



Alchemist

New and Upcoming Features

Kai Rothauge, Alex Gittens, Michael W. Mahoney



- Apache Spark shown to have **significant overheads** when performing linear algebra computations compared to MPI-based implementations
- **Alchemist** interfaces between Apache Spark and *existing* MPI-based libraries
 - Reduction in overheads
- **Best of both worlds**
 - Ease-of-use and data analysis workflow of Spark, efficiency of MPI
- **Communication via sockets**
 - Distributed datasets exchanged between *worker nodes*
 - Metadata and parameters exchanged between *driver nodes*



New and Upcoming Features

- **Server-Based Architecture**

- Alchemist runs in background on subset of cluster, client applications log on
- Allows for basic fault tolerance and elasticity in MPI-based libraries (future work)

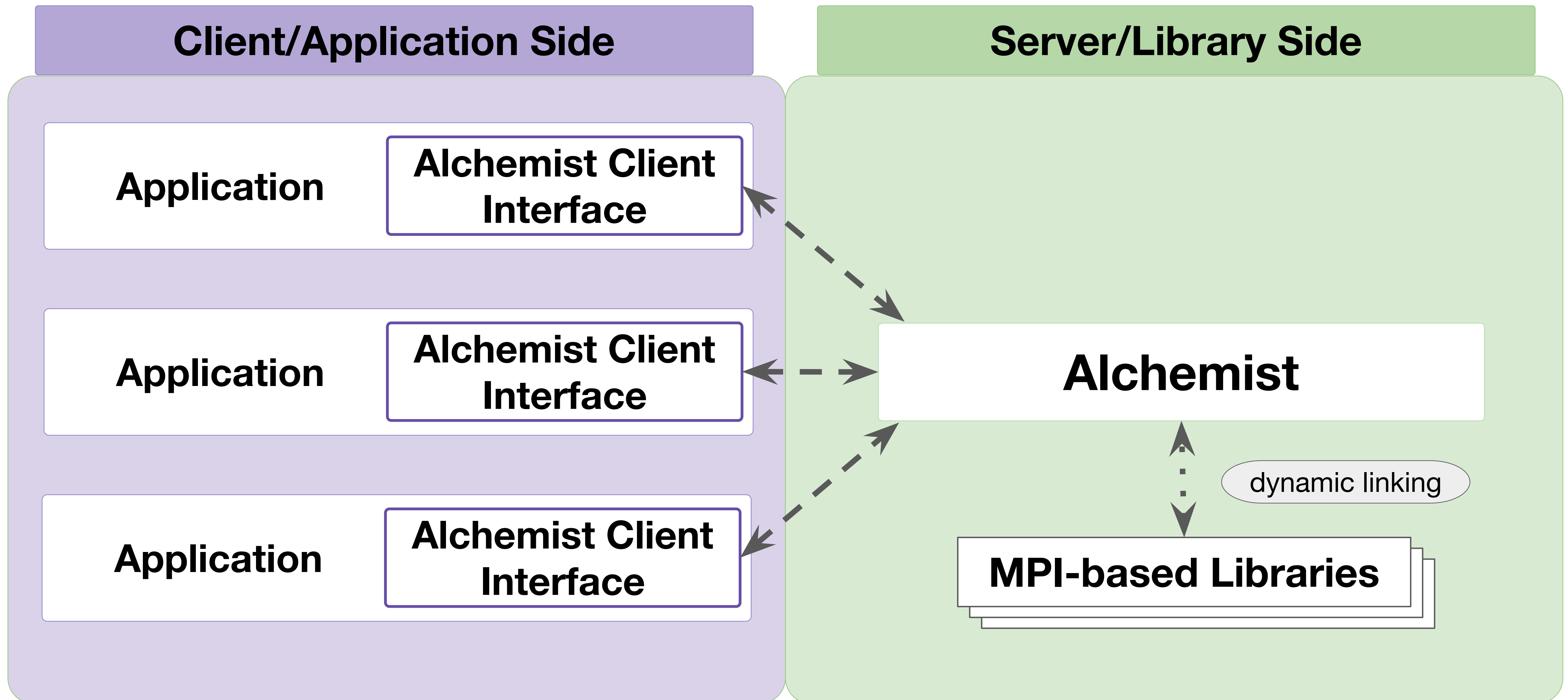
- **X ↔ MPI Interface**

- **More Functionality**

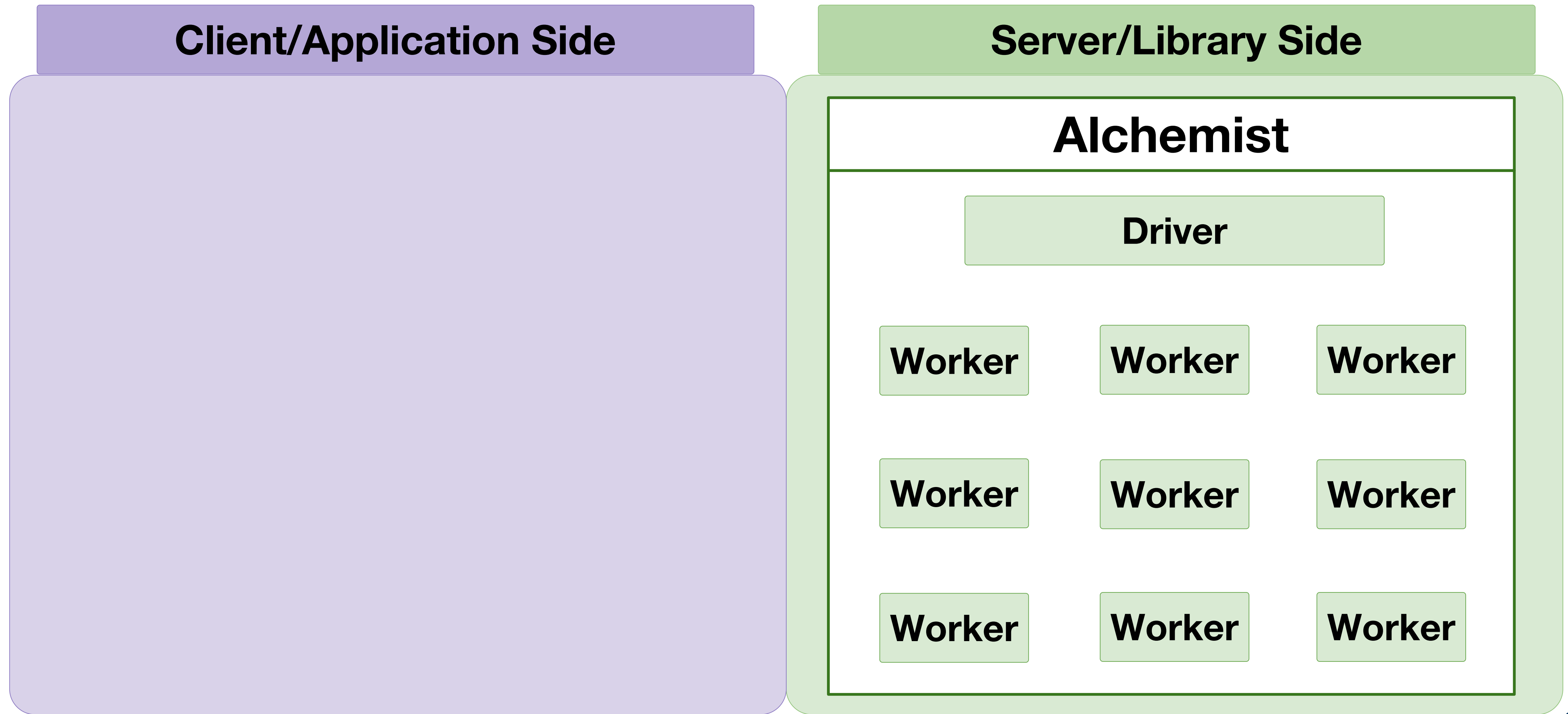
- **Alchemist and Containers**



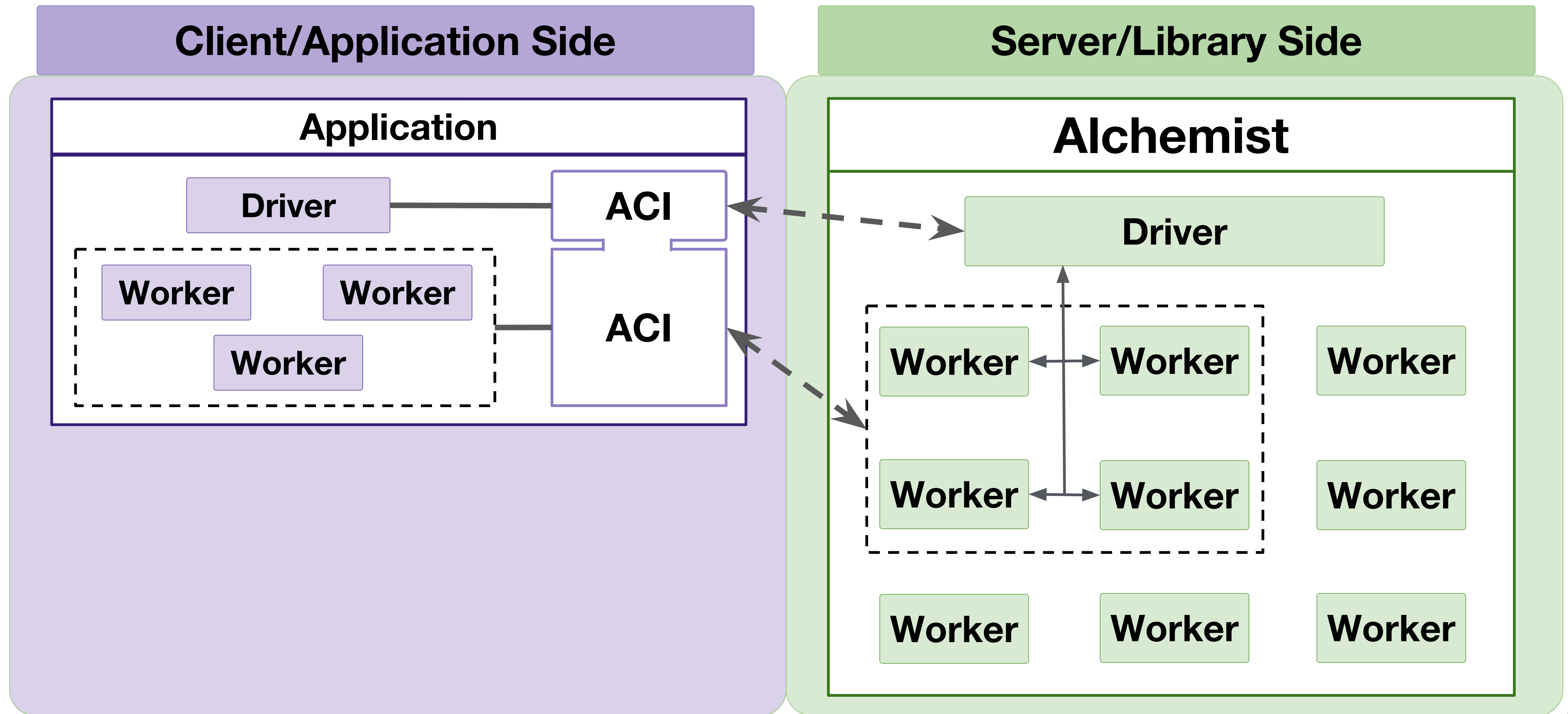
Alchemist Architecture



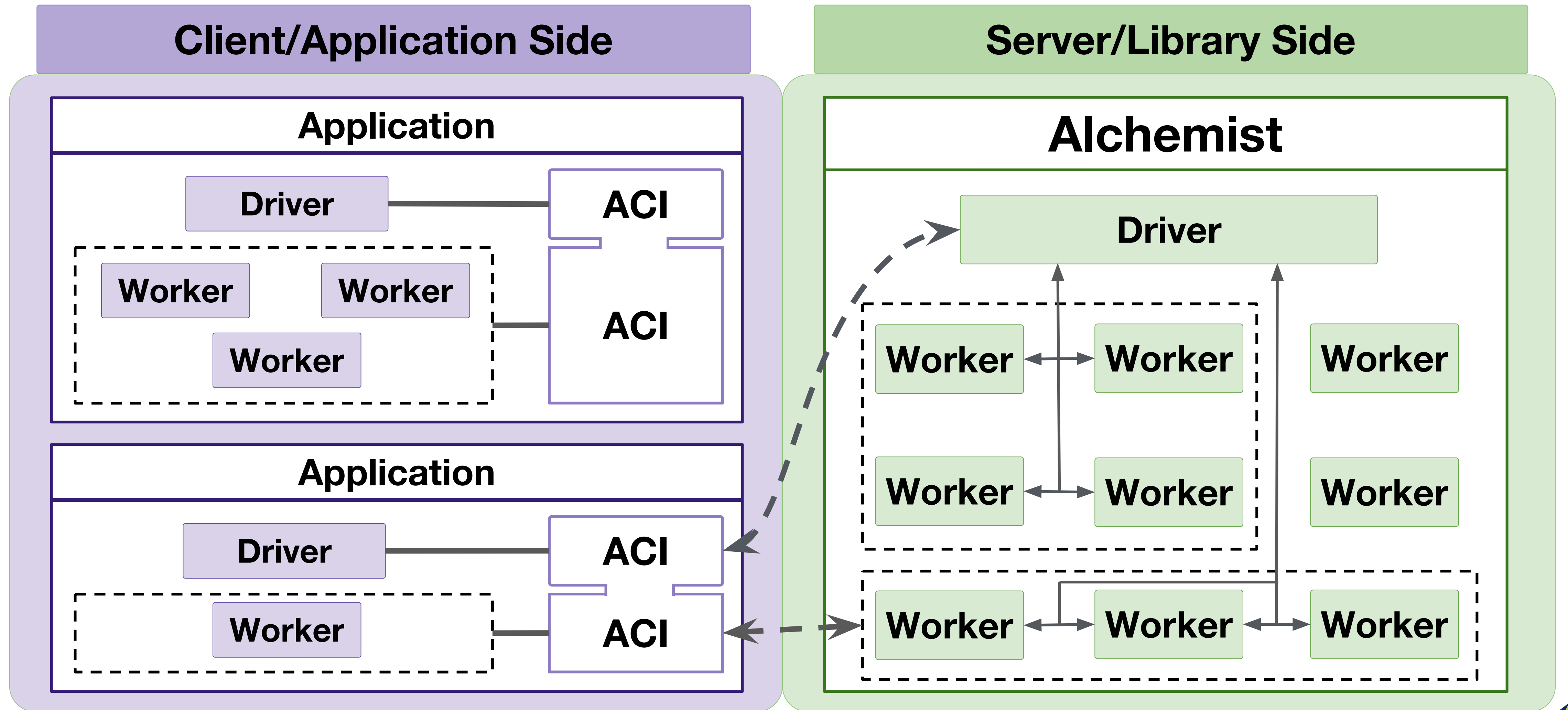
Alchemist Architecture



Alchemist Architecture



Alchemist Architecture



New and Upcoming Features

- **Server-Based Architecture**
- **X ↔ MPI Interface**
 - No longer limited to Apache Spark and Scala
 - Under development: Interfaces for Python, C++
 - Under development: Other data processing/ML frameworks
- **More Functionality**
- **Alchemist and Containers**



New and Upcoming Features

- **Server-Based Architecture**
- **X ↔ MPI Interface**
- **More Functionality**
 - Support for sparse matrices
 - Support for MPI-based libraries built on ScaLAPACK
- **Alchemist and Containers**



New and Upcoming Features

- **Server-Based Architecture**
- **X ↔ MPI Interface**
- **More Functionality**
- **Alchemist and Containers**
 - Under development: Alchemist running in Docker and Kubernetes
 - Will enable running Alchemist on the cloud



Limitations and Constraints

- Two copies of data in memory, possibly more during a relayout
- Lack of MPI-based libraries for machine learning
 - No equivalent to MLlib currently available

Future Work

- Apache Spark \Leftrightarrow X Interface
 - Interest in connecting Spark with other libraries for distributed computing (e.g. Cray Chapel, Apache REEF)
- MPI computations with fault tolerance and elasticity
- Run as network service



Kai Rothauge ▶

kai.rothauge@berkeley.edu ▶

github.com/kai-rothauge/alchemy ▶

Current Release:

github.com/alexgittens/alchemy

New Release (expected late May):

github.com/project-alchemist/