



Towards a System for Interactive Parameter Sweep Applications on The Grid

A. Wibisono, D.Vasyunin, V. Korkhov, F. Terpstra, V. Guevara, Z. Zhao, A. Belloum, M. Bubak.
IvI, Kruislaan 403, 1098 SM, Amsterdam, The Netherlands

Introduction

Parameter sweep experiment is an important task in the domains of system modeling and optimization. A straightforward approach is to execute the system with a collection of pre-selected parameters and then choose the most interesting parameter settings based on some observation metrics. Noticeable characteristics of the experiment :

- The interesting ranges of the parameters to be studied are not known in advance
- Batch processing, perform entire parameter sweep executions, analyse results, repeat from the beginning.
- Exploratory by nature, possibly trial and error of various existing methods to guide parameter search, different sampling strategy, or even different objective/observation.

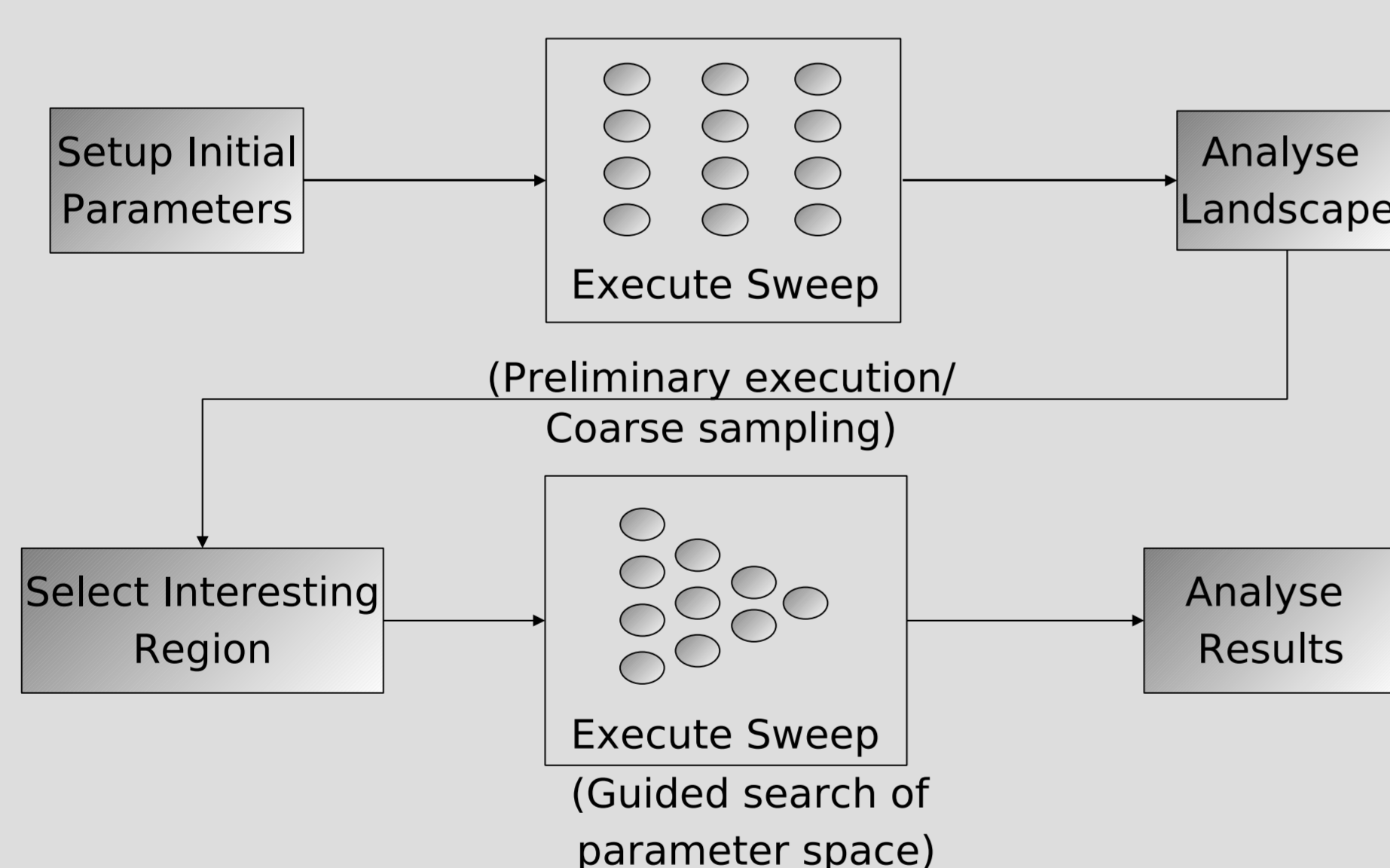
A typical parameter sweeping tools in general has two parts:

- **Job submission system** to schedule the tasks based on different combinations of parameter values.
- **The parameter searching system** to optimize the regions of parameter space and provides feedback to the task planner.

Most of the existing systems focus on the job submission system, e.g. in Nimrod/G [1], APST [2]. Little attention has been given to the searching component. Because the searching component has little knowledge on the specific problem domain, static choice of the optimization algorithm will not be sufficient for a large problem. *Including human expertise* in the runtime loop of parameter sweeping will be crucial.

We studied common practices in performing parameter sweep experiments based on literatures [4-6] and experiences from the project (VL-e) use cases [3]. From this study we have gathered evidence for the need of interactivity during execution of parameter sweep experiments.

Common Practice

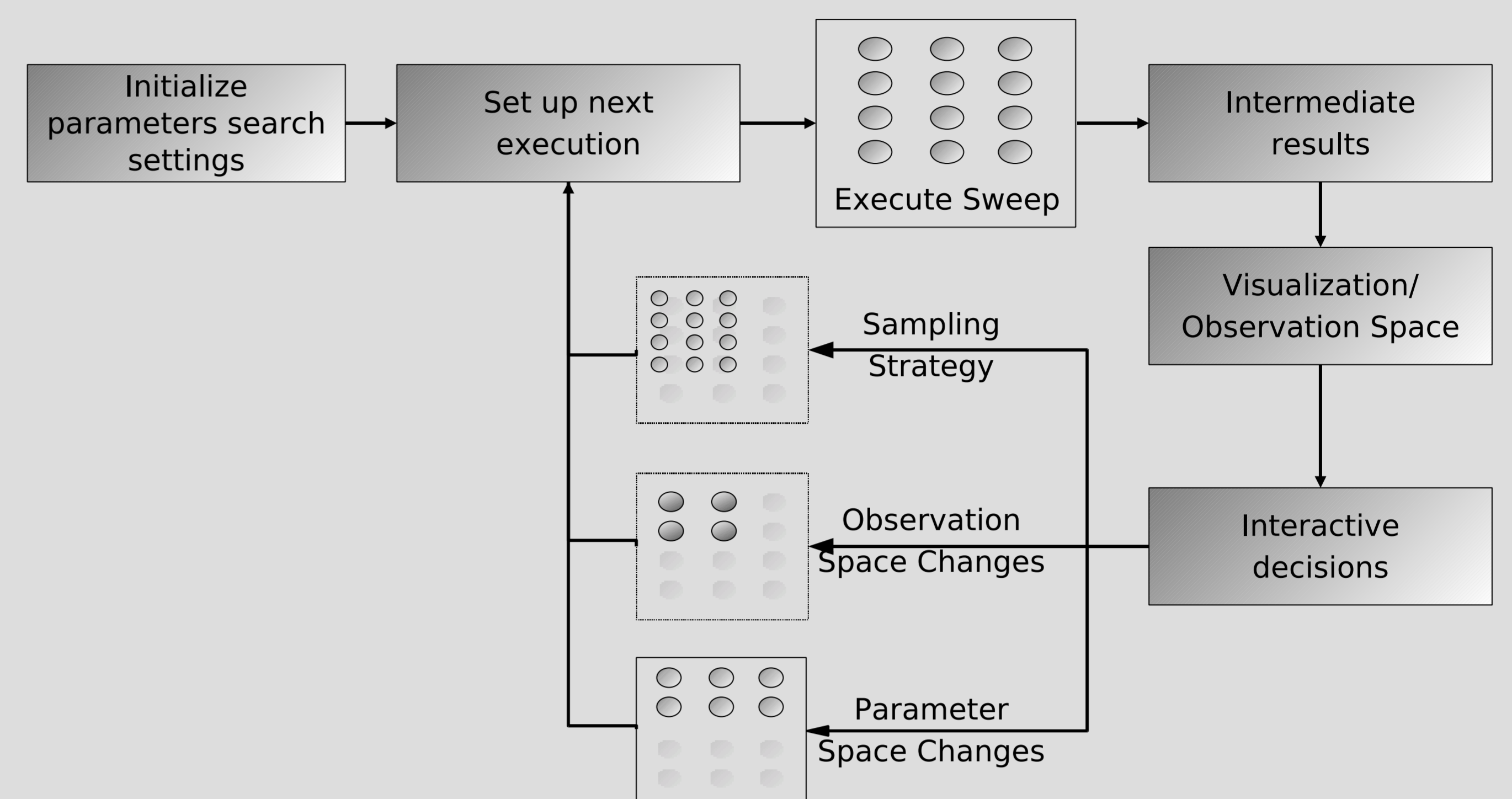


Parameter sweep experiments are performed as subsequent batch jobs submissions. Scientist waits until the entire experiment finished before analysing and performing further refinement/changes on experiment settings, and then the whole life cycle might need to be repeated from the beginning.

References

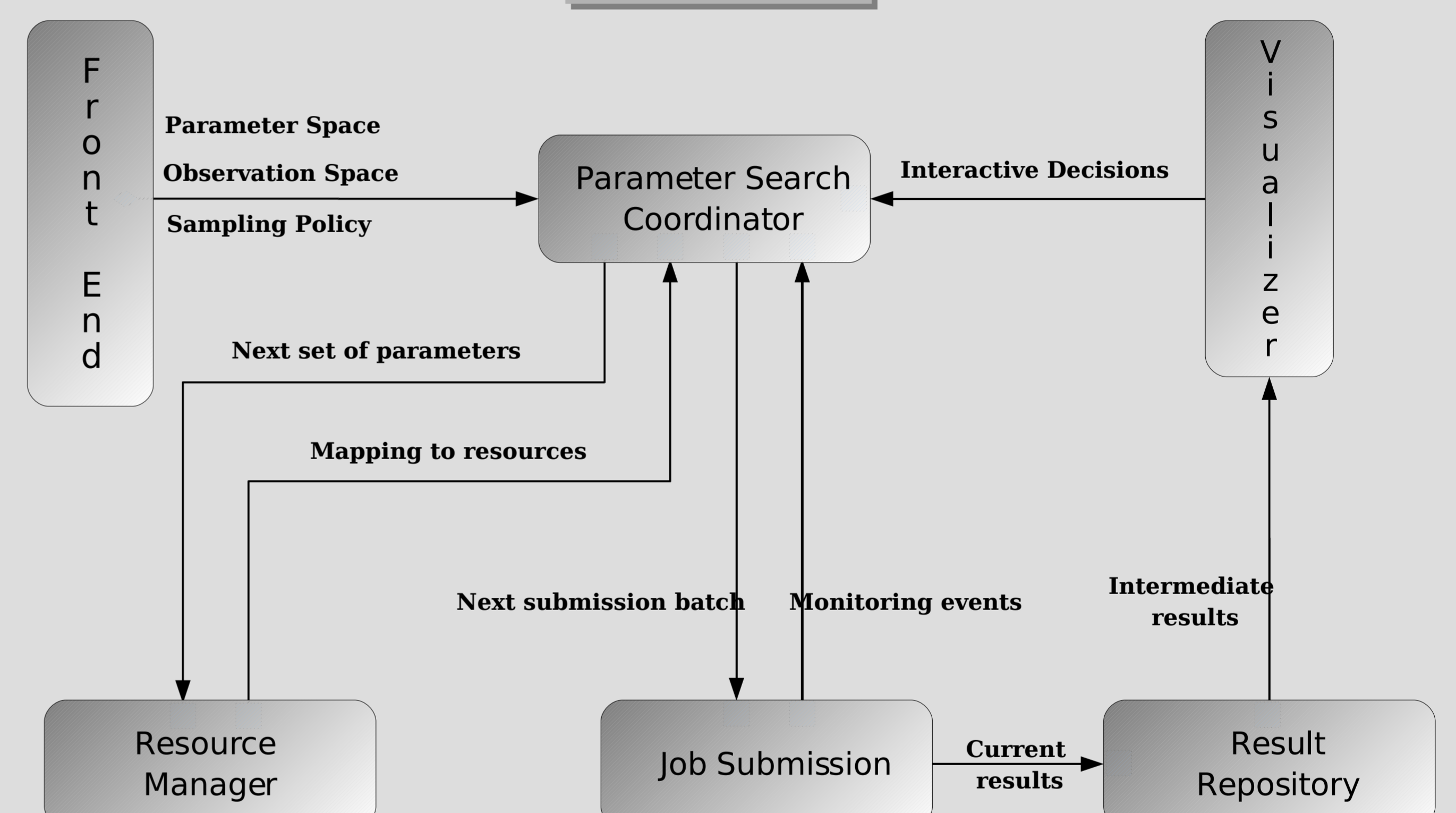
- [1] R. Buyya, D. Abramson, and J. Giddy. Nimrod/G: an architecture for a resource management and scheduling system in a global computational grid. In High Performance Computing in the Asia-Pacific Region, 2000. Proceedings. The Fourth International Conference/Exhibition on, volume 1, pages 283-289vol.1, 14-17 May 2000.
- [2] H. Casanova, F. Berman, G. Obertelli, and R. Wolski. The apples parameter sweep template: User-level middleware for the grid. In Supercomputing, ACM/IEEE 2000 Conference, pages 60-60, 04-10 Nov. 2000.
- [3] S. D. Olabarriga, A. J. Nederveen, and B. O'Nuallain. Parameter sweeps for functional mri research in the "virtual laboratory for e-science" project. In Cluster Computing and the Grid, 2007. CCGRID 2007. Seventh IEEE International Symposium on, pages 685-690, May 2007.
- [4] A. Salman, S. Turovets, A. Malony, J. Eriksen, and D. Tucker. Computational modeling of human head conductivity. In International Conference on Computational Science (1), 2005. LNCS vol 3514, pages 631-638, May 2005.
- [5] V. Stankovski and W. Dubitzky. Special section: Data mining in grid computing environments. Future Generation Computer Systems, 23(1):31-33, Jan. 2007.
- [6] W. Sudholt, K. K. Baldrige, D. Abramson, C. Enticott, S. Garic, C. Kondric, and D. Nguyen. Application of grid computing to parameter sweeps and optimizations in molecular modeling. Future Generation Computer Systems, 21(1):27-35, Jan. 2005.
- [7] A. Wibisono, D. Vasunin, V. Korkhov, Z. Zhao, A. Belloum, C. de Laat, P. W. Adriaans, Bob Hertzberger: WS-VLAM: A GT4 Based Workflow Management System. In International Conference on Computational Science (3), LNCS vol 4489, pages 191-198, 2007

Interactive Parameter Sweep



User does not have to wait until all experiments finished to make interactive decisions. Decisions could be made based on intermediate results : changes in parameter space ranges, optimizer used, or sampling strategy.

Initial Design



Front end, to submit description of parameter sweep (parameter space, observation space and sampling policy).

Parameter search coordinator, to manage the parameter space exploration.

Visualizer (and interaction), to allow user to have feed back from intermediate results of current execution and to allow the steering of the execution, changes of parameter space range, observation space, and sampling policy.

Result Repository, to manage results from previous runs and provides provenance, and avoid performing unnecessary execution.

Resource Manager and Job Submission to the grid resource will be based on components from WS-VLAM [7], a grid enabled workflow system.

Applications

Prototype will be applied to the following applications:

- **Amsterdam DDA**: package to simulate scattering and absorption of electromagnetic waves by particles of arbitrary geometry, applied to red blood cells. Tuning parameters to match experimental/observation data.
- **Biomedical application**: study in functional Magnetic Resonance Imaging of brain activity due to stimulation, investigate the effect of one parameter (delay in the hemodynamic response function) in the analysis result [3].
- **Data mining applications**: experiments based on existing classification algorithms on WEKA Toolkit (J-48 C-45) [5].

