Towards a Unified CPU-GPU code hybridization: A GPU Based Optimization Strategy Efficient on Other Modern Architectures
Ludomir Oteski, Guillaume Colin de Verdiere, Sylvain Contassot-Vivier, Stephane Vialle, Juliette Ryan

To cite this version:
Ludomir Oteski, Guillaume Colin de Verdiere, Sylvain Contassot-Vivier, Stephane Vialle, Juliette Ryan. Towards a Unified CPU-GPU code hybridization: A GPU Based Optimization Strategy Efficient on Other Modern Architectures. Parallel Computing Conference (Parco 2017), Sep 2017, Bologne, Italy. hal-01651961

HAL Id: hal-01651961
https://hal-centralesupelec.archives-ouvertes.fr/hal-01651961
Submitted on 29 Nov 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Towards a Unified CPU-GPU code hybridization: A GPU Based Optimization Strategy Efficient on Other Modern Architectures

L. Oteski, G. Colin de Verdière, S. Contassot-Vivier, S. Vialle, J. Ryan

Acks.: CEA/DIF, IDRIS, GENCI, NVIDIA, Région Lorraine
Introduction and context

Most scientific simulations
• Require intensive computations
• Are frequently based on parallel iterative processes

Classical development process:
1. Sequential (optimized) version
2. CPU multi-threaded version
3. GPU version + multi-machines version (if needed)

We propose to invert the parallel development process:
• Begin with the GPU version
• To rapidly derive efficient CPU versions: vectors + threads (+ MPI)
Test case application: Discontinuous Galerkin schemes for Computational Fluid Dynamics

Representative of some compute-intensive applications
We consider the 2D time-dependant compressible Navier-Stokes equations

Example:
Yee's vortex with the full NS equations

Numerical method:
• **Space**: 2nd order Discontinuous Galerkin,
• **Time**: 3rd order TVD Runge-Kutta,
• **Boundaries**: X and Y-periodic.
Test case application: Discontinuous Galerkin schemes for Computational Fluid Dynamics

Representative of some compute-intensive applications
We consider the 2D time-dependant compressible Navier-Stokes equations

→ Storage space per array of variables: \( N_{\text{coefs}} \times N_{\text{eq}} \times N_{\text{cells}} \)
  \( N_{\text{coefs}} \): number of polynomial coefficients (6 for 2D second order polynomials)
  \( N_{\text{eq}} \): number of equations (4 in our case)
  \( N_{\text{cells}} = N_x \times N_y \): number of cells in the x and y directions

→ 3 double precision arrays: current state, previous time-step and time-derivatives

→ Memory cost \( \approx 3 \times (6 \times 4 \times N_{\text{cells}}) \times 8 \) Bytes
  Ex : \( 2001 \times 2001 \) mesh \( \geq 2.3 \) GBytes
  \( 2731 \times 2731 \) mesh \( \geq 4.3 \) GBytes
Unified Development Approach

GPU and CPU have different architectures, **BUT**:

Today both use **vectorization**:

- *Vectors* for CPU
- *Warps* for GPU

→ Efficient computing scheme on GPU should provide:

- Compliant CPU scheme with minor adaptations
- Efficient CPU execution

**Optimized GPU programming:**

- is low-level (« close to the metal »)
- low-level API (CUDA, OpenCL) allows accurate ctrl of computations

→ Clear impact of optimization attempts on GPU
→ Interest of using GPU as the first development step
Unified Development Approach

Example of GPU kernel translated into a vectorized CPU function

2D Grid of blocks of CUDA threads $\rightarrow$ CPU nested loops

<table>
<thead>
<tr>
<th>2D CUDA-GPU code</th>
<th>2D CPU code (vectorized with icc)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 void <strong>global</strong> function(...)</td>
<td>1 void function(...)</td>
</tr>
<tr>
<td>2 {</td>
<td>2</td>
</tr>
<tr>
<td>3 //Thread index in x-direction</td>
<td>3 //Loop on y-direction</td>
</tr>
<tr>
<td>4 int tidx = ...;</td>
<td>4 #pragma omp for</td>
</tr>
<tr>
<td>5 //Thread index in y-direction</td>
<td>5 for(tidy=0; tidy&lt;Ny; tidy++)</td>
</tr>
<tr>
<td>6 int tidy = ...;</td>
<td>6</td>
</tr>
<tr>
<td>7 //Logical condition</td>
<td>7 //Loop on x-direction</td>
</tr>
<tr>
<td>8 //to make computations</td>
<td>8 #pragma simd</td>
</tr>
<tr>
<td>9 if(tidx&lt;Nx &amp;&amp; tidy&lt;Ny)</td>
<td>9 for(tidx=0; tidx&lt;Nx; tidx++)</td>
</tr>
<tr>
<td>10 {</td>
<td>10</td>
</tr>
<tr>
<td>11 ... //Copy to CPU code</td>
<td>11 [...] //Insert from GPU code</td>
</tr>
<tr>
<td>12 }</td>
<td>12</td>
</tr>
<tr>
<td>13 }</td>
<td>13</td>
</tr>
<tr>
<td>14 }</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>
Optimization guideline overview

4 crucial GPU optimizations applied to the CPU version
1. Reduce accesses to distant memories
2. Merge kernels which share the same memory patterns
3. Simplify and factor computations
4. Align data (including \textit{tiling: CPU specific optimization})

+ CPU specific optimizations
  a. Vectorization tuning
  b. Data locality improvement
Optimization guideline: GPU → CPU (1)

1. Reduce accesses to distant memories
   Transfer redundant accesses to global/DRAM memory into registers
   → GPU avoids unnecessary accesses to global memory
   → CPU improves data cache locality

   __global__
   void func(double *results, int N) {
       int tidx = threadIdx.x;
       ...
       for(int i=0; i<N; i++) {
           ...
           results[tidx] += ...
       }
   }

   __global__
   void func(double *__restrict__ results, int N) {
       int tidx = threadIdx.x;
       ...
       double loc_result = results[tidx];
       for(int i=0; i<N; i++) {
           ...
           loc_result += ...
       }
       results[tidx] = loc_result;
   }

   → Only 1 read and 1 write to results[] array

Ex: GPU perf ×3.97, CPU perf ×1.36
Optimization guideline: GPU → CPU (2)

2. Merge kernels which share the same memory patterns
   → GPU limits its number of accesses to distant memories by improving the re-use of distant variables
   → CPU increases its cache re-use

1. Compute the time-step
2. For s Runge-Kutta step:
   a. Convective fluxes in X
   b. Convective fluxes in Y
   c. Convective integral
   d. Compute local viscosity
   e. Viscous fluxes in X
   f. Viscous fluxes in Y
   g. Viscous integral
   h. Runge-Kutta propagation

Ex: GPU perf ×1.75, CPU perf ×1.37
Optimization guideline: GPU $\rightarrow$ CPU (3)

3. Simplify and factor computations
   • Suppress non essential variables
   • Store recurrent computations in intermediate vars (mult by const, const square roots,...)
   • Control whether or not a static loop should be unrolled

```c
#pragma unroll
for(int i=0; i<SIZE; i++) {
  ...
  double val = ...
  double c0 = 1/(sqrt(0.5)*val)*...
  double c1 = 1/(3*val)*...
  ...
}
```

$\rightarrow$ The most time-consuming development step

```c
double isqrt0p5 = 1/sqrt(0.5);
double inv3 = 1/3;
#pragma unroll //Keep it ?
for(int i=0; i<SIZE; i++) {
  ...
  double val = ...
  double ival = 1/val;
  double c0 = isqrt0p5*ival*...
  double c1 = inv3*ival*...
  ...
}
```

Ex: GPU perf $\times 1.21$, CPU perf $\times 1.82$
Optimization guideline: GPU $\rightarrow$ CPU (4)

4. Align data

- GPU: align data access on warp size: 32 (static size)
  $\rightarrow$ ensures coalescent memory accesses
  $\rightarrow$ GPU perf $\times 1.03$ on our problem

- CPU:
  1. Add a tiling algorithm inside each thread computations
     $\rightarrow$ CPU perf $\times 1.23$
  2. Use static tile size
     $\rightarrow$ CPU perf $\times 1.19$

Static tile size $\rightarrow$ static nb of loop iterations $\rightarrow$ better vectorization
Optimization guideline: CPU specific (1)

a. Vectorization tuning

Previous optimizations steps tend to create huge vectorized loops!
→ high increase of register pressure
→ not suitable for CPU vectorization...

So ... we adjusted our CPU vectorization strategy
1. by splitting SIMD loops into smaller loops
2. #pragma simd → #pragma vector always
   tells the compiler to perform auto-vectorization if the loop does not carry dependencies

→ CPU perf ×1.03

Compromise distant memory accesses reduction / register pressure for optimal CPU vectorization
Optimization guideline: CPU specific (2)

b. Data locality improvement

MPI-OpenMP version of the CPU code → to improve data locality

- Memory locations of data used in (only) one MPI process and its threads will be close to their associated cores
- This optimization is well suited to NUMA architectures (and many machines are becoming NUMA machines!)

→ CPU perf ×1.03

The domain decomposition has been performed along the y axis by using a ghost-cell technique to share boundaries of neighbouring domains between MPI processes
## Experimental steps & performances (1)

<table>
<thead>
<tr>
<th>Code versions</th>
<th>CPU E5-1650v3 (6th)</th>
<th>Speedup vs seq. CPU</th>
<th>Progressive speedup</th>
<th>GPU K20Xm</th>
<th>Speedup vs seq. CPU</th>
<th>Progressive speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq. CPU v (1th)</td>
<td>8.35 x10⁴ cus</td>
<td>1.00</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CPU v1 (OMP+simd pragmas)</td>
<td>0.59 x10⁶ cus</td>
<td>7.01</td>
<td>7.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPU v1 (CUDA)</td>
<td></td>
<td></td>
<td></td>
<td>1.44 x10⁶ cus</td>
<td>17.21</td>
<td>17.21</td>
</tr>
</tbody>
</table>

**Note:**
- cus: Cell Update/s
- 2001 × 2001 cells
- 100 steps
### Experimental steps & performances (2)

<table>
<thead>
<tr>
<th>Code versions</th>
<th>CPU E5-1650v3 (6th)</th>
<th>Speedup vs seq. CPU</th>
<th>Progressive speedup</th>
<th>GPU K20Xm</th>
<th>Speedup vs seq. CPU</th>
<th>Progressive speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq. CPU v (1th)</td>
<td>8.35 x10⁴ cus</td>
<td>1.00</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CPU v1 (OMP+simd pragmas)</strong></td>
<td>0.59 x10⁶ cus</td>
<td>7.01</td>
<td>7.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GPU v1 (CUDA)</strong></td>
<td></td>
<td></td>
<td></td>
<td>1.44 x10⁶ cus</td>
<td>17.21</td>
<td>17.21</td>
</tr>
<tr>
<td>Reduced accesses to distant mem.</td>
<td>0.81 x10⁶ cus</td>
<td>9.63</td>
<td>1.36</td>
<td>5.71 x10⁶ cus</td>
<td>68.36</td>
<td>3.97</td>
</tr>
<tr>
<td>Merged funcs with same mem pattern</td>
<td>1.09 x10⁶ cus</td>
<td>13.05</td>
<td>1.37</td>
<td>10.0 x10⁶ cus</td>
<td>119.62</td>
<td>1.85</td>
</tr>
<tr>
<td>Simplification of computations</td>
<td>2.01 x10⁶ cus</td>
<td>24.05</td>
<td>1.82</td>
<td>12.1 x10⁶ cus</td>
<td>145.00</td>
<td>1.21</td>
</tr>
<tr>
<td><strong>CPU Only: Tiling (cache optim)</strong></td>
<td>2.48 x10⁶ cus</td>
<td>29.72</td>
<td>1.23</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Data alignment</td>
<td>2.86 x10⁶ cus</td>
<td>34.25</td>
<td>1.19</td>
<td>12.5 x10⁶ cus</td>
<td>149.53</td>
<td>1.03</td>
</tr>
</tbody>
</table>

**cus:** Cell Update/s
2001 × 2001 cells
100 steps
# Experimental steps & performances (3)

<table>
<thead>
<tr>
<th>Code versions</th>
<th>CPU E5-1650v3 (6th)</th>
<th>Speedup vs seq. CPU</th>
<th>Progressive speedup</th>
<th>GPU K20Xm</th>
<th>Speedup vs seq. CPU</th>
<th>Progressive speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Seq. CPU v (1th)</strong></td>
<td>8.35 $\times 10^4$ cus</td>
<td>1.00</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CPU v1 (OMP+simd pragmas)</strong></td>
<td>0.59 $\times 10^6$ cus</td>
<td>7.01</td>
<td>7.01</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GPU v1 (CUDA)</strong></td>
<td></td>
<td></td>
<td>1.44 $\times 10^6$ cus</td>
<td>17.21</td>
<td>17.21</td>
<td></td>
</tr>
<tr>
<td>Reduced accesses to distant mem.</td>
<td>0.81 $\times 10^6$ cus</td>
<td>9.63</td>
<td>1.36 $\times 10^6$ cus</td>
<td>68.36</td>
<td>3.97</td>
<td></td>
</tr>
<tr>
<td>Merged funcs with same mem pattern</td>
<td>1.09 $\times 10^6$ cus</td>
<td>13.05</td>
<td>1.37 $\times 10^6$ cus</td>
<td>119.62</td>
<td>1.85</td>
<td></td>
</tr>
<tr>
<td>Simplification of computations</td>
<td>2.01 $\times 10^6$ cus</td>
<td>24.05</td>
<td>1.82 $\times 10^6$ cus</td>
<td>145.00</td>
<td>1.21</td>
<td></td>
</tr>
<tr>
<td><strong>CPU Only: Tiling (cache optim)</strong></td>
<td>2.48 $\times 10^6$ cus</td>
<td>29.72</td>
<td>1.23 $\times 10^6$ cus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Data alignment</td>
<td>2.86 $\times 10^6$ cus</td>
<td>34.25</td>
<td>1.19 $\times 10^6$ cus</td>
<td>149.53</td>
<td>1.03</td>
<td></td>
</tr>
<tr>
<td><strong>CPU Only: Tuning on vectorization</strong></td>
<td>2.96 $\times 10^6$ cus</td>
<td>35.44</td>
<td>1.03 $\times 10^6$ cus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>CPU Only: MPI + OMP (3 proc x 2th)</strong></td>
<td>3.05 $\times 10^6$ cus</td>
<td>36.52</td>
<td>1.03 $\times 10^6$ cus</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

**Cus:** Cell Update/s 2001 × 2001 cells 100 steps
Experimental testbed

Hardware:
- Bi-socket E5-2698v3
- Bi-socket E5-2680v4
- KNL 7250
- K20Xm
- K40
- P100 540GB/s
- P100 720GB/s

Application:
- 2D Discontinuous Galerkin (NS equations)
- Grid size: 2731×2731

Operating System:
- Linux

Compilers:
- CUDA 8
- ICC

ICC options:
- -O3
- -march=native
- -fma
- -align
### Experimental final performances (1)

<table>
<thead>
<tr>
<th></th>
<th>CPU</th>
<th>GPU</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNL 7250</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16GB MCDRAM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(~500 GB/s)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

→ a little bit faster than GPU K20Xm and GPU K40

Applying our optimisation guideline

<table>
<thead>
<tr>
<th></th>
<th>cus/Watts* ($\times 10^4$)</th>
<th>cus ($\times 10^6$)[OpenMP+MPI]</th>
<th>cus ($\times 10^6$)[CUDA]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>45</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>2xE52680v4</td>
<td>2xE52698v3</td>
<td>Knl 7250</td>
<td></td>
</tr>
<tr>
<td>DDR4</td>
<td>DDR4</td>
<td>MCDRAM (flat)</td>
<td></td>
</tr>
<tr>
<td>2xKML</td>
<td>2xKML</td>
<td>MCDRAM (cache)</td>
<td></td>
</tr>
<tr>
<td>K20Xm</td>
<td>K20Xm</td>
<td>GDDR5</td>
<td></td>
</tr>
<tr>
<td>K40</td>
<td>K40</td>
<td>GDDR5</td>
<td></td>
</tr>
</tbody>
</table>
Experimental final performances (2)

But GPU P100 great winner!!

CPU

GPU

P100

P100 720GB/s
HBM2, NVLINK

P100 540GB/s
HBM2

cus/Watts* \((\times 10^4)\)
cus \((\times 10^6)\) [OpenMP+MPI]
cus \((\times 10^6)\) [CUDA]
Experimental single-threaded cache roofline profile

Perf (GFLOPS) on one E5-2680v4 core (measured with Intel Advisor 2017)

Arithmetic Intensity (FLOPS/Byte)

Global application

Time consuming loops
Above L2 BdW limit (good)

Storage function (not significant)
Foundation of Hybrid solution

GPU
A large block of $n_{th}$ light threads

An array of $n$ data

$n = n_{th} \times VSIZ$
Ex: $12 = 12 \times 1$

CPU
One CPU thread with VSIZ vector units

An array of $n$ data

$n = n_{th} \times n_{steps} \times VSIZ$
Ex: $12 = 1 \times 3 \times 4$
Hybrid implementation

**Common src code for CPU & GPU**

<table>
<thead>
<tr>
<th>2D CUDA-GPU code (GPU.cu)</th>
<th>Common vectorized kernel (kernel.h)</th>
<th>2D CPU code (CPU.cpp)</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>#define VSIZ 1</code></td>
<td><code>template&lt;const int VSIZ&gt;</code></td>
<td><code>#define VSIZ 32</code></td>
</tr>
<tr>
<td><code>//Include the kernel</code></td>
<td><code>//Conditional compilation</code></td>
<td><code>//Include the kernel</code></td>
</tr>
<tr>
<td><code>#define DEFGPU</code></td>
<td><code>#ifdef DEFGPU</code></td>
<td><code>#include &quot;kernel.h&quot;</code></td>
</tr>
<tr>
<td><code>#include &quot;kernel.h&quot;</code></td>
<td><code>__device__</code></td>
<td><code>void cpu_function</code></td>
</tr>
<tr>
<td></td>
<td><code>inline void kernel()</code></td>
<td><code>{</code></td>
</tr>
<tr>
<td>`void <strong>global</strong> gpu_function(</td>
<td><code>double *__restrict__ val)</code></td>
<td><code>if(tidx&lt;Nx)</code></td>
</tr>
<tr>
<td>double *<strong>restrict</strong> val)`</td>
<td><code>const int tidx)</code></td>
<td><code>for(int tidx=0;</code></td>
</tr>
<tr>
<td></td>
<td><code>//Vectorized loop</code></td>
<td><code>tidx&lt;Nx;</code></td>
</tr>
<tr>
<td></td>
<td><code>#pragma vector always</code></td>
<td><code>tidx+=VSIZ)</code></td>
</tr>
<tr>
<td></td>
<td><code>#pragma unroll</code></td>
<td><code>}</code></td>
</tr>
<tr>
<td>6</td>
<td><code>for(vec=0; vec&lt;VSIZ; vec++)</code></td>
<td><code>kernel&lt;VSIZ&gt;(val,tidx);</code></td>
</tr>
<tr>
<td>7</td>
<td><code>{</code></td>
<td><code>}</code></td>
</tr>
<tr>
<td>8</td>
<td><code>...</code></td>
<td><code>val[VSIZ+tidx+vec] = ...;</code></td>
</tr>
<tr>
<td>15</td>
<td><code>}</code></td>
<td><code>}</code></td>
</tr>
</tbody>
</table>

- **C++ template** code, parametered with the nb of elements processed by each thread
- **Compiler directives** (ignored when not using the right compiler and architecture)
Hybrid system

MPi Process 0

GPU 0

GPU 1

CPU 0

CPU 1

CPU 2

CPU 3

CPU 4

CPU 5

MPi Process 1

MPi Process 2

MPi Process 3
Sketch of hybridized code

Parco - September 2017 – Bologna, Italy
## Experimental Hybrid Performance

<table>
<thead>
<tr>
<th>Devices</th>
<th>CPU run: 2xE5-2680v4</th>
<th>GPU run: P100 (540GB/s)</th>
<th>Hybrid run: 2xE5-2680v4 + P100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perf (x10^6 cus)</td>
<td>10.4</td>
<td>37.0</td>
<td>46.0</td>
</tr>
</tbody>
</table>

Maximum perf: $10.4 + 37.0 = 47.4$

Close to ideal (hybrid) performances
Conclusion & Perspective

• General programming guideline for modern computing devices
  CPU, NUMA CPU, GPU, Xeon-Phi KNL

• Highly efficient with the Discontinuous Galerkin problem
  Should also be efficient on other highly vectorizable problems

• Common (hybrid) computing kernel for all devices
  Inserted in a hybrid source code (for CPU+GPU machines)

→ GPU development also valuable for other (vector) targets
→ Increases the interest of developing first on GPU

Perspective:
  (Half-) Automatic code generation for CPU from the GPU one
Towards a Unified CPU-GPU code hybridization: A GPU Based Optimization Strategy Efficient on Other Modern Architectures

Questions?

We propose to invert the parallel development process:
• Begin with the GPU version
• To rapidly derive efficient CPU versions: vectors + threads (+ MPI)