

# *Altocumulus*: Harvesting Computational Resources from Devices at the Edge of the Cloud

---

YoungHoon Jung, Marcin Szczodrak,  
Richard Neill, and Luca Carloni

Department of Computer Science  
Columbia University in the City of New York

# ***Altocumulus*: Harvesting Computational Resources from Devices at the Edge of the Cloud**

---

- 1. Trends in Cloud Computing and Embedded Systems**
- 2. The Altocumulus Idea**
- 3. Our Research Efforts Toward Altocumulus**

# Rising Embedded Systems amid the Clouds



## Cloud World

Amazon starts **Web-based Retail Services**

**Hadoop** is developed

Amazon opens Web Services (IaaS, PaaS)

**AppStore** for smart devices opens

Google App Engine starts

iCloud Service opens



## Embedded World

3G/IMT-2000 (**2Mbps**/384Kbps)

PDA has **300MHz** CPU clock frequency

2004

DOCSIS 2.0 deployment (43Mbps)

iPhone is released

Android phone is released

1Ghz Mobile CPU Clock

2010

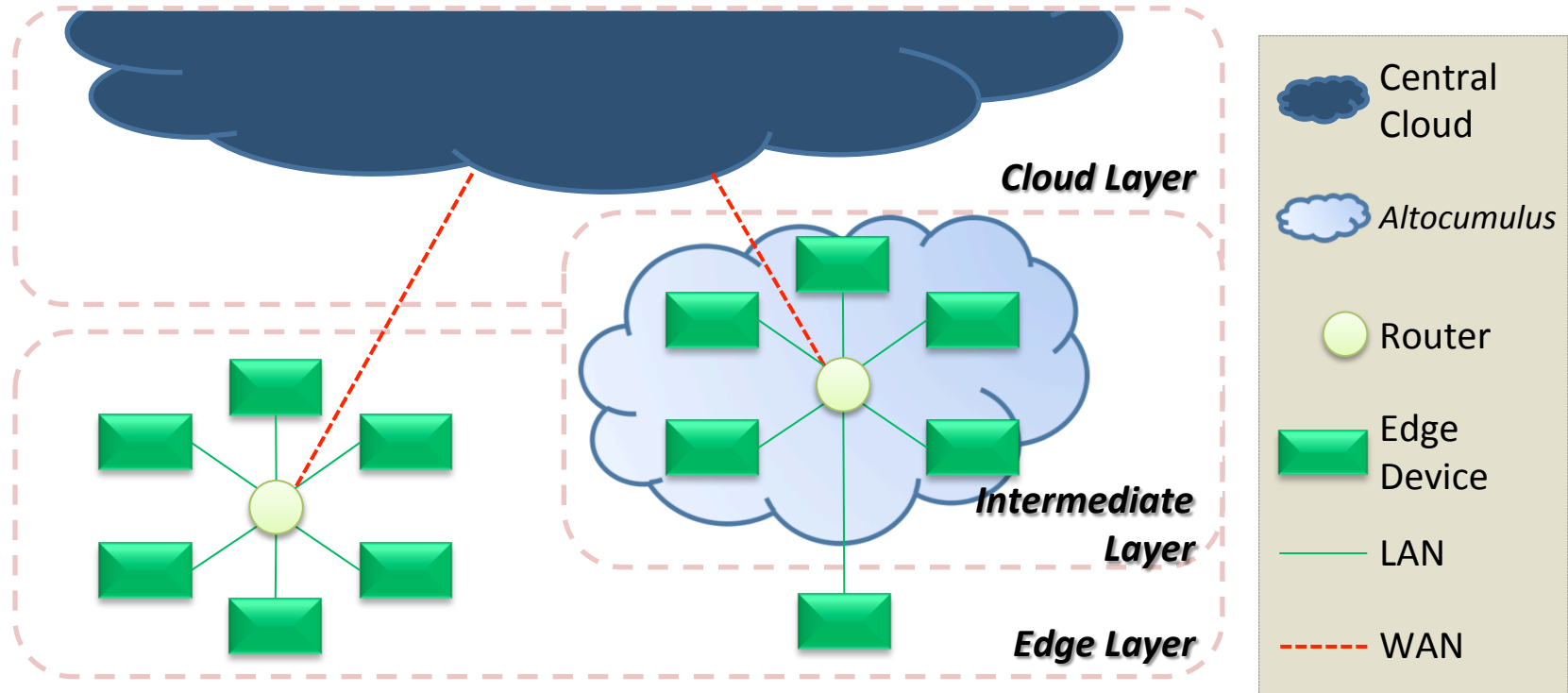
DOCSIS 3.0 (**250Mbps**), LTE (**300Mbps**)

2012

**2Ghz** Mobile CPU Clock

2014

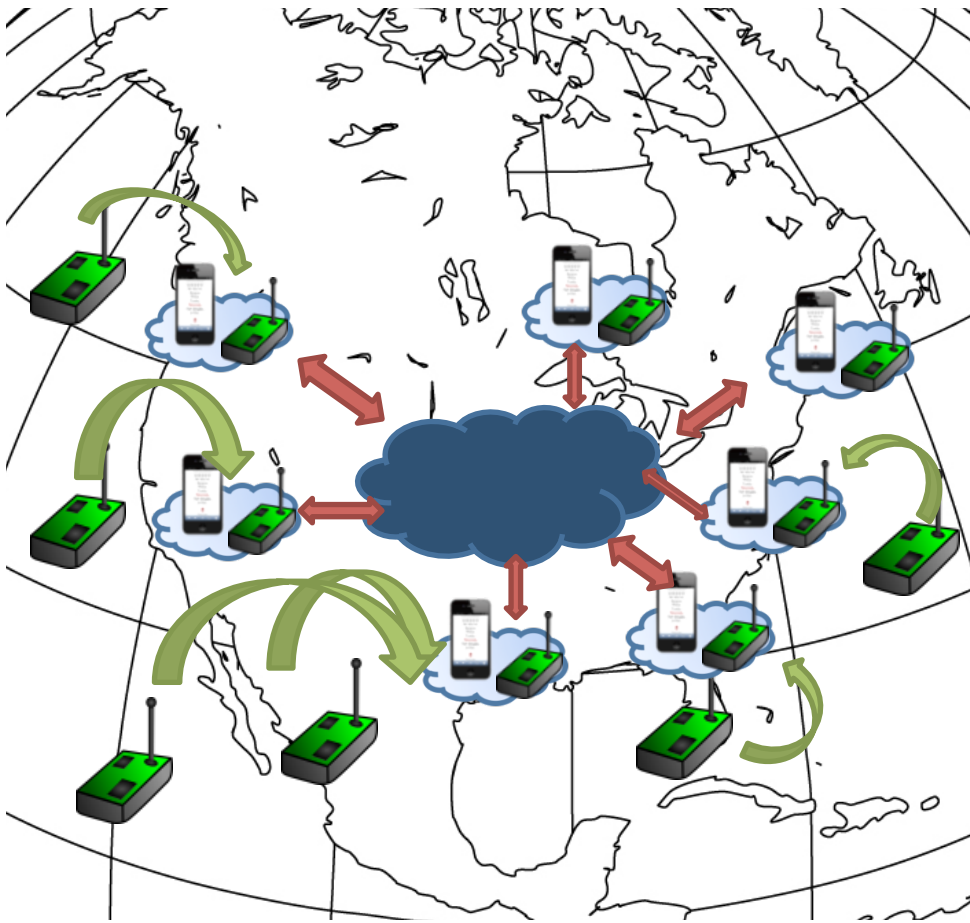
# Altocumulus: an Intermediate Cloud



- **A cloud formed by collecting resources from edge devices**
  - Small-Sized
  - Less Costs
  - Short Latency due to Close Distance
  - Pre-processing, Data Compression, Edge-Analytics

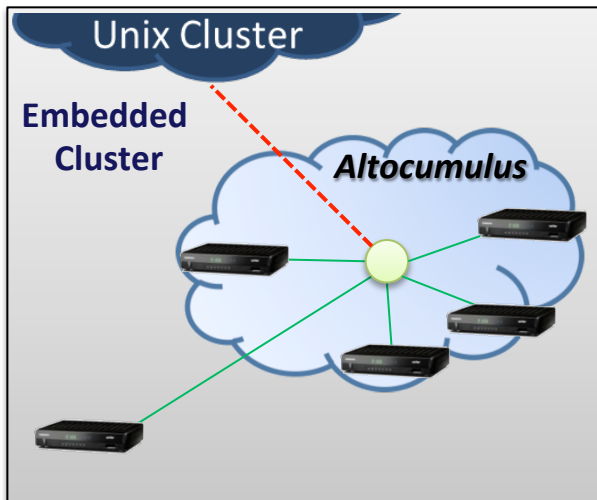
# Altocumulus: a Potential Application

- **Weather Data Analysis**



- **Traditional Approach**
  - Analyzes the National Climatic Data Center (NCDC) dataset
    - Sky ceiling height, Visibility distances, Temperature, and Atmospheric pressure
  - From more than 10,000 Weather Stations
  - **Central Data Processing**
- **With Altocumulus**
  - Regional Altocumuli
  - Altocumulus does data-collecting, pre-processing
  - **Local + Central Processing**

# OpenMPI on a Broadband Network of STBs



## A Heterogeneous System

- A Unix Cluster
- An Embedded Cluster with 64 Set-top Boxes
- DOCSIS 2.0 Network
- Complete Vertical Subset of Cable Head-end System



## DNA Sequencing Case Study

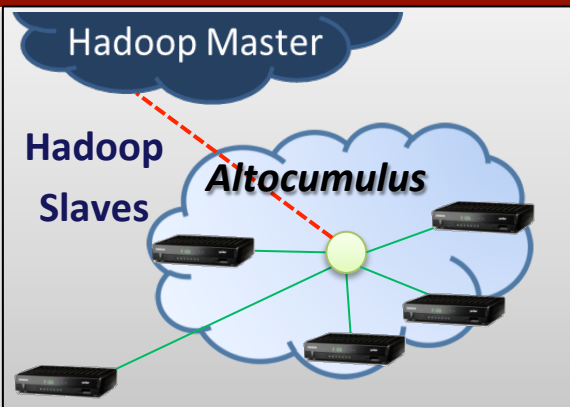
- Workers execute Parallel MSA Algorithm
- Distributed Computing using Open MPI Operations
- Scales similarly to the Unix Cluster

```
ATGGTTGCTACGCTTCGCAGCATTACCAT
--GGT-GCT-CG--TCGC-----CCAGCAAGG
ATCGTTCCT-----CAG-----
TGAAAGGAT
GAGGTGGATA-----
CCGTGTCAGGCATGTA
GAGCT---T-----CC----A--CATGT-
GGCCTG
GAGGTTGCTAGT--CCCATCGCATTGAACA
```

R. Neill, A. Shabarshin, and L. P. Carloni, "A Heterogeneous Parallel System Running Open MPI on a Broadband Network of Embedded Set-Top Devices", ACM International Conference on Computing Frontiers (CF) 2010

R. Neill, L. P. Carloni et al., "Embedded Processor Virtualization for Broadband Grid Computing", IEEE/ACM International Conference on Grid Computing (Grid) 2011

# MapReduce Computing System on an Embedded Cluster



- **Hadoop**

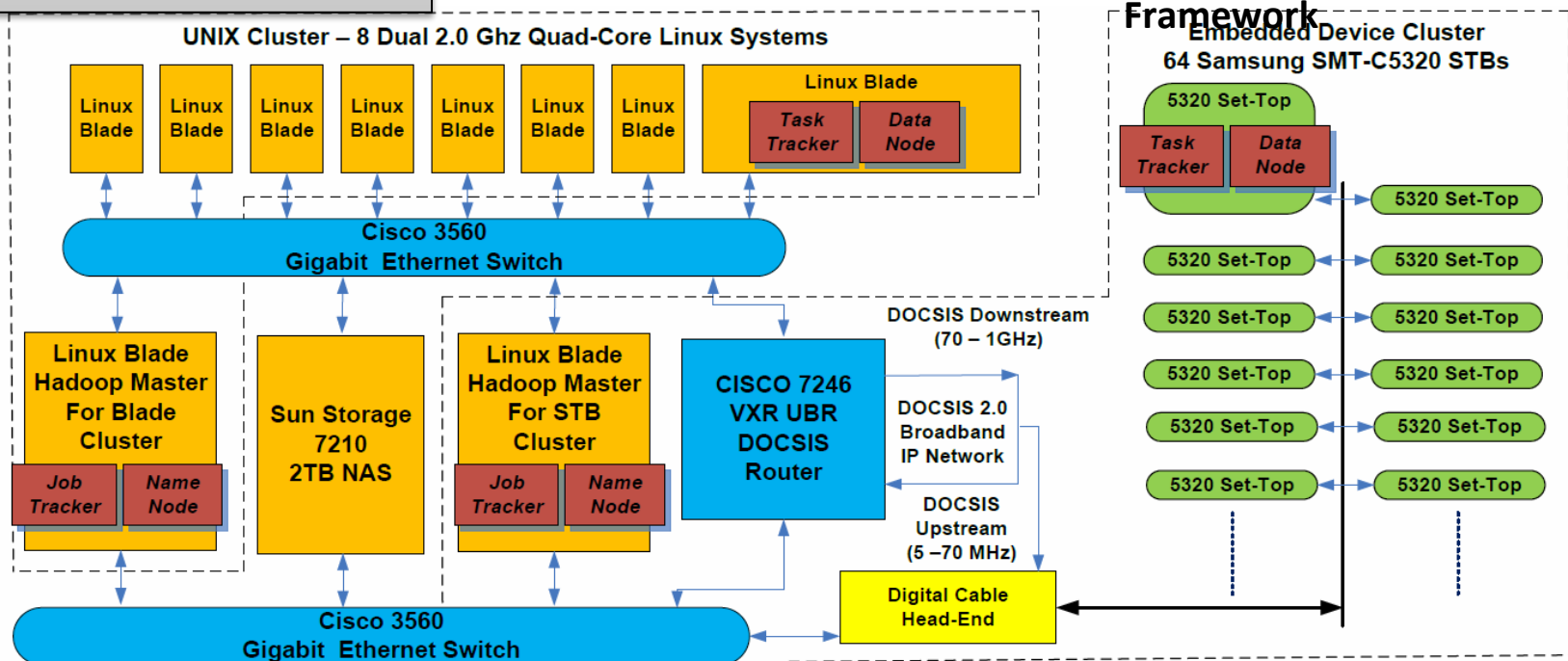
- A Distributed Data Processing Framework
- For Large Data Sets
- High-Available

- **Hadoop Distributed File System (HDFS)**

- Replicated, Reliable, and Scalable

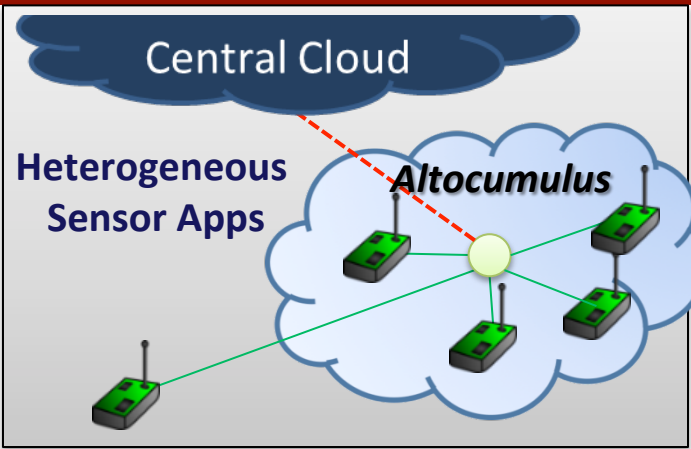
- **MapReduce**

- Simple Parallel Data Processing Framework



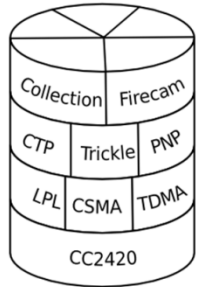
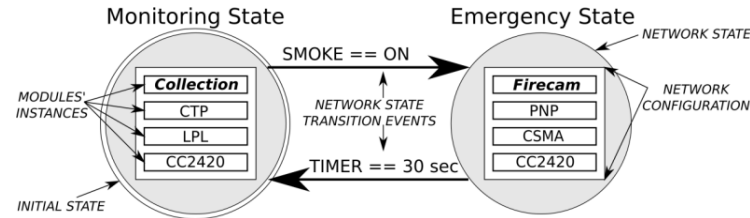
Y. Jung, R. Neill, and L. P. Carloni, "A Broadband Embedded Computing System for MapReduce Utilizing Hadoop", IEEE International Conference on Cloud Computing Technology and Science (CloudCom) 2012, Best Paper Award

# Fennec FOX: Enabling execution of multiple heterogeneous applications on the low-power wireless network



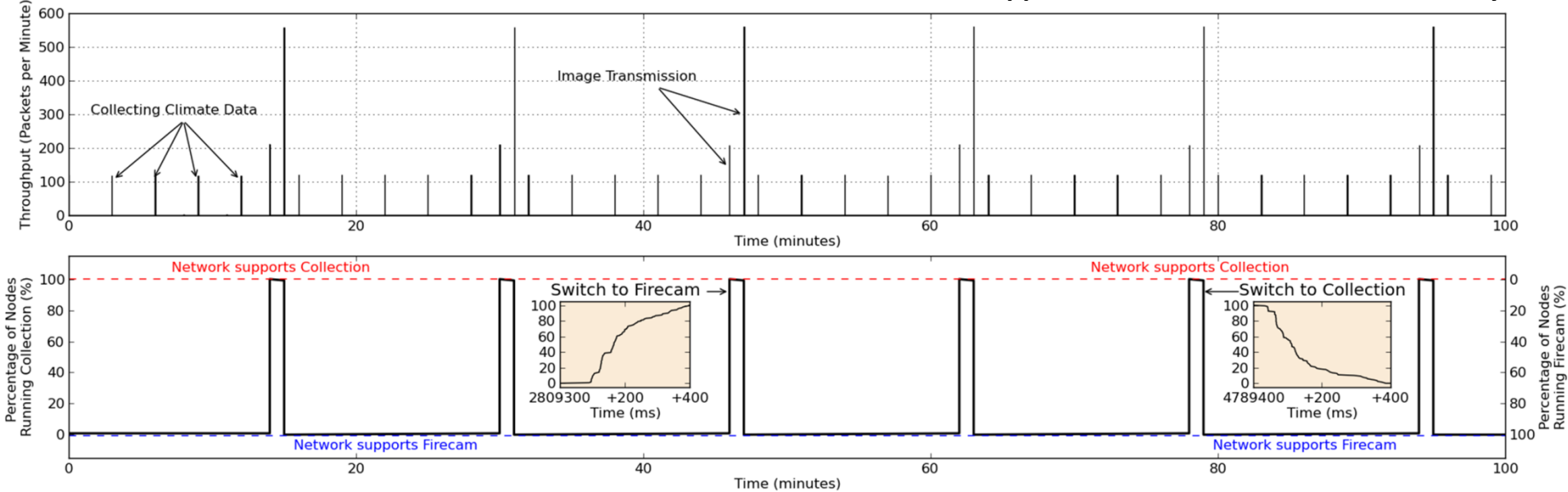
## Programming Language – Swift Fox

- Models run-time execution as finite state machine



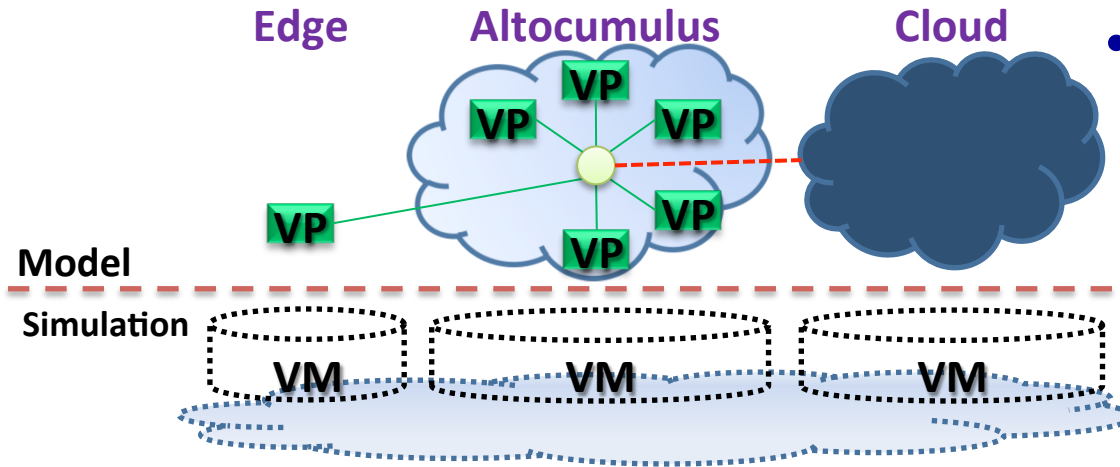
## Protocol Stack

- layered system design of application and communication protocols
- consists of application, network, MAC and radio layers

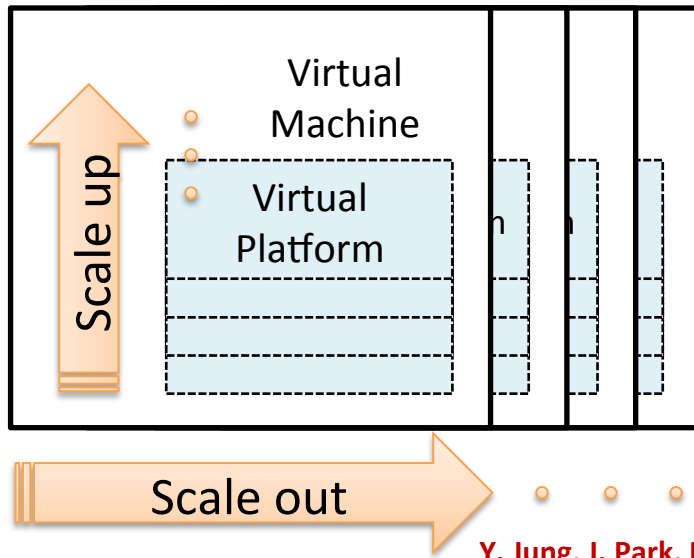




# netShip: a CAD tool for Large-scale, Heterogeneous Distributed Embedded Systems



- **A Simulation Environment**
  - For Altocumulus
  - Heterogeneity
  - Scalability
  - Based on virtual platform
  - Supports Design, Test, Simulation



- **The VP-on-VM model**
  - **Horizontal Scalability**
    - By adding VM
    - Uses dynamic VM creation
  - **Vertical Scalability**
    - By adding VP
    - Uses dynamic resource management

Y. Jung, J. Park, M. Petracca, and L. P. Carloni, "netShip: A Networked Virtual Platform for Large-Scale Heterogeneous Distributed Embedded System", Design Automation Conference (DAC) 2013

# Conclusion

---

- **Altocumulus is**
  - Small-sized
  - Low-latency
  - Low-costs
  - On the edge of the cloud
- **We are developing supporting frameworks**
  - OpenMPI
  - Hadoop Distributed File System & MapReduce
  - FennecFox
- **and CAD tool for Altocumulus systems**
  - netShip allows us to design, simulate, and analyze heterogeneous applications without physical deployment