.1i.63</code></td>
</tr>
<tr>
<td>1.541299</td>
<td>20062500</td>
<td>12</td>
<td>0</td>
<td>12</td>
<td><code>riemann_LOOP.3.1i.64</code></td>
</tr>
<tr>
<td>0.863656</td>
<td>1687500</td>
<td>104</td>
<td>0</td>
<td>108</td>
<td><code>parabola_LOOP.6.1i.67</code></td>
</tr>
</tbody>
</table>
The CCE Program Library (PL)

- **An application wide repository** for compiler and tools information
  - Allows the user to specify a repository of compiler information for an application build

- **Provides the framework for application analysis**
  - Whole application IPA information for optimization
  - Automatic whole application inlining and cloning
  - Various inter-procedural optimizations
  - Whole application static error detection

- **Provides ability for tools to annotate loops with runtime feedback and other performance hints without source change**
  - Support for the Cray refactoring tool, Reveal.
Generate a Program Library

- `> cc -h pl=himeno.pl -hwp* himeno.c`
- `> ftn -h pl=vhone.pl file1.f90`

* Optionally add whole program analysis for additional inlining.
Launch Reveal

● Use with compiler information only (no need to run program):

  > reveal vhone.pl

● Use with compiler + loop work estimates (include performance data):

  > reveal vhone.pl vhone_loops.ap2
Visualize Compiler and Performance Information
Access Cray Compiler Message Information

Access integrated message 'explain' support by right clicking on message.
Navigate Loops through Call Chain
Navigate Code via Compiler Messages

Choose “Compiler Messages” view to access message filtering.

Default filter: Loops that didn’t vectorize. Can select other filters.
View Pseudo Code for Inlined Functions

- Inlined call sites marked
- Expand to see pseudo code
- Search code with Ctrl-F

March 2016
© Cray Inc. Proprietary
Hybridization Step 2: Scope Selected Loop(s)
Review Scoping Results

Loops with scoping information are flagged. Red needs user assistance.

Parallelization inhibitor messages are provided to assist user with analysis.

User addresses issues for variables with FAIL status.
Review Scoping Results (2)

Variable from inlining – hover over ‘I’ to see what symbol means

See where variable came from (@function_name)

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Scope</th>
<th>Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>ar@parabola_1</td>
<td>Scalar</td>
<td>Unresolved</td>
<td>Possible recurrence involving this object.</td>
</tr>
<tr>
<td>da@parabola_1</td>
<td>Scalar</td>
<td>Unresolved</td>
<td>Possible resolvable recurrence involving this object.</td>
</tr>
<tr>
<td>delta@remap_1</td>
<td>Scalar</td>
<td>Unresolved</td>
<td>Possible recurrence involving this object.</td>
</tr>
<tr>
<td>dvol_1</td>
<td>Array</td>
<td>Unresolved</td>
<td>Possible resolvable recurrence involving this object.</td>
</tr>
<tr>
<td>dx</td>
<td>Array</td>
<td>Unresolved</td>
<td>Possible recurrence involving this object.</td>
</tr>
<tr>
<td>dxO</td>
<td>Array</td>
<td>Unresolved</td>
<td>Possible resolvable recurrence involving this object.</td>
</tr>
<tr>
<td>e</td>
<td>Array</td>
<td>Unresolved</td>
<td>Possible recurrence involving this object.</td>
</tr>
</tbody>
</table>

Find Name: 

Insert Directive Show Directive

© Cray Inc. Proprietary

March 2016
Review Scoping Results (3)

Reveal identifies shared reductions down the call chain

Reveal identifies calls that prevent parallelization
Hybridization Step 3: Generate OpenMP Directives

! Directive inserted by Cray Reveal. May be incomplete.

!$OMP parallel do default(none) &
!$OMP& unresolved (dvoldx,dx0,e,f,flat,p,para,q,r,radius,svel,u,v,w, &
!$OMP& xa,xa0) &
!$OMP& private (i,j,k,m,n,$$_n,delp2,delp1,shock,temp2,old_flat, &
!$OMP& onemfl,hdt,sinx0,gamfac1,gamfac2,dtheta,deltx,fractn, &
!$OMP& ekin) &
!$OMP& shared (gamm,isy,js,ks,mpey,ndim,ngeomy,nlefty,npey,nrighty, &
!$OMP& recv1,send2,zdy,zxc,zya)
do k = 1, ks
do i = 1, isy
  radius = zxc(i+mpey*isy)
enddo

! Put state variables into 1D arrays, padding with 6 ghost zones
do m = 1, npey
  do j = 1, js
    n = j + js*(m-1) + 6
  enddo
  r(n) = recv1(1,k,j,i,m)
  p(n) = recv1(2,k,j,i,m)
  u(n) = recv1(4,k,j,i,m)
  v(n) = recv1(5,k,j,i,m)
  w(n) = recv1(3,k,j,i,m)
  f(n) = recv1(6,k,j,i,m)
enddo

endo do j = 1, jmax
  n = j + 6
Reveal generates OpenMP directive with illegal clause marking
variables that need addressing
Or Validate User Inserted Directives

User inserted directive with mis-scoped variable ‘n’
Hybridization Step 4: Performance Analysis

Choose “Compiler Messages” view to access message filtering.

See loops that didn’t vectorize. Can select other filters.

See all compiler messages for a loop nest.
Focus on Relevant Loops (June’16)

Compiler messages sorted by time
Hybridization Step 4: Performance Analysis

Observations and suggestions

D1 cache utilization:
61.7% of total execution time was spent in 1 functions with D1 cache hit ratios below the desirable minimum of 90.0%. Cache utilization might be improved by modifying the alignment or stride of references to data arrays in these functions.

<table>
<thead>
<tr>
<th>D1</th>
<th>Time%</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cache</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ratio</td>
<td></td>
</tr>
<tr>
<td>74.3%</td>
<td>61.7%</td>
<td>calc3_</td>
</tr>
</tbody>
</table>

D1 + D2 cache utilization:
61.7% of total execution time was spent in 1 functions with combined D1 and D2 cache hit ratios below the desirable minimum of 97.0%. Cache utilization might be improved by modifying the alignment or stride of references to data arrays in these functions.

<table>
<thead>
<tr>
<th>D1+D2</th>
<th>Time%</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>cache</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hit</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ratio</td>
<td></td>
</tr>
<tr>
<td>96.6%</td>
<td>61.7%</td>
<td>calc3_</td>
</tr>
</tbody>
</table>
Summary

● Reveal can be used to simplify the task of adding OpenMP to MPI programs.

● The result is performance portable code: OpenMP directives (programs can be built with any compiler that supports OpenMP)

● Can be used as a stepping stone for codes targeted for nodes with higher core counts and as the first step in adding directives to applications to target GPUs

● Moving to OpenMP 4.0 accelerator directives or OpenACC via OpenMP is a good idea
  ● Same work required
  ● Can have both (conditionally compile one or other or none)
  ● First level of debugging on multicore CPU