Accelerating Dynamic Programming algorithms using Massively Parallel Processors

Luke Cartey

Programming Tools Group, Computing Laboratory, University of Oxford

19th November 2010

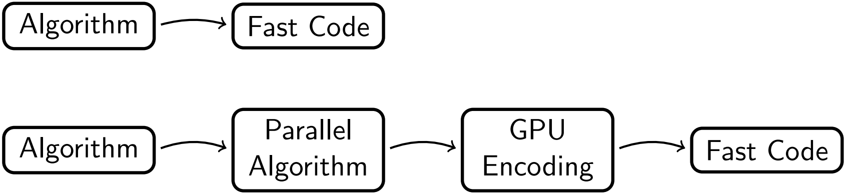
Why?

Massively Parallel Processing now **matters**.

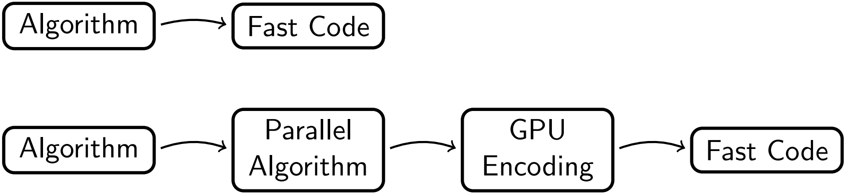
GPUs now power **3 of the top 5** supercomputers
Installed base ranges from supercomputers to netbooks

Mapping problems is hard

Mapping problems is hard

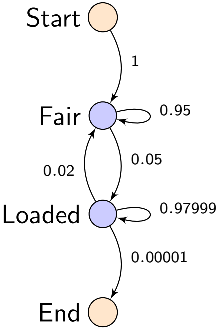


Mapping problems is hard



General-purpose automated parallelisation is difficult.
Instead, lets focus on **domains**.

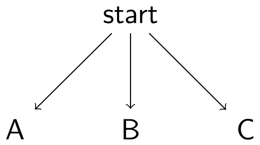
Progress so far



HMMingbird

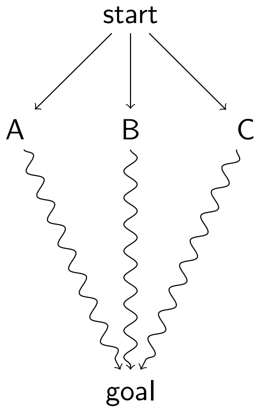
- A Hidden Markov Model compiler
- Targets GPUs
- Using a fixed set of algorithms

Optimisation Problems



Choices

Optimisation Problems



Choices

Optimal Substructure

Edit distance

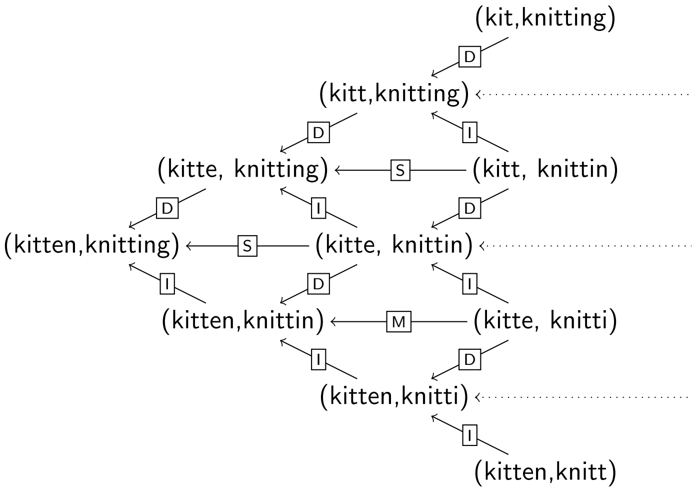
Minimum operations required to transform one string to another.
e.g

$$\text{kitten} \implies \text{kitten}\textbf{g}\ (\text{I})$$

$$\text{kitteng} \implies \text{kitt}\textbf{i}\text{ng}\ (\text{S})$$

$$\text{kitting} \implies \text{k}\textbf{n}\text{itting}\ (\text{I})$$

Graph problem



Recursion

$$f(x,y) = \begin{cases} x & \text{if } y = 0 \\ y & \text{if } x = 0 \\ f(x-1,y-1) & \text{if } s[x] = t[y] \\ min(f(x-1,y), \\ \qquad f(x,y-1), & \text{otherwise} \\ \qquad f(x-1,y-1)) + 1 & \end{cases}$$

Partitions

		k	n	i	t	t	i	n	g
	0	← 1	← 2	← 3	← 4	← 5	← 6	← 7	← 8
k	1	← 0	← 1	← 2	← 3	← 4	← 5	← 6	← 7
i	2	← 1	← 1	← 1	← 2	← 3	← 4	← 5	← 6
t	3	← 2	← 2	← 2	← 1	← 2	← 3	← 4	← 5
t	4	← 3	← 3	← 3	← 2	← 1	← 2	← 3	← 4
e	5	← 4	← 4	← 4	← 3	← 3	← 2	← 3	← 4
n	6	← 5	← 4	← 5	← 4	← 4	← 4	← 2	← 3

Partitions

		k	n	i	t	t	i	n	g
	0	1	2	3	4	5	6	7	8
k	1	0	1	2	3	4	5	6	7
i	2	1	1	1	2	3	4	5	6
t	3	2	2	2	1	2	3	4	5
t	4	3	3	3	2	1	2	3	4
e	5	4	4	4	3	3	2	3	4
n	6	5	4	5	4	4	4	2	3

Partition function

Describe each partition based on the parameters e.g

$$P_f(x, y) = x + y$$

Proving suitability of a partition function

Consider each recursive call in turn using descent functions.

$$f(d_x(x,y), d_y(x,y))$$

e.g

$$f(\ x-1\ ,\ y-1\ )$$

Verify the recursive equation satisfies the partition function:

$$\forall_{x,y} P_f(x,y) > P_f(d_x(x,y), d_y(x,y))$$

e.g

$$\forall_{x,y} x+y > (x-1)+(y-1)$$

Termination

Partition base case is when $P_f(x, y) = 0$:
The base case of the recursion needs to satisfy this.
We must also check the recursion never jumps out of bounds.

Further work

Tie work into higher-level domains for Bioinformatics
(e.g Hidden Markov Models)
Permit more sophisticated recursions