On the Impact of Number Representation for High-Order LES

F.D. Witherden and A. Jameson

Department of Aeronautics & Astronautics
Stanford University
Motivation

• LES is expensive...

• ...really expensive.
Computer Arithmetic

- Binary floating point following IEEE 754

- \( x = \text{sign} \cdot \text{mantissa} \cdot 2^{\text{exponent}} \)

<table>
<thead>
<tr>
<th>Format</th>
<th>Sign</th>
<th>Exponent</th>
<th>Mantissa</th>
</tr>
</thead>
<tbody>
<tr>
<td>binary32</td>
<td>1</td>
<td>8</td>
<td>23</td>
</tr>
<tr>
<td>binary64</td>
<td>1</td>
<td>11</td>
<td>52</td>
</tr>
</tbody>
</table>

(sign | exponent | mantissa)
Computer Arithmetic

• Complicated!

• If you think you understand floating point arithmetic—you don’t!
### Why Number Precision?

<table>
<thead>
<tr>
<th>Model</th>
<th>GB/s</th>
<th>Single</th>
<th>Double</th>
<th>Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMD Radeon R9 Nano</td>
<td>512</td>
<td>8.19</td>
<td>0.51</td>
<td>16</td>
</tr>
<tr>
<td>AMD FirePro W9100</td>
<td>320</td>
<td>5.24</td>
<td>2.62</td>
<td>2</td>
</tr>
<tr>
<td>Intel Xeon E5-2699 v4</td>
<td>77</td>
<td>1.55</td>
<td>0.77</td>
<td>2</td>
</tr>
<tr>
<td>Intel Xeon Phi 7120A</td>
<td>352</td>
<td>2.42</td>
<td>1.21</td>
<td>2</td>
</tr>
<tr>
<td>NVIDIA Tesa K40c</td>
<td>288</td>
<td>4.29</td>
<td>1.43</td>
<td>3</td>
</tr>
<tr>
<td>NVIDIA Tesa M40</td>
<td>288</td>
<td>7.00</td>
<td>0.21</td>
<td>32</td>
</tr>
</tbody>
</table>
Potential Speedups

• If a code region is limited by…

  • FLOPs = 2x to 32x

  • Memory bandwidth = 2x

  • Disk I/O = 2x

  • Latency (memory, disk, network, …) = 1x
The Status Quo

- Extensive research in bars indicates that, if given the choice between a single and a double measure, the double wins every time.

- CFD codes are no exception.
Do We Need Double Precision?

• Very little research in the CFD space.

• Results mostly limited to steady state computations where double precision does appear to be necessary.
Methodology

• Rerun several of our previous published test cases using single precision arithmetic.

• Compare the results and assess the performance.
Experiments

- Using PyFR we have evaluated several unsteady viscous test cases.
  - Taylor–Green vortices.
  - Flow over a circular cylinder.
  - Flow over a NACA 0021.
3D Taylor–Green Vortex

- Standard test case for DG.
3D Taylor–Green Vortex

- Four structured grids with roughly constant DOF count.

<table>
<thead>
<tr>
<th>Order</th>
<th>$N_E$</th>
<th>$\sum N_u$</th>
<th>Memory / GiB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = 2$</td>
<td>$86^3$</td>
<td>$258^3$</td>
<td>Single: 6.4, Double: 12.2</td>
</tr>
<tr>
<td>$\phi = 3$</td>
<td>$64^3$</td>
<td>$256^3$</td>
<td>Single: 5.4, Double: 10.3</td>
</tr>
<tr>
<td>$\phi = 4$</td>
<td>$52^3$</td>
<td>$260^3$</td>
<td>Single: 5.1, Double: 9.8</td>
</tr>
<tr>
<td>$\phi = 5$</td>
<td>$43^3$</td>
<td>$258^3$</td>
<td>Single: 4.6, Double: 9.0</td>
</tr>
</tbody>
</table>
3D Taylor–Green Vortex

- Consider kinetic energy decay rate.
- Compare with van Rees et al.
- No difference between single and double.
3D Taylor–Green Vortex

- Performance on a two NVIDIA K40c’s with GiMMiK.

<table>
<thead>
<tr>
<th>Order</th>
<th>GFLOP / stage</th>
<th>$t_w/ \sum N_u / 10^{-9}$ s</th>
<th>GFLOP / s</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi = 2$</td>
<td>$1.84 \times 10^1$</td>
<td>Single: 4.8, Double: 8.9</td>
<td>Single: 222.1, Double: 120.5</td>
<td>1.84</td>
</tr>
<tr>
<td>$\phi = 3$</td>
<td>$1.82 \times 10^1$</td>
<td>Single: 4.2, Double: 7.9</td>
<td>Single: 252.3, Double: 134.6</td>
<td>1.88</td>
</tr>
<tr>
<td>$\phi = 4$</td>
<td>$1.92 \times 10^1$</td>
<td>Single: 4.4, Double: 8.6</td>
<td>Single: 255.9, Double: 129.7</td>
<td>1.97</td>
</tr>
<tr>
<td>$\phi = 5$</td>
<td>$1.96 \times 10^1$</td>
<td>Single: 4.5, Double: 13.1</td>
<td>Single: 250.8, Double: 87.0</td>
<td>2.88</td>
</tr>
</tbody>
</table>
Flow Over a Cylinder
Flow Over a Cylinder

- Cylinder at $Re = 3900$, and $Ma = 0.2$ with $p = 4$.

- Mixed prism/tet grid of span $\pi D$. 
Flow Over a Cylinder

- Pressure coefficient on the surface.
- Compare with Lehmkuhl et al.
Flow Over a Cylinder

• Performance on a single NVIDIA K40c with GiMMiK.

• Tet operator matrices are small and prisms sparse.

• Overall speedup of \(~1.6\).

• Simulation results in heavy indirection; thus experiences less of an improvement from single precision.
NACA 0021

• Flow over a NACA 0021 at 60 degree AoA.

• $\text{Re} = 270,000$ and $\text{Ma} = 0.1$.

• Compare with experimental results of Swalwell.
NACA 0021

- **206,528** hexahedral elements.
- Span is four times the chord.
- Fourth order solution polynomials with **full anti-aliasing**.
NACA 0021

• Performance on 16 NVIDIA K80’s (32 GPUs).

• All operators are dense.

• Near the limit of strong scaling.

• Overall speedup of $\sim1.8$. 
Remarks and Closing Thoughts

For LES single precision is sufficient.
Acknowledgements

• Air Force Office of Scientific Research for their support under grant FA9550-14-1-0186.

• Margot Gerritsen for access to XStream.

• NVIDIA for hardware donations.