The Open Computing Language (OpenCL)

OpenCL consists of two parts: a C/C++-callable API and a C-ish programming language.

- The OpenCL programming language can run on NVIDIA GPUs, AMD GPUs, Intel CPUs, Intel GPUs, mobile devices, and (supposedly) FPGAs (Field-Programmable Gate Arrays).
- But, OpenCL is at its best on compute devices with large amounts of **data parallelism**, which usually implies GPU usage.
- You break your computational problem up into lots and lots of small pieces. Each piece gets farmed out to threads on the GPU.
- Each thread wakes up and is able to ask questions about where it lives in the entire collection of (thousands of) threads. From that, it can tell what it is supposed to be working on.
- OpenCL can share data, and interoperate, with OpenGL.
- There is a JavaScript implementation of OpenCL, called WebCL.
- There is a JavaScript implementation of OpenGL, called WebGL.
- WebCL can share data, and interoperate, with WebGL.
- The GPU does not have a stack, and so the OpenCL C-ish programming language cannot do recursion and cannot make function calls. It also can’t use pointers.

The Khronos Group

http://www.khronos.org/opencl/
http://en.wikipedia.org/wiki/OpenCL

Who is Part of the Khronos Group?
Example of using OpenCL in a System-on-a-Chip: Qualcomm Node – Full Linux and OpenCL

OpenCL wants you to break the problem up into Pieces

If you were writing in C/C++, you would say:

```c
void ArrayMult( int n, float *a, float *b, float *c)
{
    for ( int i = 0;  i < n;  i++ )
        c[i] = a[i] * b[i];
}
```

If you were writing in OpenCL, you would say:

```cl
kernel void ArrayMult( global float *dA, global float *dB, global float *dC)
{
    int gid = get_global_id ( 0 );
}
```

Think of this as having an implied for-loop around it, looping through all possible values of gid.

The OpenCL Language also supports Vector Parallelism

OpenCL code can be vector-oriented, meaning that it can perform a single instruction on multiple data values at the same time (SIMD).

Vector data types are: char, int, float, where \( n = 2, 4, 8, \) or 16.

```cl
float4 f, g;
f = (float4)( 1.f, 2.f, 3.f, 4.f );
float16 a16, x16, y16, z16;
f.x = 0.;
f.xy = g.zw;
x16.s89ab = f;
float16 a16 = x16 * y16 + z16;
```

(Note: just because the language supports it, doesn’t mean the hardware does.)
Compute Units and Processing Elements are Arranged in Grids

A GPU **Device** is organized as a grid of **Compute Units**.

Each **Compute Unit** is organized as a grid of **Processing Elements**.

So in NVIDIA terms, their Turing GPU has 68 **Compute Units**, each of which has 64 **Processing Elements**, for a grand total of 4,352 **Processing Elements**.

---

Work-Groups are Arranged in Grids

- The GPU’s workload is divided into a **Grid of Work-Groups**
- Each Block’s workload is divided into a **Grid of Work-Items**

---

OpenCL Software Terminology:

**Work-Groups and Work-Items are Arranged in Grids**

An OpenCL program is organized as a grid of **Work-Groups**.

Each **Work-Group** is organized as a grid of **Work-Items**.

In terms of hardware, a Work-Group runs on a Compute Unit and a Work-Item runs on a Processing Element (PE).

One thread is assigned to each Work-Item.

Threads are swapped on and off the PEs.

---

OpenCL Memory Model

- **Kernel**
  - **Global Memory**
  - **Constant Memory**
- **WorkGroup**
  - **Shared Memory**
OpenCL Error Codes

- CL_SUCCESS
- CL_DEVICE_NOT_FOUND
- CL_DEVICE_NOT_AVAILABLE
- CL_COMPILER_NOT_AVAILABLE
- CL_MEM_OBJECT_ALLOCATION_FAILURE
- CL_OUT_OF_RESOURCES
- CL_OUT_OF_HOST_MEMORY
- CL_PROFILING_INFO_NOT_AVAILABLE
- CL_IMAGE_FORMAT_MISMATCH
- CL_IMAGE_FORMAT_NOT_SUPPORTED
- CL_BUILD_PROGRAM_FAILURE
- CL_MAP_FAILURE
- CL_INVALID_VALUE
- CL_INVALID_DEVICE_TYPE
- CL_INVALID_PLATFORM
- CL_INVALID_CONTEXT
- CL_INVALID_QUEUE_PROPERTIES
- CL_INVALID_BINARY
- CL_INVALID_BUILD_OPTIONS
- CL_INVALID_SAMPLER
- CL_INVALID_PROGRAM
- CL_INVALID_PROGRAM_EXECUTABLE
- CL_INVALID_KERNEL_NAME
- CL_INVALID_KERNEL_DEFINITION
- CL_INVALID_KERNEL
- CL_INVALID_ARG_INDEX
- CL_INVALID_ARG_VALUE
- CL_INVALID_ARG_SIZE
- CL_INVALID_WORK_DIMENSION
- CL_INVALID_MIP_LEVEL
- CL_INVALID_GLOBAL_WORK_SIZE

This way of querying information is a recurring OpenCL pattern (get used to it):

```c
cl_uint numPlatforms;
status = clGetPlatformIDs( 0, NULL, &numPlatforms );
if( status != CL_SUCCESS )
    printf( stderr, "clGetPlatformIDs failed (1)/n" );
    fprintf( stderr, "Number of Platforms = %d
", numPlatforms );
cl_platform_id * platforms = new cl_platform_id[ numPlatforms ];
status = clGetPlatformIDs( numPlatforms, platforms, NULL );
if( status != CL_SUCCESS )
    fprintf( stderr, "clGetPlatformIDs failed (2)/n" );
```

A Way to Print OpenCL Error Codes – get this from the Class Announcements

```c
struct errorcode
{
    cl_int statusCode;
    char * meaning;
}
ErrorCodes[] =
    { CL_SUCCESS,                                        " "                                                },
    { CL_DEVICE_NOT_FOUND,                    "Device Not Found"                    },
    { CL_DEVICE_NOT_AVAILABLE,              "Device Not Available"                },
    . . .
    { CL_INVALID_MIP_LEVEL,                       "Invalid MIP Level"                     },
    { CL_INVALID_GLOBAL_WORK_SIZE,      "Invalid Global Work Size"         },
};
void
PrintCLError( cl_int errorCode,  char * prefix,  FILE *fp )
{
    const int numErrorCodes = sizeof( ErrorCodes ) / sizeof( struct errorcode );
    char * meaning = " ";
    for( int i = 0; i < numErrorCodes; i++ )
    {
        if( errorCode == ErrorCodes[i].statusCode )
        {
            meaning = ErrorCodes[i].meaning;
            break;
        }
    }
    fprintf( fp, "%s %s
", prefix, meaning );
}
```

Querying the Number of Platforms (usually one)
// find out how many devices are attached to each platform and get their ids:
status = clGetDeviceIDs(platform, CL_DEVICE_TYPE_ALL, 0, NULL, &numDevices);
devices = new cl_device_id[numDevices];
status = clGetDeviceIDs(platform, CL_DEVICE_TYPE_ALL, numDevices, devices, NULL);

// find out how many platforms are attached here and get their ids:
cl_uint numPlatforms;
status = clGetPlatformIDs(0, NULL, &numPlatforms);
if (status != CL_SUCCESS)
fprintf(stderr, "clGetPlatformIDs failed (1)\n");
fprintf(OUTPUT, "Number of Platforms = %d\n", numPlatforms);
cl_platform_id *platforms = new cl_platform_id[numPlatforms];
status = clGetPlatformIDs(numPlatforms, platforms, NULL);
if (status != CL_SUCCESS)
fprintf(stderr, "clGetPlatformIDs failed (2)\n");

cl_uint numDevices;
cl_device_id *devices;
for (int i = 0; i < (int)numPlatforms; i++)
{
fprintf(OUTPUT, "Platform #%d:\n", i);
size_t size;
char *str;
clGetPlatformInfo(platforms[i], CL_PLATFORM_NAME, 0, NULL, &size);
str = new char[size];
clGetPlatformInfo(platforms[i], CL_PLATFORM_NAME, size, str, NULL);
fprintf(OUTPUT, "\tName = '%s'\n", str);
delete[] str;
clGetPlatformInfo(platforms[i], CL_PLATFORM_VENDOR, 0, NULL, &size);
str = new char[size];
clGetPlatformInfo(platforms[i], CL_PLATFORM_VENDOR, size, str, NULL);
fprintf(OUTPUT, "\tVendor = '%s'\n", str);
delete[] str;
clGetPlatformInfo(platforms[i], CL_PLATFORM_VERSION, 0, NULL, &size);
str = new char[size];
clGetPlatformInfo(platforms[i], CL_PLATFORM_VERSION, size, str, NULL);
fprintf(OUTPUT, "\tVersion = '%s'\n", str);
delete[] str;
clGetPlatformInfo(platforms[i], CL_PLATFORM_PROFILE, 0, NULL, &size);
str = new char[size];
clGetPlatformInfo(platforms[i], CL_PLATFORM_PROFILE, size, str, NULL);
fprintf(OUTPUT, "\tProfile = '%s'\n", str);
delete[] str);

// find out how many devices are attached to each platform and get their ids:
status = clGetDeviceIDs(platforms[i], CL_DEVICE_TYPE_ALL, 0, NULL, &numDevices);
if (status != CL_SUCCESS)
fprintf(stderr, "clGetDeviceIDs failed (2)\n");
devices = new cl_device_id[numDevices];
status = clGetDeviceIDs(platforms[i], CL_DEVICE_TYPE_ALL, numDevices, devices, NULL);

for (int j = 0; j < (int)numDevices; j++)
{
fprintf(OUTPUT, "\tDevice #%d:\n", j);
size_t size;
cl_device_type type;
cl_uint ui;
sizes[3] = { 0, 0, 0, 0 };
clGetDeviceInfo(devices[j], CL_DEVICE_TYPE, sizeof(type), &type, NULL);
switch (type)
{
case CL_DEVICE_TYPE_CPU:
fprintf(OUTPUT, "CL_DEVICE_TYPE_CPU\n");
break;
case CL_DEVICE_TYPE_GPU:
fprintf(OUTPUT, "CL_DEVICE_TYPE_GPU\n");
break;
case CL_DEVICE_TYPE_ACCELERATOR:
fprintf(OUTPUT, "CL_DEVICE_TYPE_ACCELERATOR\n");
break;
default:
fprintf(OUTPUT, "Other...\n");
break;
}
clGetDeviceInfo(devices[j], CL_DEVICE_VENDOR_ID, sizeof(ui), &ui, NULL);
fprintf(OUTPUT, "\tDevice Vendor ID = 0x%04x\n", ui);
clGetDeviceInfo(devices[j], CL_DEVICE_MAX_COMPUTE_UNITS, sizeof(ui), &ui, NULL);
fprintf(OUTPUT, "\tDevice Maximum Compute Units = %d\n", ui);
clGetDeviceInfo(devices[j], CL_DEVICE_MAX_WORK_ITEM_DIMENSIONS, sizeof(ui), &ui, NULL);
fprintf(OUTPUT, "\tDevice Maximum Work Item Dimensions = %d x %d x %d\n", ui, ui, ui);
clGetDeviceInfo(devices[j], CL_DEVICE_MAX_WORK_GROUP_SIZE, sizeof(size), &size, NULL);
fprintf(OUTPUT, "\tDevice Maximum Work Group Size = %d\n", size);
clGetDeviceInfo(devices[j], CL_DEVICE_MAX_CLOCK_FREQUENCY, sizeof(size), &size, NULL);
fprintf(OUTPUT, "\tDevice Maximum Clock Frequency = %d MHz\n", size);"}
Number of Platforms = 1
Platform #0:
  Name = 'NVIDIA CUDA'
  Vendor = 'NVIDIA Corporation'
  Version = 'OpenCL 1.1 CUDA 4.1.1'
  Profile = 'FULL_PROFILE'
Device #0:
  Type = 0x0004 = CL_DEVICE_TYPE_GPU
  Device Vendor ID = 0x10de
  Device Maximum Compute Units = 15
  Device Maximum Work Item Dimensions = 3
  Device Maximum Work Item Sizes = 1024 x 1024 x 64
  Device Maximum Work Group Size = 1024
  Kernel Maximum Work Group Size = 1024
  Kernel Compile Work Group Size = 0 x 0 x 0
  Kernel Local Memory Size = 0

Typical Values from Querying the Device

size_t extensionSize;
clGetDeviceInfo( device, CL_DEVICE_EXTENSIONS, 0, NULL, &extensionSize);
char *extensions = new char [extensionSize];
clGetDeviceInfo( device, CL_DEVICE_EXTENSIONS, extensionSize, extensions, NULL);
fprintf( stderr, "Device Extensions:
" );
for( int i = 0; i < (int)strlen(extensions); i++ )
{
  if( extensions[ i ] == ' ' )
    extensions[ i ] = '
';
} fprintf( stderr, "%s
", extensions );
delete [] extensions;

Querying to see what extensions are supported on this device

This is the big one you are looking for. It shows that this OpenCL system can interoperate with OpenGL.

This one is handy too. It shows that this OpenCL system can support 64-bit floating point (i.e., double precision).

Steps in Creating and Running an OpenCL program

1. Program header
2. Allocate the host memory buffers
3. Create an OpenCL context
4. Create an OpenCL command queue
5. Allocate the device memory buffers
6. Write the data from the host buffers to the device buffers
7. Read the kernel code from a file
8. Compile and link the kernel code
9. Create the kernel object
10. Setup the arguments to the kernel object
11. Enqueue the kernel object for execution
12. Read the results buffer back from the device to the host
13. Clean everything up
# include <stdio.h>
# include <math.h>
# include <string.h>
# include <stdlib.h>
# include <omp.h> // for timing

1. .cpp Program Header

```c
// allocate the host memory buffers:
float * hA = new float[NUM_ELEMENTS];
float * hB = new float[NUM_ELEMENTS];
float * hC = new float[NUM_ELEMENTS];
```

2. Allocate the Host Memory Buffers

This could have also been done like this:

```c
float hA[NUM_ELEMENTS];
```

Global memory and the heap typically have lots more space than the stack does. So, typically, you do not want to allocate a large array like this as a local variable.

(Here, it's being done on the heap. It could also have been done in global memory.)

3. Create an OpenCL Context

```c
cl_context context = clCreateContext(NULL, 1, &device, NULL, NULL, &status);
```

4. Create an OpenCL Command Queue

```c
cl_command_queue cmdQueue = clCreateCommandQueue(context, device, 0, &status);
```
5. Allocate the Device Memory Buffers

- Allocate memory buffers on the device:
  
  ```
  cl_mem dA = clCreateBuffer( context, CL_MEM_READ_ONLY, dataSize, NULL, &status );
  cl_mem dB = clCreateBuffer( context, CL_MEM_READ_ONLY, dataSize, NULL, &status );
  cl_mem dC = clCreateBuffer( context, CL_MEM_WRITE_ONLY, dataSize, NULL, &status );
  ```

6. Write the Data from the Host Buffers to the Device Buffers

- Enqueue the two commands to write data into device buffers:
  
  ```
  status = clEnqueueWriteBuffer( cmdQueue, dA, CL_FALSE, 0, dataSize, hA, 0, NULL, NULL );
  status = clEnqueueWriteBuffer( cmdQueue, dB, CL_FALSE, 0, dataSize, hB, 0, NULL, NULL );
  ```

- The read and write terminology is with respect to the OpenCL device. So, CL_MEM_READ_ONLY means that the OpenCL device can only get this data — it can’t send it back to the host CPU. Other options are CL_MEM_WRITE_ONLY and CL_MEM_READ_WRITE.

Enqueuing Works Like a Conveyor Belt

- Write Buffer dA
- Write Buffer dB
- Execute Kernel
- Read Buffer dC

The .cl File

```c
kernel void ArrayMult( global const float *dA, global const float *dB, global float *dC ) {
  int gid = get_global_id( 0 );
}
```

Which dimension's index are we fetching?

- 0 = X
- 1 = Y
- 2 = Z

Since this is a 1D problem, X is the only index we need to get.
OpenCL code is compiled in the Driver...

```c
// Application Program

OpenCL Driver

OpenCL code in a separate file

GPU

```

7. Read the Kernel Code from a File into a Character Array

```c
const char *CL_FILE_NAME = { "arraymult.cl" }; ... FILE *fp = fopen( CL_FILE_NAME, "r" ); if (fp == NULL) { fprintf(stderr, "Cannot open OpenCL source file '%s'
", CL_FILE_NAME); return 1; } // read the characters from the opencl kernel program:

```

GLSL code is compiled in the Driver...

```c
// Application Program

GLSL Driver

GLSL code in a separate file

GPU

```

A Warning about Editing on Windows and Running on Linux

Some of you will end up having strange, unexplainable problems with your csh scripts, .cpp programs, or .cl programs. This could be because you are typing your code in on Windows (using Notepad or Wordpad or Word) and then running it on Linux. Windows likes to insert an extra carriage return (\r) at the end of each line, which Linux interprets as a garbage character.

You can test this by typing the Linux command:
```
od -c  loop.csh
```
which will show you all the characters, even the \r (which you don't want) and the \n (newlines, which you do want).

To get rid of the carriage returns, enter the Linux command:
```
tr -d \r' < loop.csh > loop1.csh
```

Then run loop1.csh

Or, on some systems, there is a utility called dos2unix which does this for you:
```
dos2unix < loop.csh > loop1.csh
```

Sorry about this. Unfortunately, this is a fact of life when you mix Windows and Linux.
### Something new: Intermediate Compilation

- You pre-compile your OpenCL code with an external compiler
- Your OpenCL code gets turned into an intermediate form known as SPIR-V
- SPIR-V gets turned into fully-compiled code at runtime

**Advantages:**
1. Software vendors don’t need to ship their OpenCL source
2. Syntax errors appear during the SPIR-V step, not during runtime
3. Software can launch faster because half of the compilation has already taken place
4. This guarantees a common front-end syntax
5. This allows for other language front-ends

---

### 8. Compile and Link the Kernel Code

```cpp
// create the kernel program on the device:
char * strings[1];
strings[0] = clProgramText;
cl_program program = clCreateProgramWithSource(context, 1, (const char **) strings, NULL, &status);
delte[ ] clProgramText;

// build the kernel program on the device:
char *options = { "" },
status = clBuildProgram(program, 1, &device, options, NULL, NULL);
if( status != CL_SUCCESS )
{
    // retrieve and print the error messages:
    size_t size;
    clGetProgramBuildInfo(program, devices[0], CL_PROGRAM_BUILD_LOG, 0, NULL, &size);
    cl_char *log = new cl_char[ size ];
    clGetProgramBuildInfo(program, devices[0], CL_PROGRAM_BUILD_LOG, size, log, NULL);
    fprintf(stderr, "clBuildProgram failed:
%s
", log);
    delete[ ] log;
}
```

### Why use an array of strings to hold the OpenCL program, instead of just a single string?

1. You can use the same OpenCL source and insert the appropriate “#defines” at the beginning
2. You can insert a common header file (= .h file)
3. You can simulate a “#include” to re-use common pieces of code
9. Create the Kernel Object

```c
cl_kernel kernel = clCreateKernel( program, "ArrayMult", &status );
```

10. Setup the Arguments to the Kernel Object

```c
status = clSetKernelArg( kernel, 0, sizeof(cl_mem), &dA );
status = clSetKernelArg( kernel, 1, sizeof(cl_mem), &dB );
status = clSetKernelArg( kernel, 2, sizeof(cl_mem), &dC );
```

11. Enqueue the Kernel Object for Execution

```c
size_t globalWorkSize[3] = { NUM_ELEMENT, 1, 1 };
size_t localWorkSize[3] = { LOCAL_SIZE, 1, 1 };
Wait( cmdQueue ); // will be covered in the OpenCL event notes
double time0 = omp_get_wtime();
status = clEnqueueNDRangeKernel( cmdQueue, kernel, 1, NULL, globalWorkSize, localWorkSize, 0, NULL, NULL );
Wait( cmdQueue ); // will be covered in the OpenCL event notes
double time1 = omp_get_wtime();
```

Work-Groups, Local IDs, and Global IDs

```
# dimensions
# events
event object

status = clEnqueueNDRangeKernel( cmdQueue, kernel, 1, NULL, globalWorkSize, localWorkSize, 0, NULL, NULL );
```

```
G_x = 20  
W_x = 5  
L_y = 4  

# WorkGroups = \frac{GlobalIndexSpaceSize}{WorkGroupsSize}

5 = \frac{20}{4}
```

"NDRange Index Space" can be 1D, 2D, or 3D. This one is 1D.
Work-Groups, Local IDs, and Global IDs

"NDRange Index Space" can be 1D, 2D, or 3D. This one is 2D.

\[
\frac{\text{WorkGroups} \times \text{Lx} \times \text{Ly}}{\text{GlobalIndexSpaceSize} \times \text{WorkGroupSize}}
\]

\[
5 \times 4 = \frac{20 \times 12}{4 \times 3}
\]

Figuring Out What Thread You Are and What Your Thread Environment is Like

uint get_work_dim( ) ;
size_t get_global_size( uint dimindx ) ;

size_t get_global_id( uint dimindx ) ;
size_t get_local_size( uint dimindx ) ;
size_t get_local_id( uint dimindx ) ;
size_t get_num_groups( uint dimindx ) ;
size_t get_group_id( uint dimindx ) ;
size_t get_global_offset( uint dimindx ) ;

0 \leq \text{dimindx} \leq 2

12. Read the Results Buffer Back from the Device to the Host

status = clEnqueueReadBuffer( cmdQueue, dC, CL_TRUE, 0, dataSize, hC, 0, NULL, NULL );
13. Clean Everything Up

// clean everything up:
clReleaseKernel( kernel );
clReleaseProgram( program );
clReleaseCommandQueue( cmdQueue );
clReleaseMemObject( dA );
clReleaseMemObject( dB );
clReleaseMemObject( dC );
delete [] hA;
delete [] hB;
delete [] hC;  
Do this because we created these arrays with new

Array Multiplication Performance:
What is a Good Work-Group Size?

<table>
<thead>
<tr>
<th>Array Size (K)</th>
<th>Work-Group Size</th>
<th>GigaMultiplications/Second</th>
</tr>
</thead>
<tbody>
<tr>
<td>8192</td>
<td>64</td>
<td>160</td>
</tr>
<tr>
<td>4096</td>
<td>64</td>
<td>150</td>
</tr>
<tr>
<td>2048</td>
<td>64</td>
<td>140</td>
</tr>
<tr>
<td>1024</td>
<td>64</td>
<td>130</td>
</tr>
<tr>
<td>512</td>
<td>64</td>
<td>120</td>
</tr>
<tr>
<td>256</td>
<td>64</td>
<td>110</td>
</tr>
<tr>
<td>128</td>
<td>64</td>
<td>100</td>
</tr>
<tr>
<td>64</td>
<td>64</td>
<td>90</td>
</tr>
<tr>
<td>32</td>
<td>64</td>
<td>80</td>
</tr>
<tr>
<td>16</td>
<td>64</td>
<td>70</td>
</tr>
<tr>
<td>8</td>
<td>64</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>64</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>64</td>
<td>40</td>
</tr>
<tr>
<td>1</td>
<td>64</td>
<td>30</td>
</tr>
</tbody>
</table>

Writing the .cl Program's Binary Code

size_t binary_sizes;
status = clGetProgramInfo( Program, CL_PROGRAM_BINARY_SIZES, 0, NULL, &binary_sizes );
size_t size;
status = clGetProgramInfo( Program, CL_PROGRAM_BINARY_SIZES, sizeof(size_t), &size, NULL );
unsigned char *binary = new unsigned char [ size ];
status = clGetProgramInfo( Program, CL_PROGRAM_BINARIES, size, &binary, NULL );
FILE *fpbin = fopen( "particles.nv", "wb" );
if( fpbin == NULL )
{
    fprintf( stderr, "Cannot create 'particles.bin'
" );
}
else
{
    fwrite( binary, 1, size, fpbin );
    fclose( fpbin );
}
delete [] binary;
Importing that Binary Code back In:

Instead of doing this:
```c
char *strings[] = { clProgramText; cl_program program = clCreateProgramWithSources( context, 1, (const char **)&strings, NULL, &status ); delete [] clProgramText; }
```

You would do this:
```c
unsigned char byteArray[ numBytes ]; cl_program program = clCreateProgramWithBinary( context, 1, &device, &numBytes, &byteArray, &binaryStatus, &status ); delete [] byteArray; 
```

And you still have to do this:
```c
char *options[] = { ""}; status = clBuildProgram( program, 1, &device, options, NULL, &status );
if( status != CL_SUCCESS ) {
    size_t size;
    clGetProgramBuildInfo( program, device, CL_PROGRAM_BUILD_LOG, 0, NULL, &size );
    cl_char *log = new cl_char[ size ];
    clGetProgramBuildInfo( program, device, CL_PROGRAM_BUILD_LOG, size, log, NULL );
    fprintf( stderr, "clBuildProgram failed in
%s
", log );
    delete [] log; 
}
```

Instead of doing this:
```c
char * strings[ 1 ]; strings[0] = clProgramText; cl_program program = clCreateProgramWithSource( context, 1, (const char **)strings, NULL, &status ); delete [] clProgramText; 
```