MPI usage at NERSC: Present and Future

Alice Koniges,
Brandon Cook, Jack Deslippe, Thorston Kurth, Hongzhang Shan

National Energy Research Scientific Computing Center
Lawrence Berkeley National Laboratory, USA

September, 2016
See the results of two user surveys – MPI at NERSC, and MPI Alternatives

NERSC Users by Country

Europe:  
United Kingdom  
Germany  
France  
Italy  
Switzerland  
Spain  
Denmark  
Poland  
Czech Republic  
Russian Federation  
Norway  
Sweden  
Netherlands  
Finland  
Greece  
Portugal  
Belgium  
Cyprus  
Ireland  
Turkey  
Sweden  
Austria  
Serbia  
Ukraine

Cray XC40 with Knights Landing

Cray XC30

For Example:

Blocking MPI_Send / MPI_Recv  
Blocking MPI_Bsend / MPI_Recv  
Blocking MPI_Ssend / MPI_Recv  
Blocking MPI_SendRecv

Nonblocking MPI_Isend and MPI_Recv  
Nonblocking MPI_Send and MPI_Irecv

How do you implement point-to-point communication?

- Blocking MPI_Send / MPI_Recv: 18 (62.1%)
  - Blocking MPI_Bsend / MPI_Recv: 2 (6.9%)
  - Blocking MPI_Ssend / MPI_Recv: 1 (3.4%)
  - Blocking MPI_SendRecv: 5 (17.2%)

- Nonblocking MPI_Isend and MPI_Recv: 14 (48.3%)
  - Nonblocking MPI_Send and MPI_Irecv: 13 (44.8%)

0 2 4 6 8 10 12 14 16 18
See what MPI alternatives people are using and some recent comparisons with MPI

Which parallel programming model(s)/approach(s) do you use?

<table>
<thead>
<tr>
<th>Model/Approach</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>UPC</td>
<td>8 (30.8%)</td>
</tr>
<tr>
<td>UPC++</td>
<td>9 (34.6%)</td>
</tr>
<tr>
<td>Charm++</td>
<td>5 (19.2%)</td>
</tr>
<tr>
<td>Kokkos</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Chapel</td>
<td>1 (3.8%)</td>
</tr>
<tr>
<td>Coarray Fortran</td>
<td>3 (11.5%)</td>
</tr>
<tr>
<td>HPX</td>
<td>-2 (7.7%)</td>
</tr>
<tr>
<td>OCR -based</td>
<td>8 (30.8%)</td>
</tr>
<tr>
<td>Legion</td>
<td>1 (3.8%)</td>
</tr>
<tr>
<td>Parallel Python</td>
<td>3 (11.5%)</td>
</tr>
<tr>
<td>Hadoop/MapReduce</td>
<td>1 (3.8%)</td>
</tr>
<tr>
<td>other</td>
<td>7 (26.9%)</td>
</tr>
</tbody>
</table>

AMR with UPC++ Compared to MPI

MiniGhost with HPX Compared to MPI

(40 variables - 20 timesteps - 200x200x200 - 10% reduction)