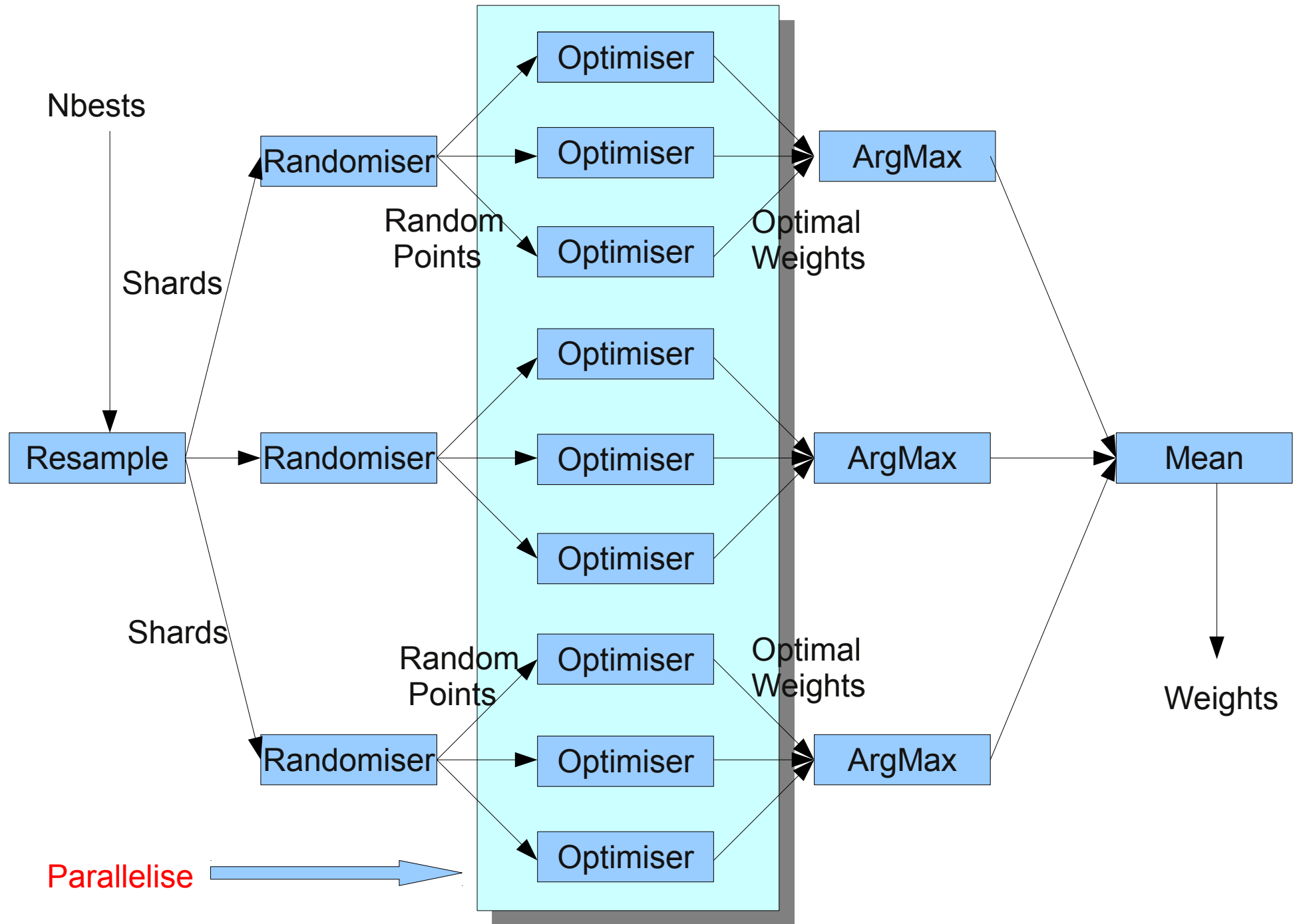


Parallelisation of MERT – Barry Haddow



Parallelisation of MERT

- Parallel random restarts – in trunk
- Sharding/Resampling implemented - in branch
 - Experiments running
 - Hope to improve stability
 - But affects convergence criteria
 - And dev score calculation
- Works for Line Search – Will add PRO Support
 - Requires better integration of PRO in mert

Parallelisation of Scoring

- Discussed but not Addressed
- Probably slowest, non-parallelised step
- Can be done with Hadoop (Map-Reduce)
 - Multi-thread map-reduce libraries available
- Need to deal with bounded memory sorting
 - Kenneth working on related code
- Text format of extract file is wasteful
 - Encoded/binary or stream