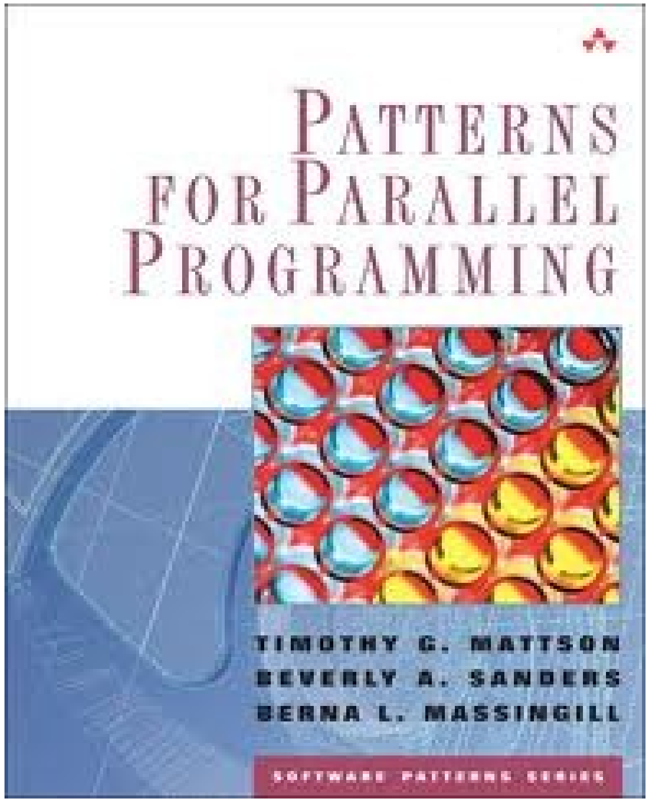
Patterns for Parallel Programming

Tim Mattson
Intel
69 PUBLICATIONS **1,326** CITATIONS
SEE PROFILE

Berna L. Massingill
Trinity University
36 PUBLICATIONS **571** CITATIONS
SEE PROFILE

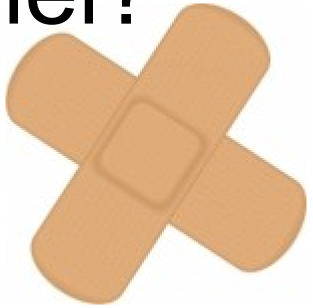
Patterns for Parallel Programming



Timothy Mattson , Beverly Sanders , Berna Massingill,
Patterns for parallel programming, Addison-Wesley Professional, 2004
ISBN-13: 978-0321228116

Sticking Plaster Pitfall

"Could you just tweak my serial code to make it run in parallel?"



Why Bother With This Book?

- Recipe based
  - Recipes guide our thinking
  - Help us not to forget
- Introduces recurrent themes and terminology
  - e.g. (memory) latency, “loop parallelism”
- Emphasises ***design***
  - Amdahl's law highlights the pitfalls of looking for sticking-plaster speed-ups in serial programs – design for concurrency

Familiar Mantras - ..only more so

Flexibility

Environments will be more heterogeneous.

Efficiency

We're going parallel for a speed-up, right?

But more pitfalls (latency, thread overheads etc.)

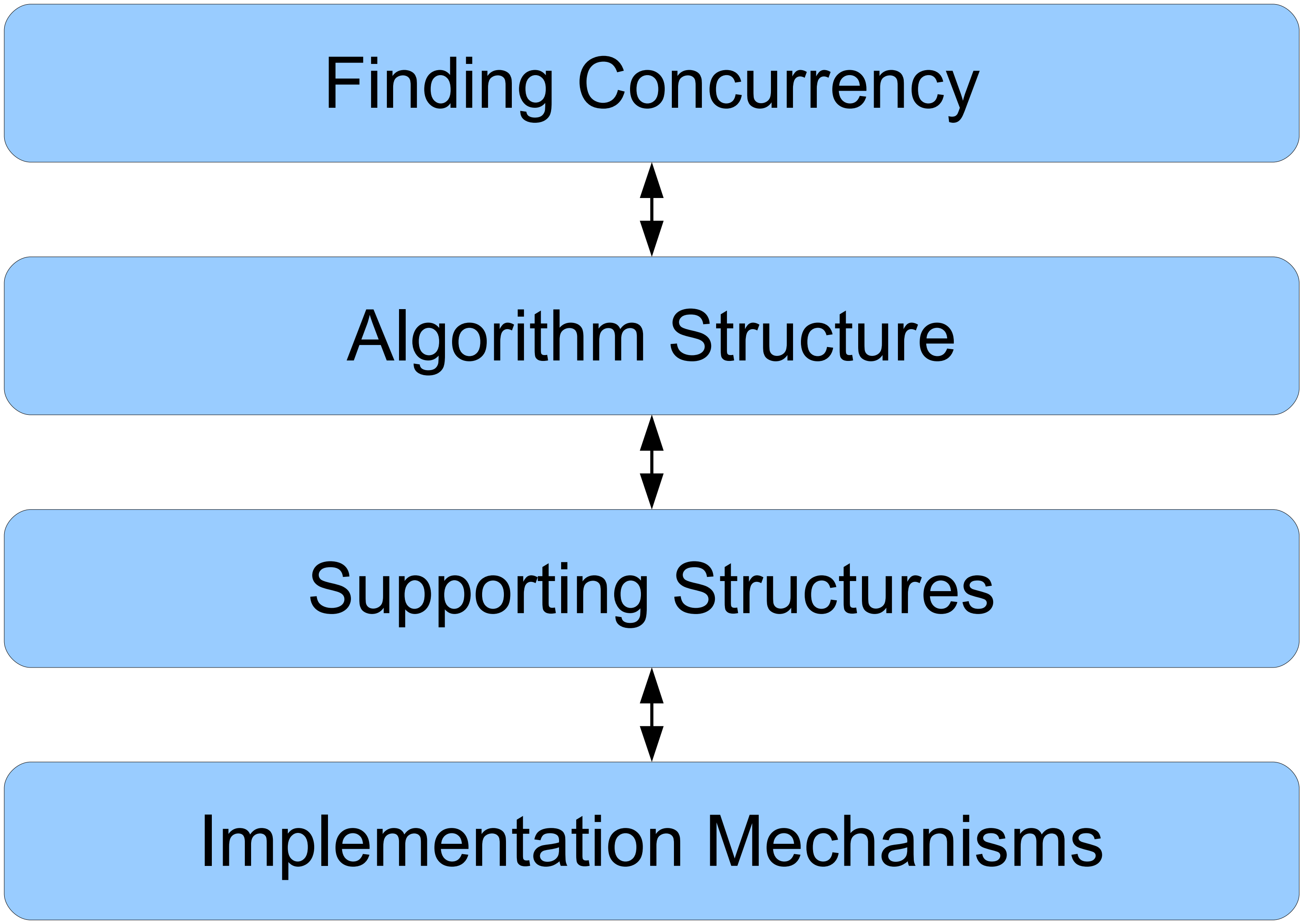
Simplicity

Parallel codes will be more complicated.

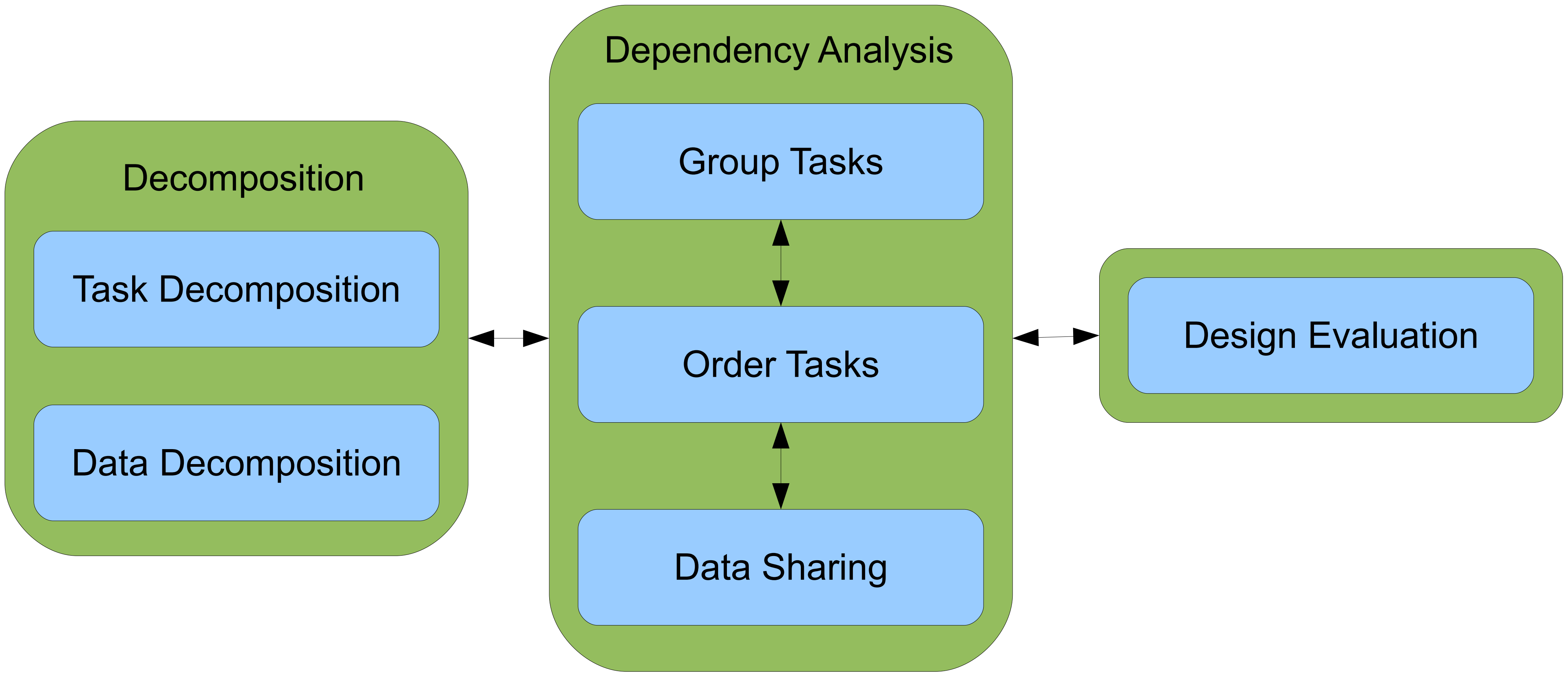
All the more reason to strive for maintainable, understandable programs.

University of BRISTOL

Four Design Spaces



Finding Concurrency

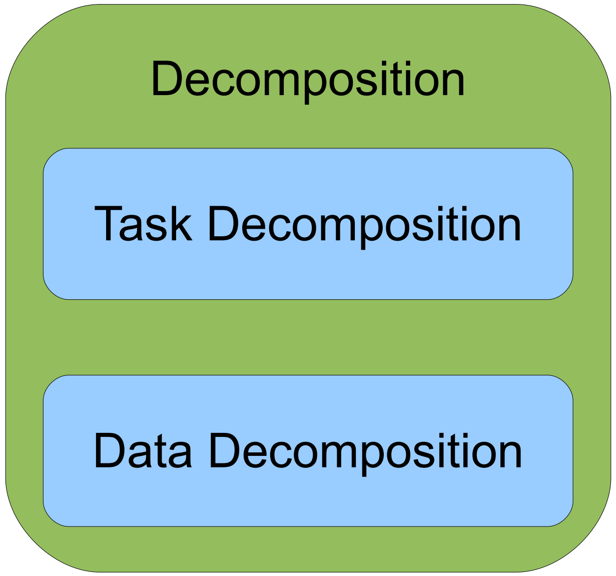


Examples

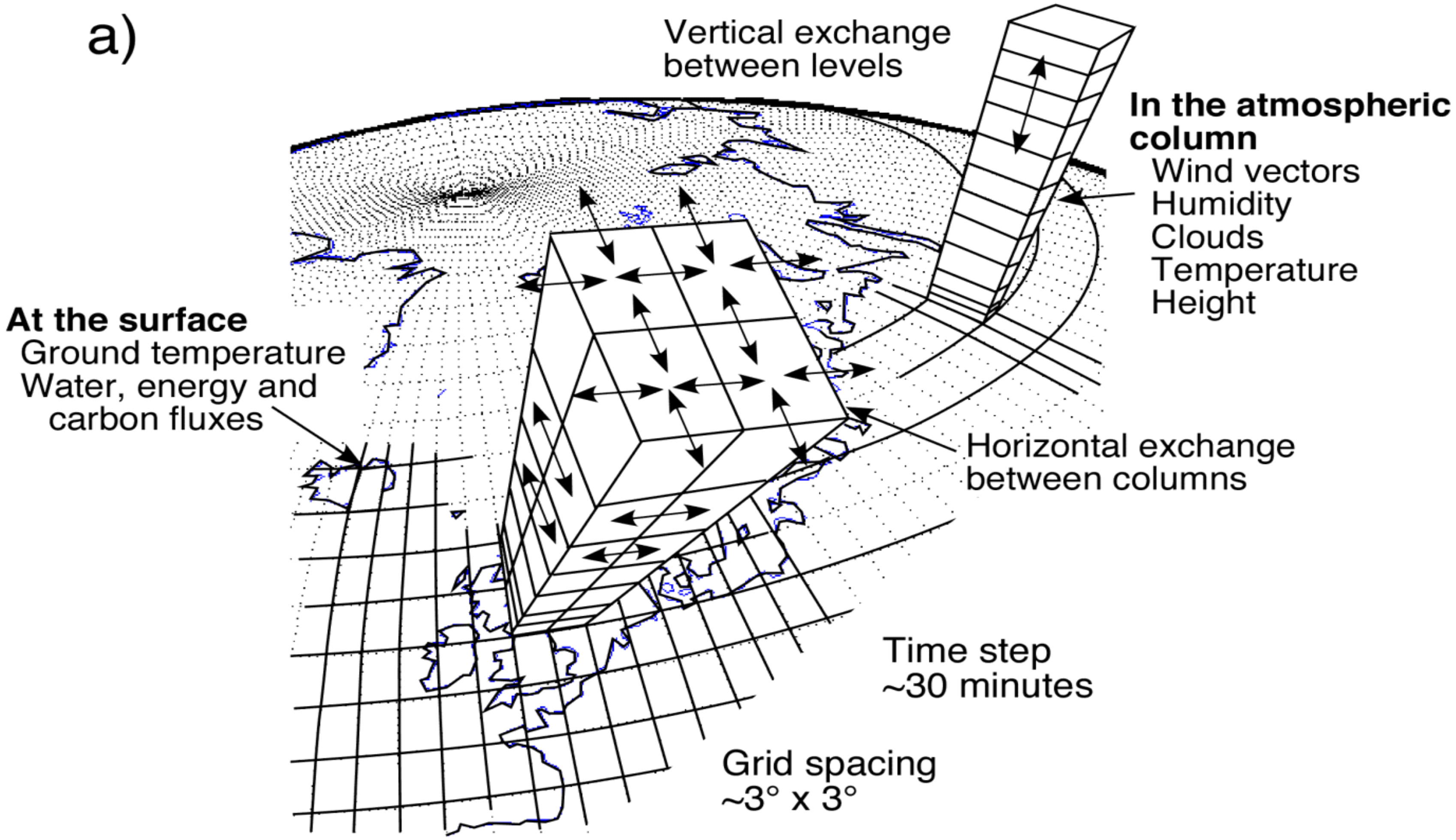
• HPC:	A Climate Model
• Embedded Systems:	A Speech Recogniser
• The Cloud:	Document Search

Highlights the fact that parallel programming is emerging everywhere..

Task vs. Data Decomposition

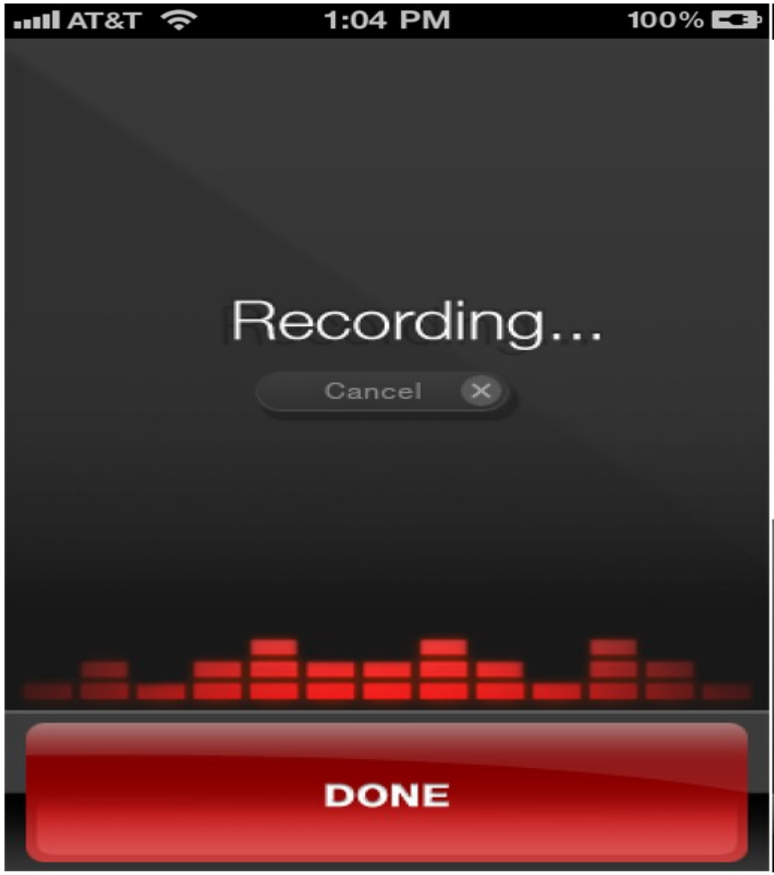


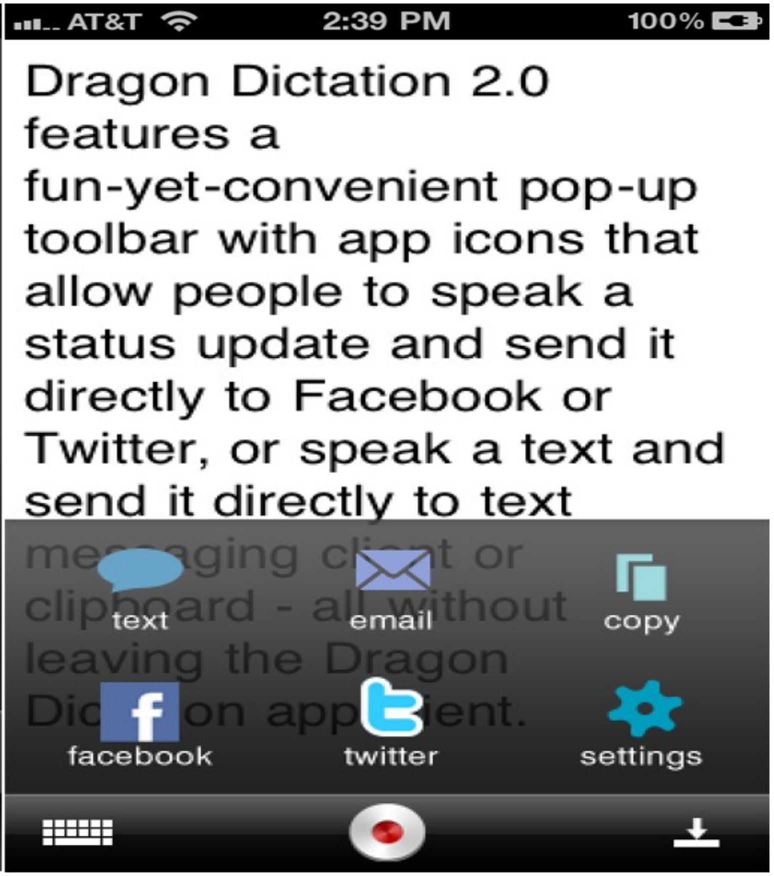
Data Decomposition (trad. HPC): A Climate Model


a)
Vertical exchange between levels
In the atmospheric column
Wind vectors
Humidity
Clouds
Temperature
Height
At the surface
Ground temperature
Water, energy and carbon fluxes
Horizontal exchange between columns
Time step ~30 minutes
Grid spacing ~3° x 3°

Data Parallel over grid cells

Task Decomposition (Embedded): A Speech Recogniser



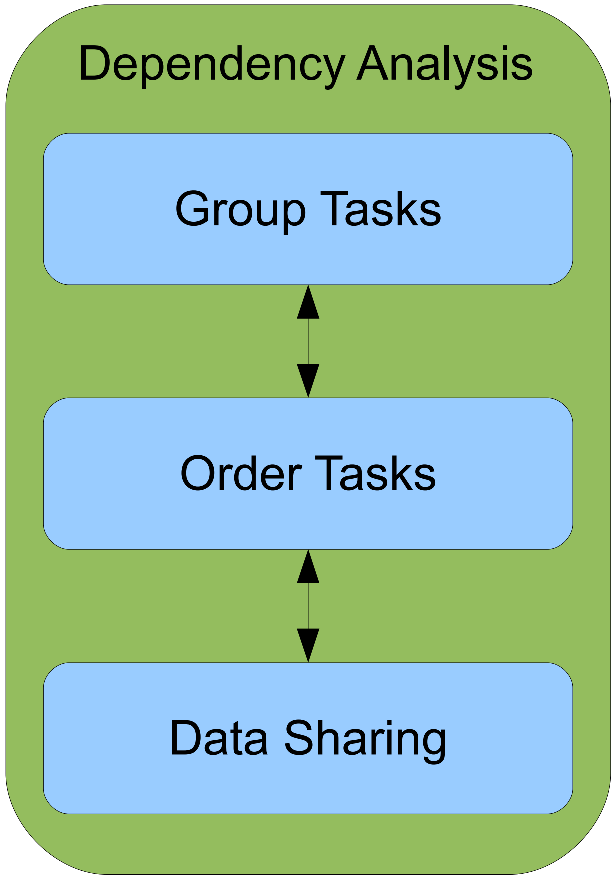


Dragon Dictation 2.0 features a fun-yet-convenient pop-up toolbar with app icons that allow people to speak a status update and send it directly to Facebook or Twitter, or speak a text and send it directly to text messaging client or clipboard - all without leaving the Dragon Dictation application.

Acoustic Analysis: concurrency in stages and components
Pattern Matching: search over many possible word matches

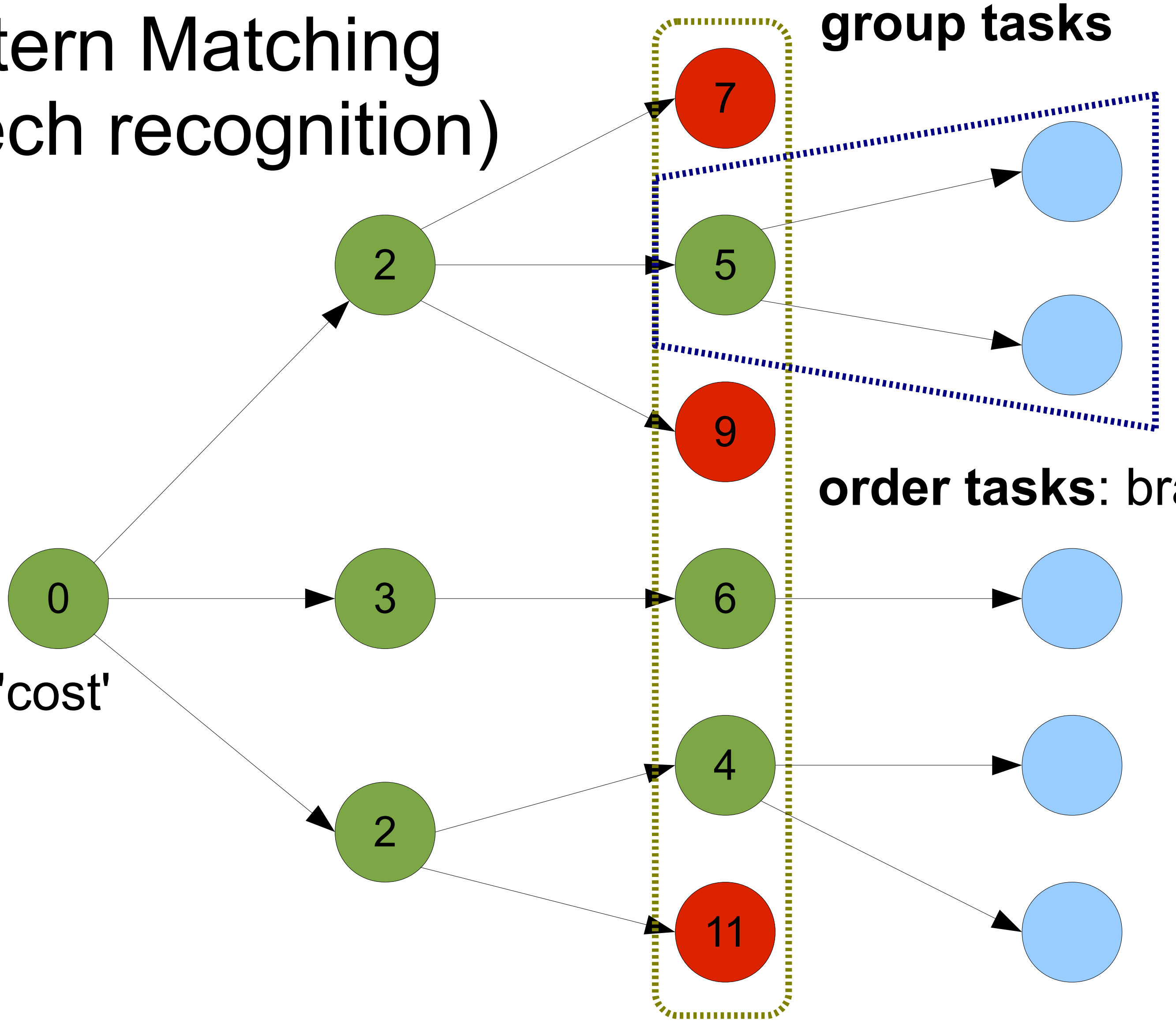
University of BRISTOL

Finding Relationships between Concurrent Tasks



Dependency Analysis

Pattern Matching
(speech recognition)



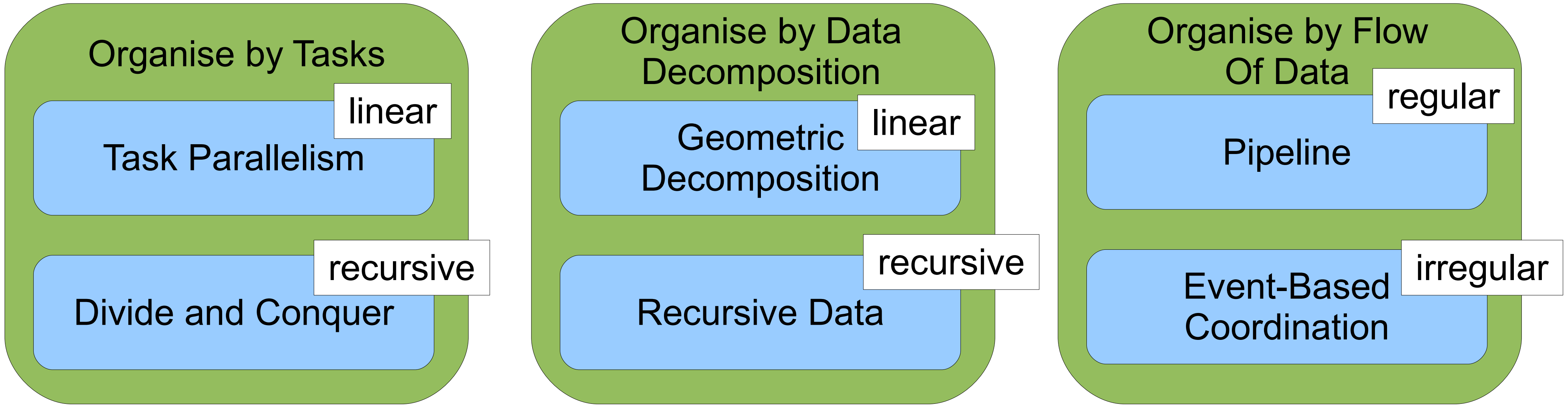
group tasks

data sharing:
- queue of paths
- legal branches
- acoustic i/p

order tasks: branch..& bound

match 'cost'

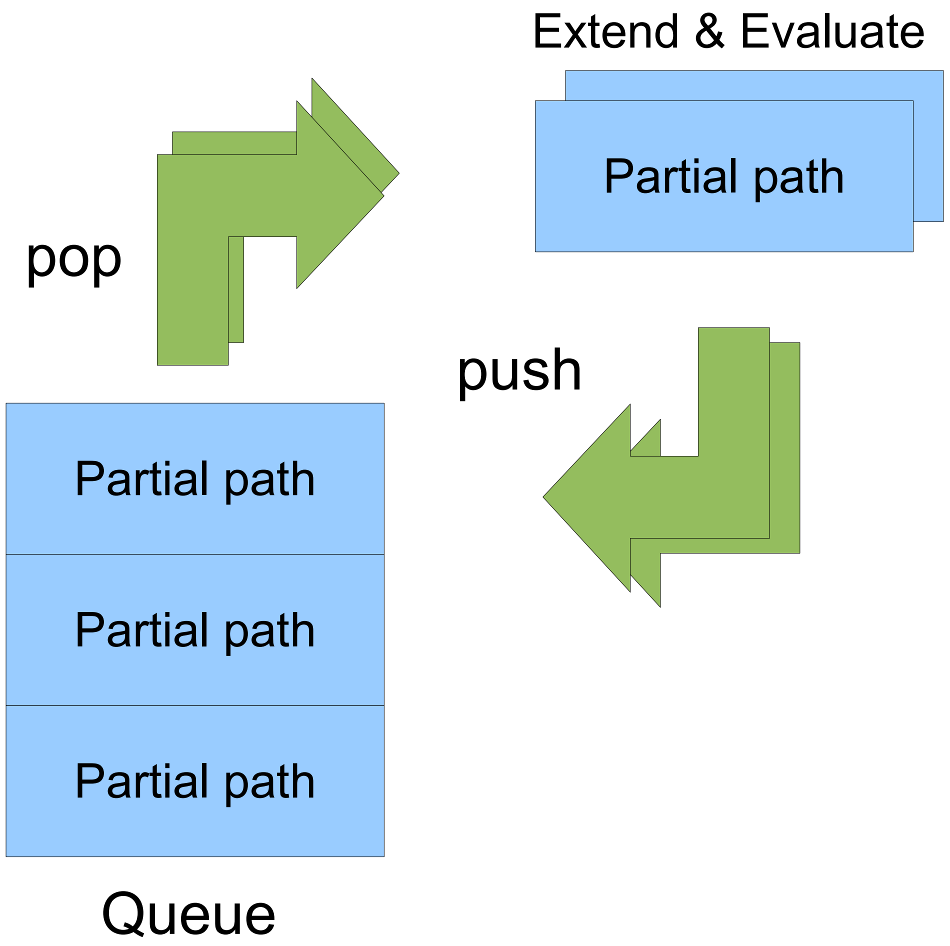
Algorithm Structure



Organise by Tasks - Some Considerations

- No dependencies between tasks
  - massively parallel (vs. embarrassingly serial!)
- Dependencies between tasks
  - Temporal (e.g. speech: real-time constraints)
  - Separable – 'reductions' (we'll see later)
- Cost of setting up task vs. amount of work done
  - See thresholds to switch to serial work (we'll see this in e.g. quicksort)

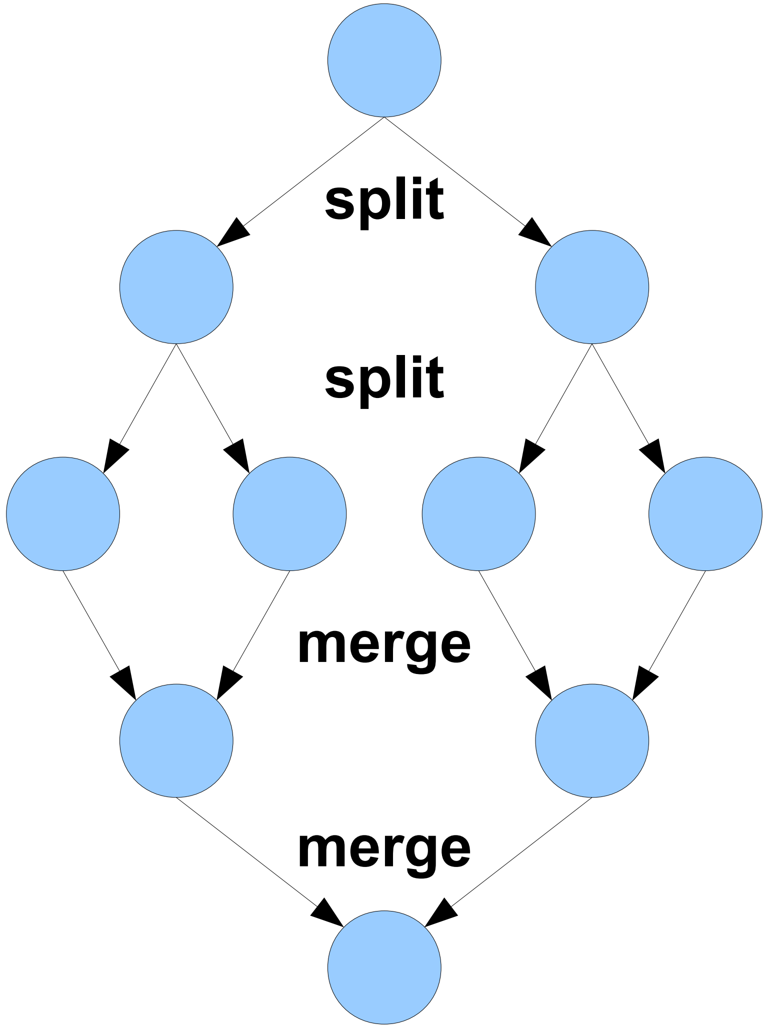
Organise by Tasks - Task Parallelism

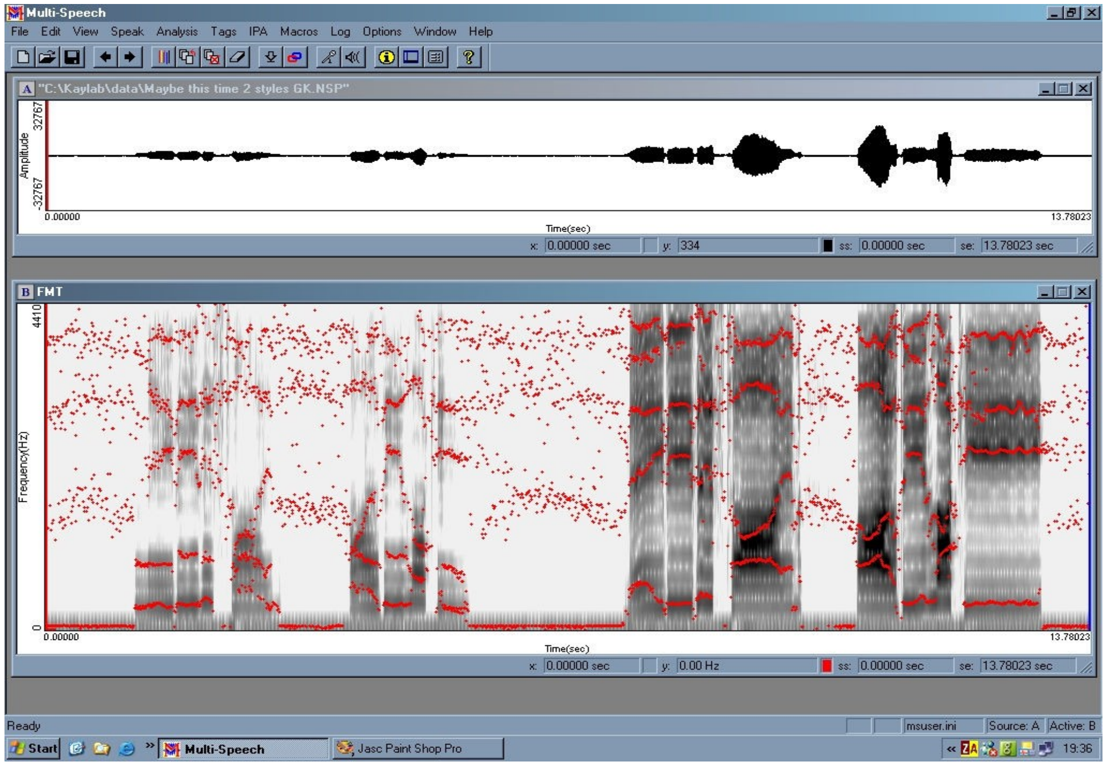


<u>Shared data:</u>
Queue of partial paths
Bounding criterion

Branch & bound implemented with a shared queue

Organise by Tasks - Divide and Conquer

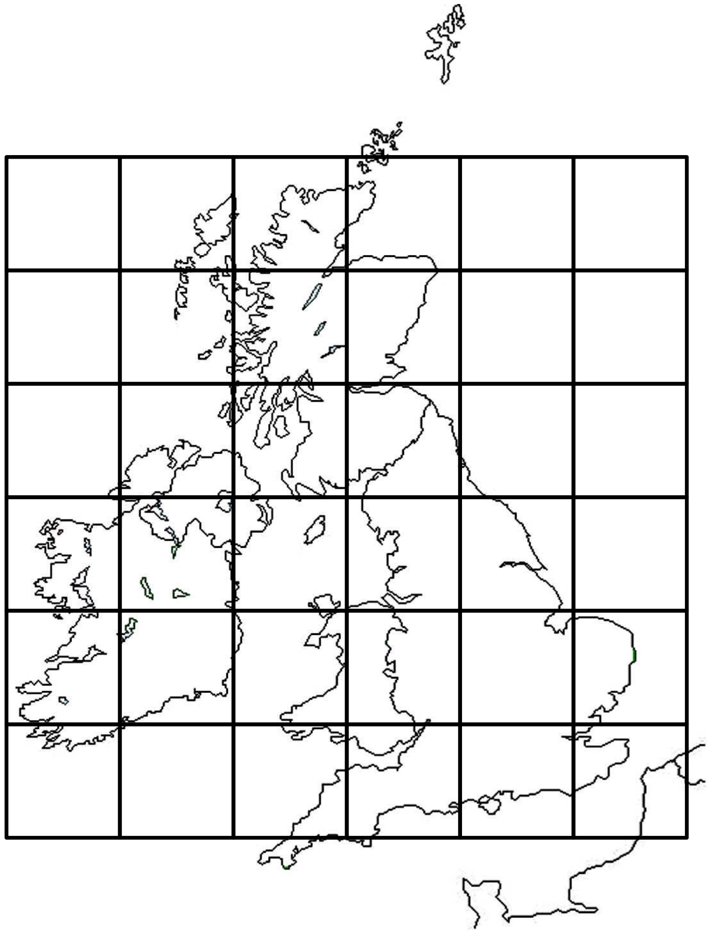




e.g. FFT for speech recognition
Sorting algorithms

University of BRISTOL

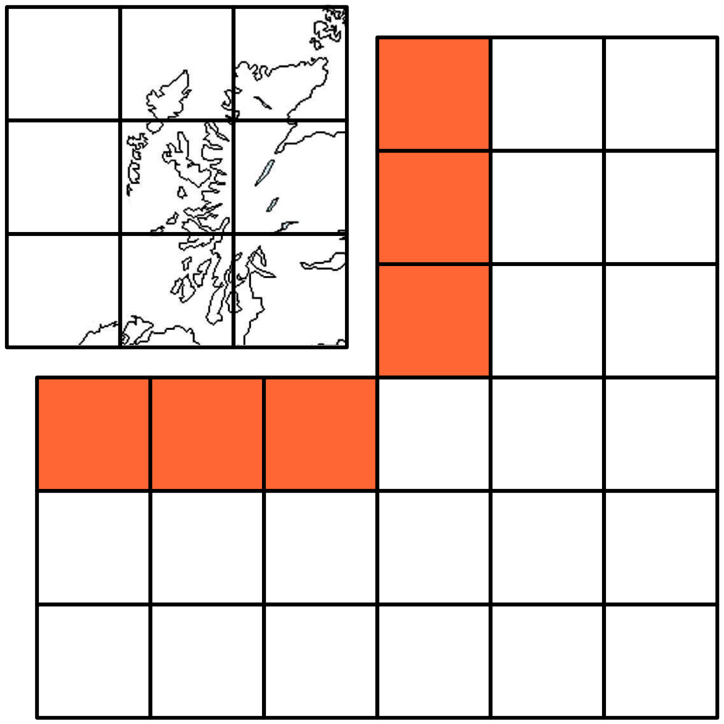
Organise by Data Decomposition - Geometric Decomposition



Grid Cells

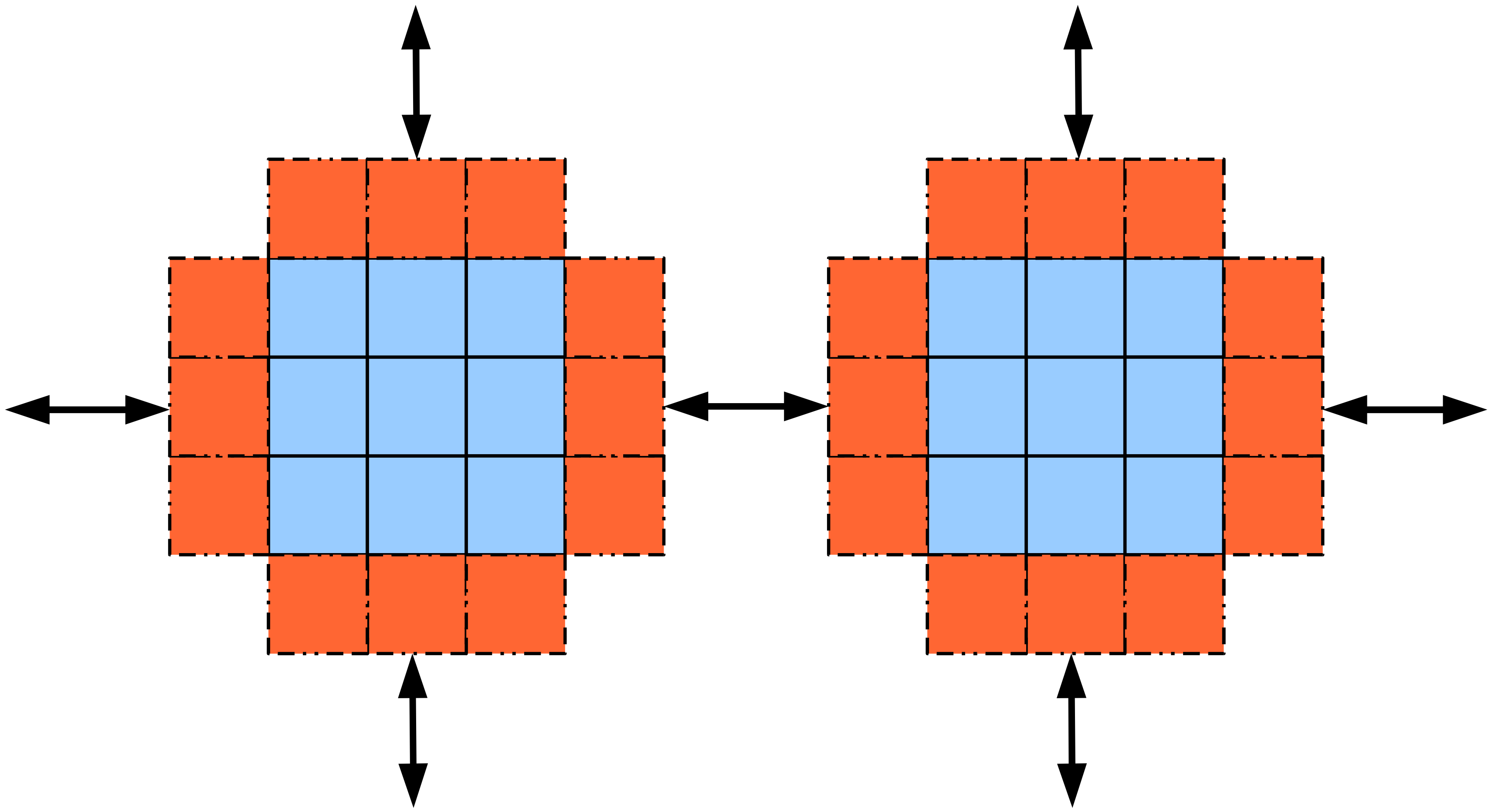
Organise by Data Decomposition - Geometric Decomposition

cells assigned to single processor



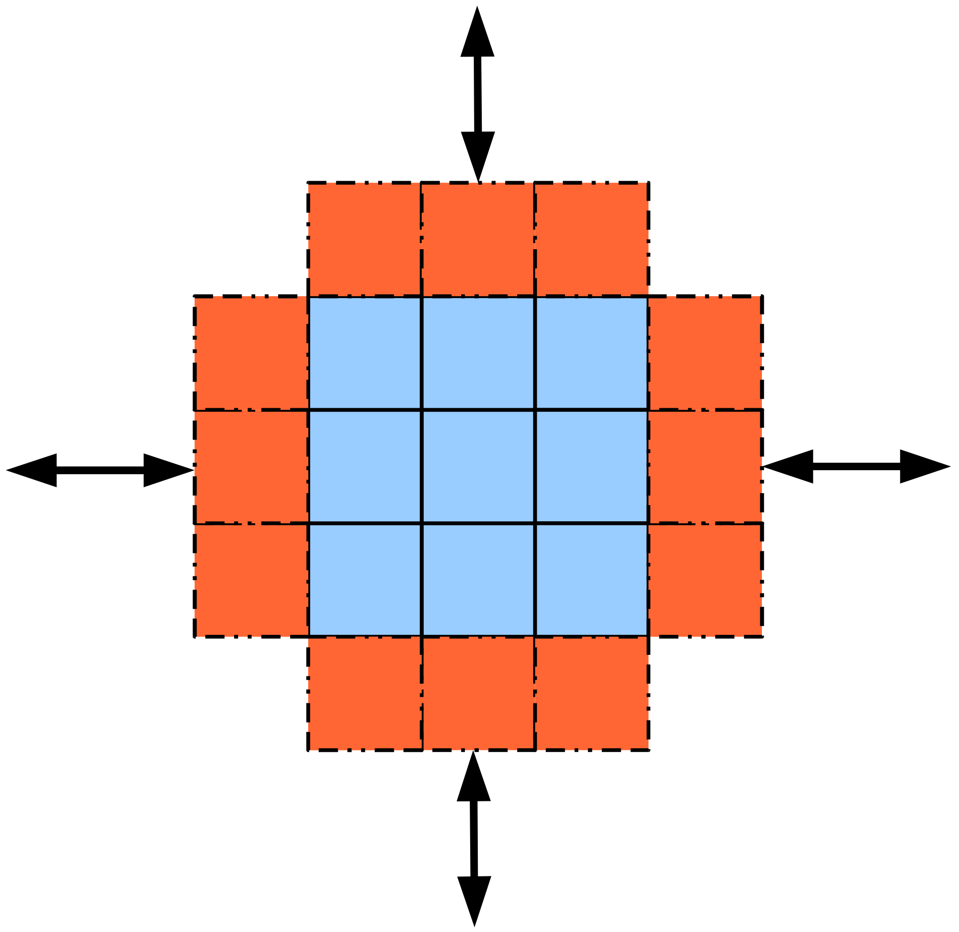
Exchange local data with neighbours

Organise by Data Decomposition - Geometric Decomposition



Halo exchange

Organise by Data Decomposition - Geometric Decomposition



<u>Benefits of halo exchange:</u>

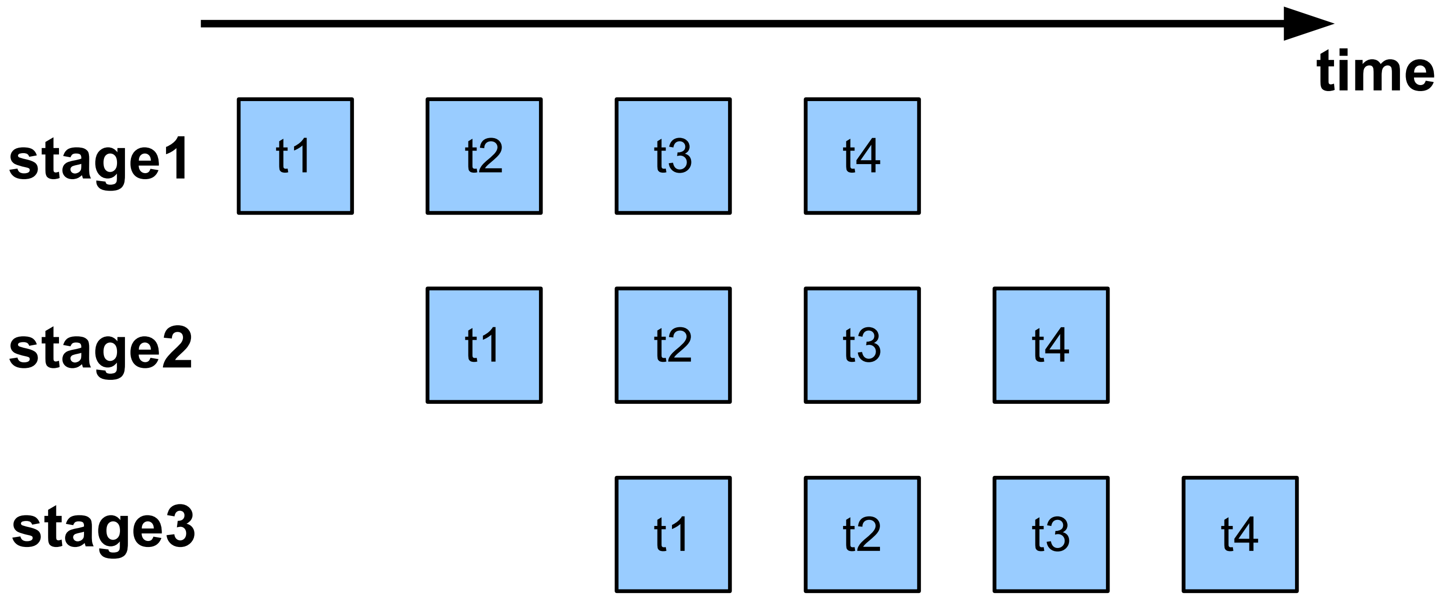
1. **Can overlap communication & computation**
2. **Compute scales with volume but communication scales with surface area**

Q. What's wrong with the number of grid cells here?

Bonus Q. Any issues with the grid on a globe? Any solutions?

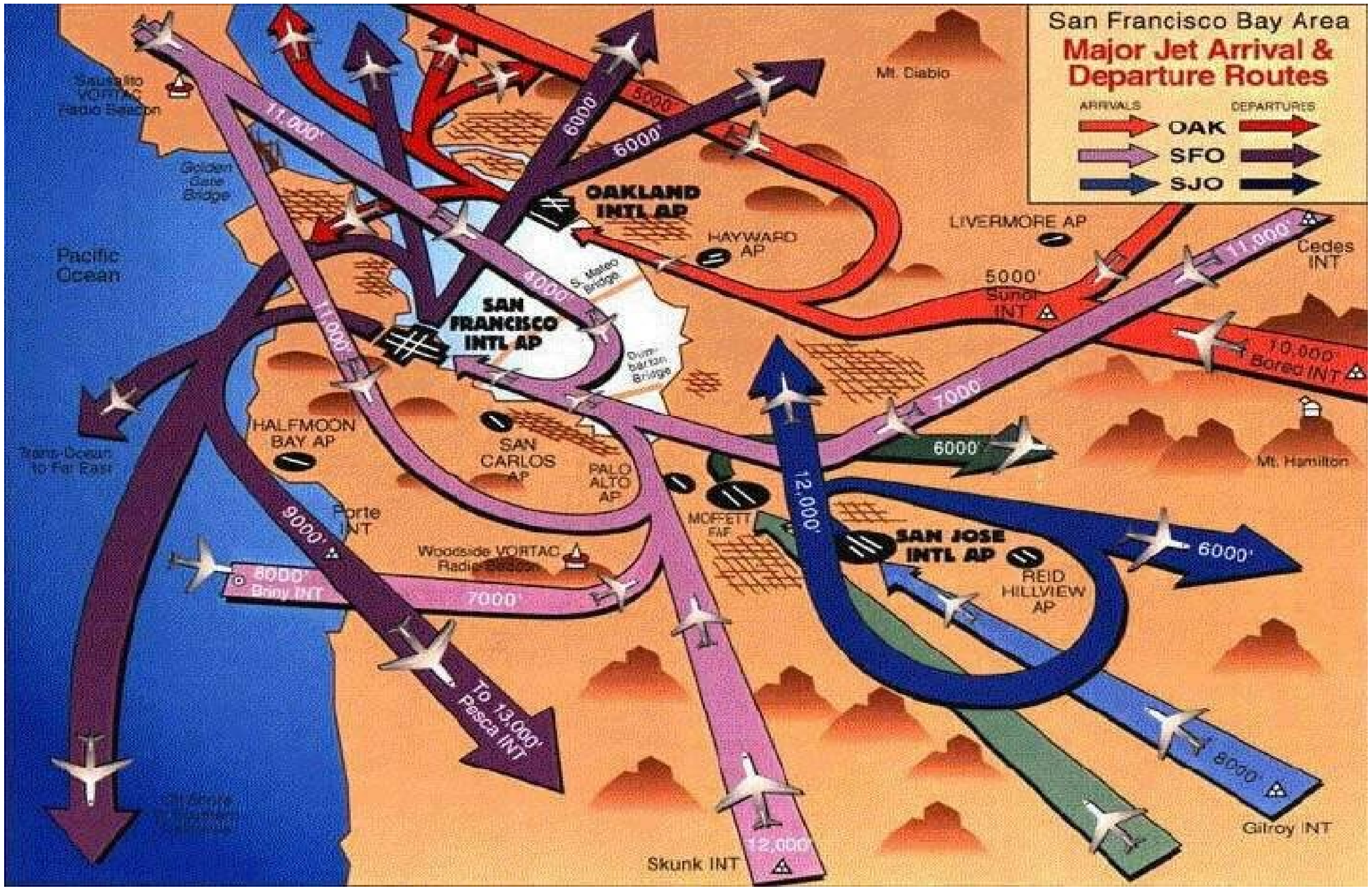
Pipeline

Pipeline



Speech recognition:
1. Discrete Fourier Transform (DFT)
2. manipulation e.g. log
3. Inverse DFT
4. Truncate 'Cepstrum' ..

Event-Based Coordination



Irregular events, ordering constraints (queues can be handy)

Supporting Structures

Program Structures

- SPMD
- Master/Worker
- Loop Parallelism
- Fork/Join

Program Structures

- Shared Data
- Shared Queue
- Distributed Array

Useful **idioms** rather than unique implementations

Single Program Multiple Data

rank = 0

```
if(rank == 0) {
printf(“MASTER\n”);
}
else {
printf(“OTHER\n”);
}
```

rank = 1

```
if(rank == 0) {
printf(“MASTER\n”);
}
else {
printf(“OTHER\n”);
}
```

- Only one program to manage
- Conditionals based on thread or process IDs
- Load balance predictable (implicit in branching)
- Plenty of examples and practice when we look at MPI

Master/Worker

Use when load balance is not predictable..

..& work cannot be distributed using loops

- PEs may have different capabilities
- A bag of independent tasks is ideal
- Workers take from bag, process, then take another

Load is **automatically** balanced in this way

Master/Worker (Cloud): MapReduce

Big Data:
Google Processed 20PB/day in 2008 using MapReduce
Also used by Yahoo, FaceBook, eBay etc.

Loop Parallelism

Use if computational expense is concentrated in loops (common in scientific code)

1. Profile code to find 'hot-spots'
2. Eliminate dependencies between iterations (e.g. private copies & reductions)
3. Parallelise loops (easy in OpenMP)
4. Tune the performance, e.g. via scheduling

We'll get plenty of practice with OpenMP

Fork/Join

Use if the number of concurrent tasks varies, e.g. if tasks are created recursively

- Beware: overhead of creating a new UEs (Uinits of Execution, e.g. thread or process)
  - Direct vs. indirect mappings from tasks to Ues
- Sorting algos are an examples

University of BRISTOL

Shared Data

- Try to avoid, as can limit scalability
- Use a concurrency-controlled (e.g. 'thread-safe') data type:
  - One-at-a-time: critical region/'mutex'
  - Look for non-interfering operations e.g. readers vs. writers
  - If pushed, finer grained critical regions, but this will increase complexity & hence the chance of a bug
- 'Shared Queue' is an instance of 'Shared Data'

Distributed Arrays

In a nutshell: partition data and distribute so that data is close to computation.

- Why? Memory access (esp. over a network) is slow relative to computation.
- Simple concept but the devil is in the details
- Some terminology:
  - 1D block, 2D block and block cyclic distribution

Libraries: e.g. ScaLAPACK

Recap of Key Points

Design..

- for massively parallel systems
- because if not today they will be tomorrow
- and in all areas of computing

Design Patterns..

- provide useful – recurring - solutions
- & structure to the process

Implementation Mechanisms

OpenMP & Pthreads

MPI

OpenCL