

# 2019 Retreat



PLATFORMLAB

# Thank You, Sponsors!



**FUTUREWEI**



**ciena**

**NEC**

**vmware**

**SIEMENS**

**FUJITSU**

**Google**

**XILINX**

**SGX**

**BROADCOM**

# Platform Lab Faculty



**Bill Dally**  
Architecture



**Sachin Katti**  
Networking



**Christos Kozyrakis**  
Architecture,  
System Software



**Phil Levis**  
Embedded Systems



**Nick McKeown**  
Networking



**John Ousterhout**  
Granular Computing  
(Fac. Director)



**Guru Parulkar**  
Networking  
(Exec. Director)



**Balaji Prabhakar**  
Networking



**Mendel Rosenblum**  
Distributed Systems,  
Networking



**Keith Winstein**  
Networking,  
Granular Apps



**Matei Zaharia**  
Big Data,  
Cloud Computing

# Lab mission:

define new hardware/software  
platforms

that enable exciting new classes  
of applications

# State of the Lab

- **14 affiliate companies**
- **11 faculty in EE and CS**
- **Dozens of research projects**

# Theme #1: Self-Programming Networks

- **Rethinking networking to raise the level of programming**
- **Desired network behavior described in high-level, declarative fashion**
- **Low-level control actions determined automatically using**
  - Desired behaviors
  - Measurements of current usage
- **Example Platform Lab projects:**
  - Applying machine learning to network control
  - Rethinking the management and usage of clocks
  - Using smart NICs

# Theme #2: Granular Computing

- **Building large-scale datacenter applications out of large numbers of very short-lived tasks**
- **Why?**
  - Analyze large datasets in real time
  - Manage robot swarms operating at millisecond timescale
- **Example Platform Lab projects**
  - Infrastructure and applications for today's serverless computing platforms (e.g. AWS Lambda: 100 ms granularity tasks)
  - Low-latency communication and dispatching
  - Millisort: sorting at scale in 1 ms

# Theme #3: Machine Learning

- **Using machine learning to build better systems:**
  - Self-programming networks
  - Improved video streaming
  - Machine learning for robotic control
- **Systems to enable better machine learning:**
  - New hardware and software approaches to accelerate machine learning
  - Distributed inference



# How We Work

- **All work is open, freely accessible**
- **Goal: results not just interesting, but actually **useful****
- **We love industrial collaboration**
  - Interested in your ideas and feedback
  - Looking for opportunities to transfer technology

# News

- **Fintech symposia (Prabhakar, Rosenblum)**
- **O-RAN Alliance ([www.o-ran.org](http://www.o-ran.org)): global consortium creating software defined architecture for mobile radio networks (Katti)**
- **New projects:**
  - Inference as a service (Kozyrakis)
  - Machine learning for the Internet (Levis, Winstein)

# Awards

- **Inaugural ACM Thacker Award: Rosenblum**
- **Sloan Fellowship: Winstein**
- **SIGCOMM “Test of Time” Award: Katti et al. (XORs in the air)**
- **Caltech Distinguished Alumni Award: Dally**
- **SIGCOMM SOSR Software Systems Award: McKeown et al. (NetFPGA)**

## **Best Paper Awards:**

- **Pantheon: Yan, Levis, Winstein (USENIX ATC)**
- **Darwin: Turakhia, Dally, Bejerano (ASPLOS)**

# Recent/Imminent PhDs

|                              |  |                          |
|------------------------------|--|--------------------------|
| <b>Grant Ayers (CK)</b>      | <b>Optimizing Processor Architectures for Warehouse-Scale Computers</b>                | <b>TBD</b>               |
| <b>Assaf Eisenman (SK)</b>   | <b>Reducing DRAM Footprint to Scale Data Store Systems</b>                             | <b>TBD</b>               |
| <b>Lavanya Jose (NM)</b>     | <b>Proactive Congestion Control</b>  | <b>Google</b>            |
| <b>Ana Klimovic (CK)</b>     | <b>Fast, Elastic Storage for the Cloud</b>   | <b>ETH Zurich</b>        |
| <b>Manikanta Kotaru (SK)</b> | <b>Harnessing Nature to Make Wireless Positioning Practical and Accurate</b>           | <b>TBD</b>               |
| <b>Behnam Montazeri (JO)</b> | <b>Designing Datacenter Transports for Low Latency and High Throughput</b>             | <b>Google</b>            |
| <b>Henry Qin (JO)</b>        | <b>Arachne: Improving Latency and Utilization through Core Aware Thread Management</b> | <b>Square</b>            |
| <b>Hang Qu (PL)</b>          | <b>High Performance Cloud Computing</b>  | <b>Facebook</b>          |
| <b>Chinmayee Shah (PL)</b>   | <b>Automatically Distributing and Load Balancing Graphical Fluid Simulations</b>       | <b>Zoox</b>              |
| <b>Yatish Turakhia (WD)</b>  | <b>Hardware Acceleration of Genomic Sequence Alignment and Applications</b>            | <b>UCSC post-doc</b>     |
| <b>Feiran Wang (BP)</b>      | <b>Building High-Performance Distributed Systems with Synchronized Clocks</b>          | <b>TBD</b>               |
| <b>Lisa Yan (NM)</b>         | <b>Tools to Help Teach Large Classes</b>   | <b>Stanford lecturer</b> |

# Thursday Agenda

|      |  |   |
|------|--|---|
| 1:00 | Welcome, Introductions, Platform Lab Overview        | John Ousterhout                           |
| 1:30 | Distributed Inference and Learning                   | Sandeep Chinchali                         |
| 2:00 | INFaaS: Managed and Model-Less Inference Serving     | Franky Romero, Qian Li, Neeraja Yadwadkar |
| 2:30 | Smart Harvesting of Spare CPU Cores in Cloud Servers | Yawen Wang                                |
| 3:00 | Break  |   |
| 3:30 | OREO: Serverless Computing with Perf. Guarantees     | Neeraja Yadwadkar                         |
| 4:00 | Continual Learning Improves Video Streaming          | Keith Winstein                            |
| 4:30 | Millisort: An Experiment in Flash Bursts             | Yilong Li                                 |
| 5:00 | Lightning talks for posters                          |   |
| 5:30 | Reception/posters                                    |   |
| 7:00 | Dinner   |   |
| 8:30 | Evening activities                                   |   |

# Friday Agenda

|       |  |                           |
|-------|--|---------------------------|
| 9:00  | <b>The World Needs a New NIC (and it needs to run Homa)</b>                          | <b>John Ousterhout</b>    |
| 9:30  | <b>Enabling Affordable Precision Agriculture by Sensing Soil Moisture Wirelessly</b> | <b>Colleen Josephson</b>  |
| 10:00 | <b>Power Clocks: Dynamic Multi-Clock Management for Embedded Systems</b>             | <b>Holly Chiang</b>       |
| 10:40 | <b>Recreation and Informal Conversations</b>   |                           |
| 12:00 | <b>Lunch</b>   |                           |
|       | <b>SDN and Programmable Forwarding</b>   | <b>Nick McKeown</b>       |
| 1:30  | <b>Self-Programming Networks: from Sensing to Control</b>                            | <b>Balaji Prabhakar</b>   |
| 2:00  | <b>Using NICs and ECC to Improve Packet Transport</b>                                | <b>Shiyu Liu</b>          |
| 2:30  | <b>CRaft: Using Accurate Clocks to Build Multi-Leader Version of Raft</b>            | <b>Feiran Wang</b>        |
| 3:00  | <b>Break</b>   |                           |
| 3:30  | <b>Deep Learning Multi-Robot Task Assignment Policies for Flexible Manufacturing</b> | <b>Patrick Washington</b> |
| 4:00  | <b>Industrial Feedback</b>   |                           |

# Questions/Discussion

