

Introduction to Parallel Matlab

Daniel D. Warner

May 31, 2010

What is Matlab?

Matlab is . . .

- an Integrated Development Environment for solving numerical problems in computational science.
- a collection of state-of-the-art algorithms for scientific computing with a particular emphasis on matrix algebra
- a programming language where matrices are first class citizens
- a facility for running Matlab programs on clusters

Parallelism in Matlab

Matlab's 5-fold way to parallelism:

- Multiple Threads – the internal routines are written to explicitly support multiple threads and multiple cores.
- Asynchronous Batch Mode - submit jobs to a worker pool
- SIMD - uses Matlab's pmode to perform the same instructions on multiple workers.
- Distributed Matrices - built-in operations on matrices distributed across multiple workers. This is actually MPI.
- GPU - the most recent version of Matlab can use GPUs via MEX files (C-code describing a GPU kernel).

Matrix Algebra

Compute $AB - BA$ using loops (time00.m)

```
C = zeros(N);
for i=1:N
    for j=1:N
        for k=1:N
            C(i,j) = C(i,j) + A(i,k)*B(k,j);
        end
        for k=1:N
            C(i,j) = C(i,j) - B(i,k)*A(k,j);
        end
    end
end
end
```

Matrix Algebra

Matrices are first class citizens of Matlab (time01.m)

$$C = A*B - B*A;$$

By leveraging a little knowledge of matrix algebra

- One line of code (10 characters) replaces 11 lines
- For $N = 1,000$ that single line tells the computer to do a billion multiplies and additions.
- For the right problems Matlab provides **Coding Power**.

Demo 1 - Multiple Cores

Daniel D. Warner

Introduction

Matlab's 5-fold
way

Basic Matlab
Example

Parallel Demos

Conclusion

Matlab automatically uses multiple cores (time01.m)

- With multiple threads disabled, Matlab takes about 16 seconds to compute $AB - BA$ a dozen times when A and B are 1200 by 1200 matrices.
- With multiple threads enabled on a 2 core machine, Matlab takes about 10 seconds to complete the same computation.
- The use of the multiple cores is easily shown with an activity monitor.
- This automatic use of multiple cores depends on the fact that Mathworks has coded the low level matrix operations to run efficiently on multiple cores.

Demo 2 - HTC parfor

With the **parfor** command Matlab can perform simple high throughput computing (time01).

- Create a pool of workers with the **matlabpool open** command
- Execute the **parfor** loop
- The code within each iteration of the parfor loop must be completely independent.
- Not really effective without separate processors
- Substantial startup overhead

Demo 3 - Multiple Tasks

Daniel D. Warner

Introduction

Matlab's 5-fold
way

Basic Matlab
Example

Parallel Demos

Conclusion

With the **batch** command Matlab can perform multiple tasks.

- Use the **batch** command to create workers that run separate tasks.
- `batchdemo1` - generates $\sin(x)$ and $\cos(x)$ as separate tasks, then plots the results together.
- The **batch** command runs tasks asynchronously - hence followed by a **wait** on that job.
- `batchdemo2` - the $\cos(x)$ task is launched with the command

```
job2 = batch('mywave03','matlabpool',2)
```

- The second task uses a `parfor` with 2 subworkers.

Demo 4 - SIMD Calculations

Daniel D. Warner

Introduction

Matlab's 5-fold
way

Basic Matlab
Example

Parallel Demos

Conclusion

With the **pmode** command Matlab can perform SIMD calculations.

- Use the **pmode** command to create interactive parallel labs.
- In Matlab **labindex** and **numlabs** are the counterparts of MPI's rank and worldsize.
- Each lab executes the same instructions on different data.
- Each lab can utilize all Matlab commands and functions.

Demo 5 - MPI Calculations

With the **pmode** command Matlab can perform MPI calculations.

- Use the **pmode** command to create interactive parallel labs.
- Use **distributed** and **distributor()** to distribute matrices across the pool of labs.
- The host still issues commands that each lab executes. This instruction sequence looks the same as SIMD **but operations are actually done in MPI mode.**
- Most **but not all** Matlab operations and functions have been rewritten to exploit distributed message passing parallelism.
- It is possible to “drill down” and write mpi code directly.

Conclusions

- Matlab is widely used in the Engineering community
These engineers want to solve bigger problems faster.
They do not want to learn C, mpi, cuda, etc.
Parallel Matlab and clusters can be the correct solution
- Matlab exploits parallelism at several levels and represents a significant level of code development
- Matlab's parallelism can be exploited on simple deskside clusters all the way up to large super clusters.
- Matlab is now beta testing GPU technology