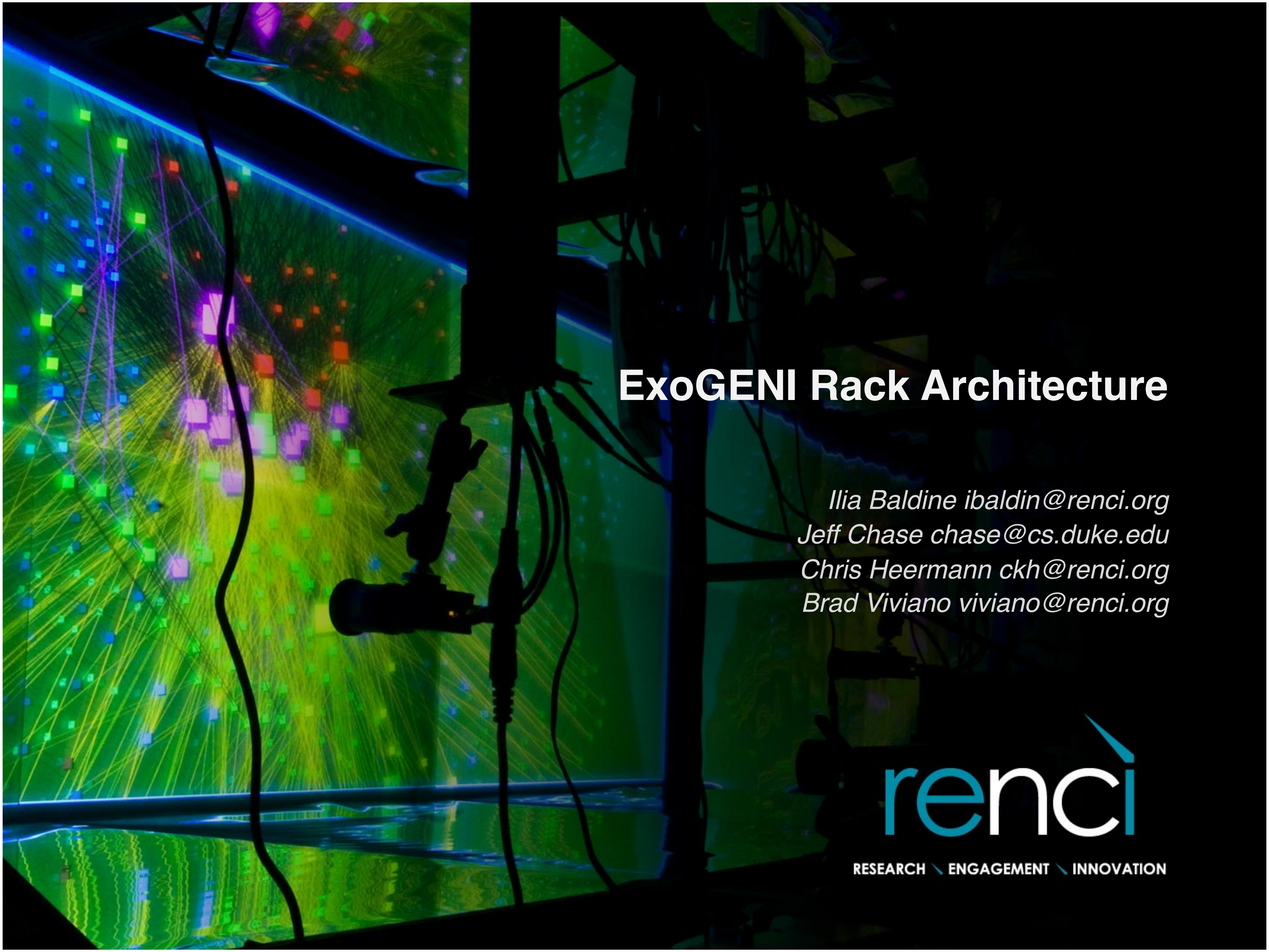


ExoGENI Rack Architecture

Ilia Baldine ibaldin@renci.org
Jeff Chase chase@cs.duke.edu
Chris Heermann ckh@renci.org
Brad Viviano viviano@renci.org

renci

RESEARCH \ ENGAGEMENT \ INNOVATION

Overview

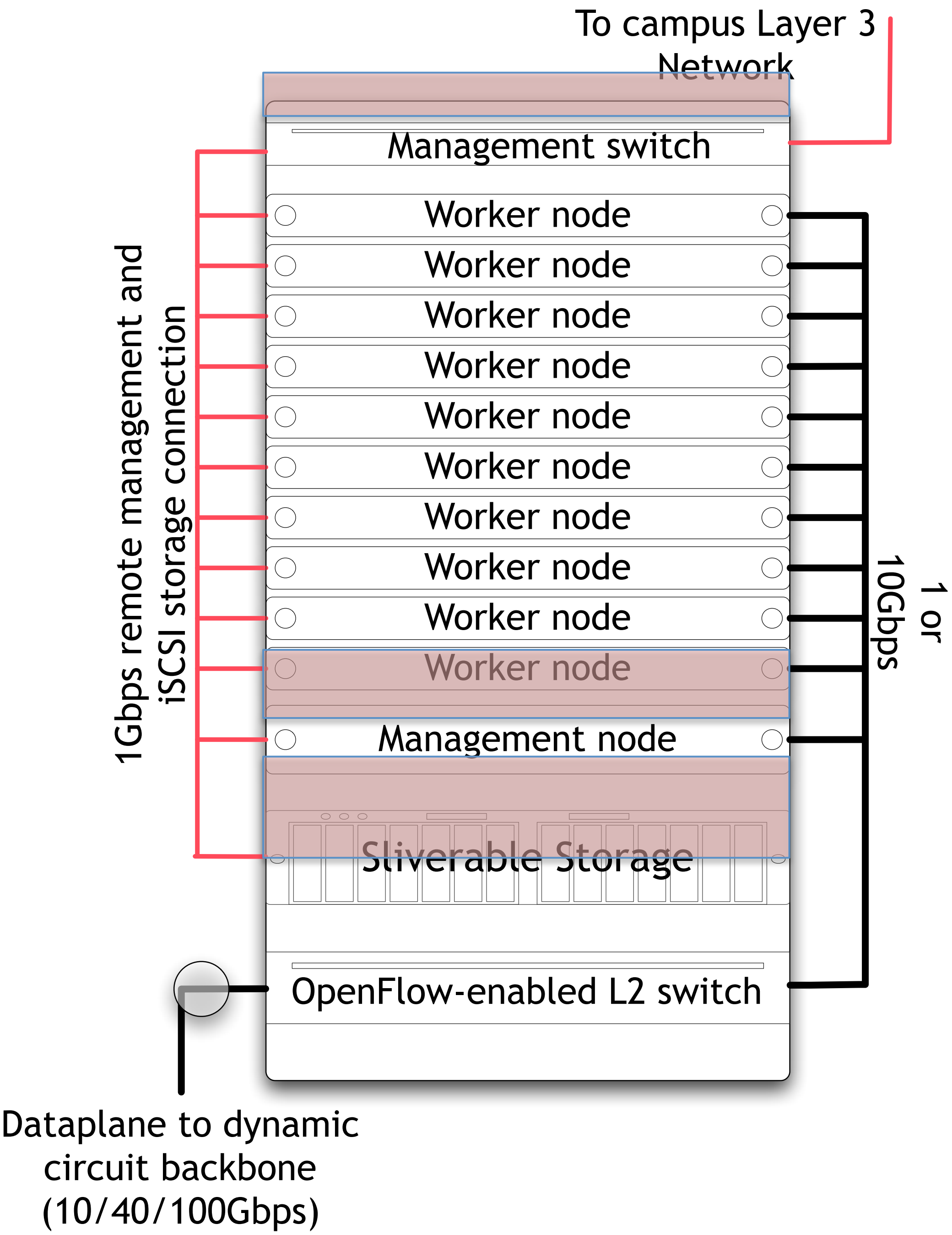
- Hardware architecture
- Software architecture
- Connectivity
- Remote management/Site Logistics/Usage
- Interaction with other projects

Introduction

- ExoGENI Racks: a partnership between RENCI, Duke and IBM
- Uses IBM x3650 M3 and M4 2U servers
  - Westmere PCIe II or SandyBridge PCIe III 4 or 6 core
- An OpenFlow switch (vendor TBD) with either
  - 1G ports with 10G uplinks
  - 10G ports with 40G uplinks
  - Depends on available bandwidth at the site
- Separate iSCSI storage for
  - User OS images
  - Measurement data
- Expandability
  - 2U servers with PCIe III for
    - GPGPUs, 10/40/100G NICs, NetFPGA 10G, ???

ExoGENI Rack

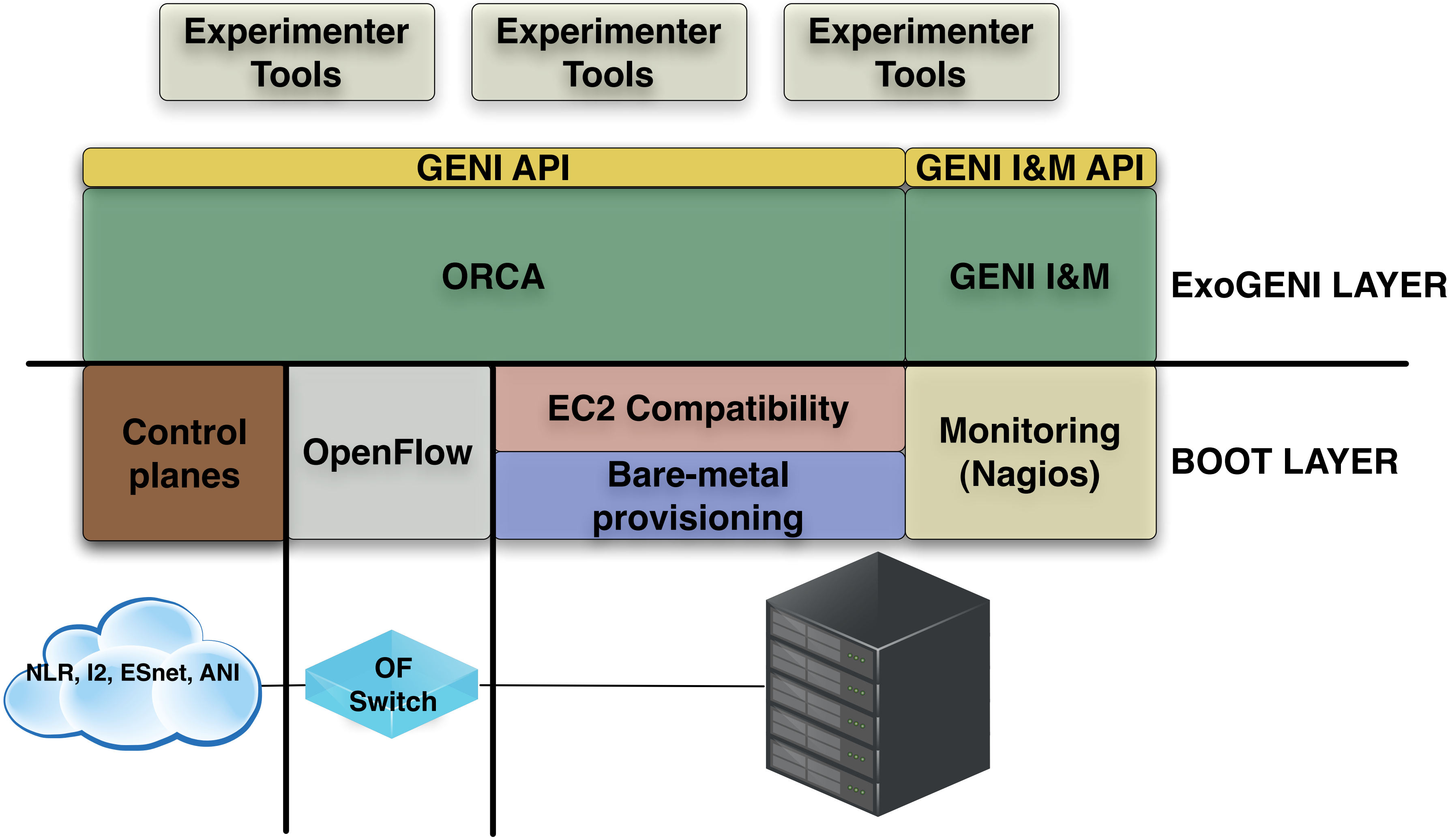
- Management node (no experimenter access)
- Worker nodes (sliverable)
  - Bare metal
  - Virtualized
- Management switch
- OpenFlow dataplane switch
- Sliverable storage



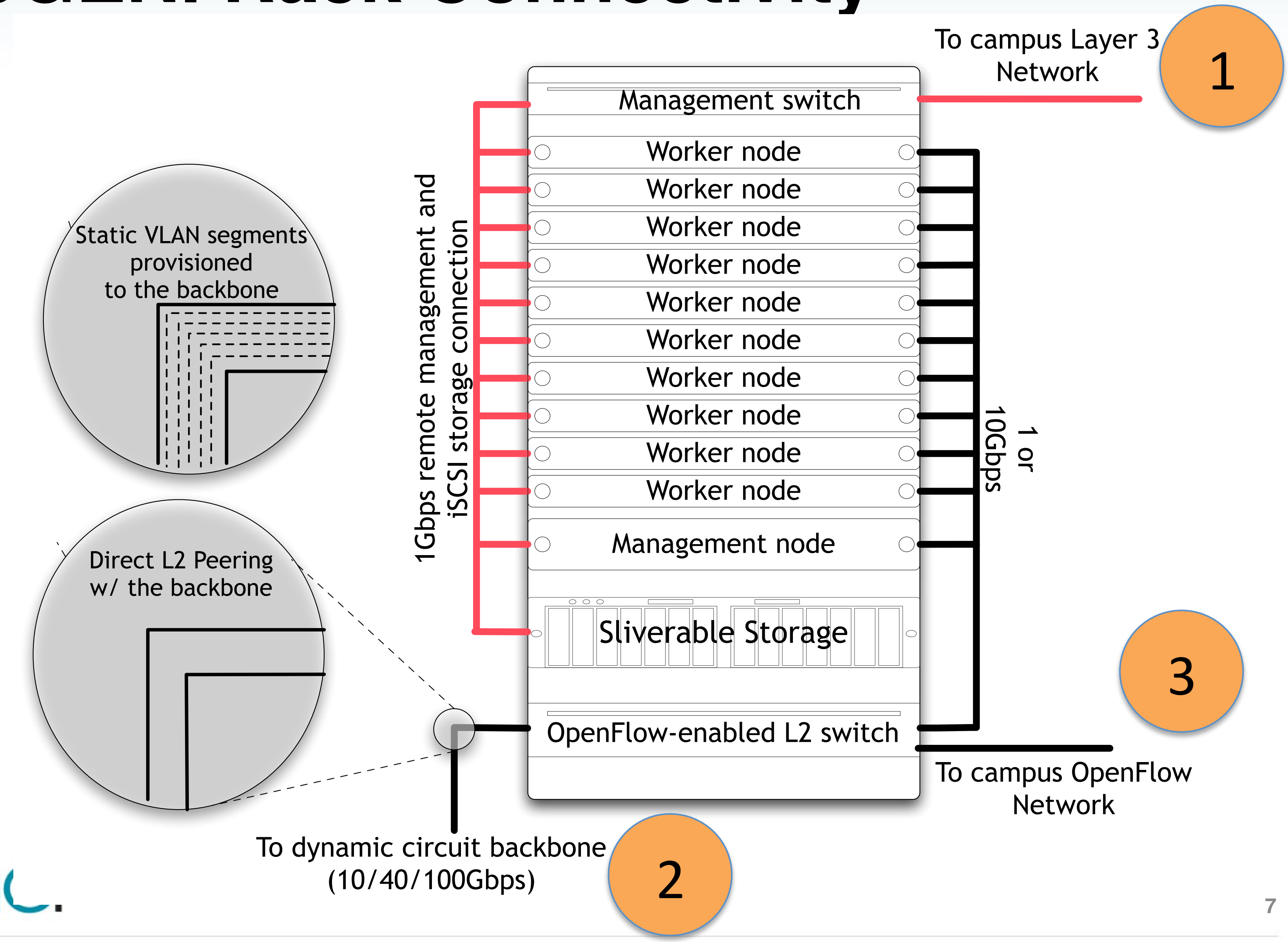
Software stack

- Design philosophy:
  - An immutable 'boot' layer built on stable tried-and true technologies
    - Resource provisioning (compute, network, storage)
    - Measurement (e.g. power via IPMI)
    - Remote management
  - *'Exo' [prefix] - **external, from the outside***
    - GENI software running on top of the boot layer uses exported provisioning functions to
  - Emphasis on virtualization technologies:
    - Hypervisors (KVM), SR-IOV
    - Hardware design goal: highest core count per rack within budget (not highest server count).
  - Use ORCA Control Framework

ExoGENI Software Stack



ExoGENI Rack Connectivity


To campus Layer 3 Network
1
Management switch
Worker node
Worker node
Worker node
Worker node
Worker node
Worker node
Worker node
Worker node
Worker node
Worker node
Management node
Sliverable Storage
OpenFlow-enabled L2 switch
1Gbps remote management and iSCSI storage connection
1 or 10Gbps
Static VLAN segments provisioned to the backbone
Direct L2 Peering w/ the backbone
3
To campus OpenFlow Network
To dynamic circuit backbone (10/40/100Gbps)
2

Connectivity details

1. Remote management connection via campus L3 network
   - Low traffic, secured with VPN
2. Rack has a connection to one of national research backbones: NLR, I2, ESnet, ANI
   - Rack has a claim on a significant fraction of 10G interface capacity
   - Some sites will offer 40G and 100G connectivity
   - Connection is direct or via RONs

- Rack either
  - has a claim on a pool of VLAN tags that are visible at the backbone negotiated with RON
  - has a direct connection to a dynamic circuit network (NLR Sherpa, I2 ION, ESnet/ANI OSCARS)
- Experiment topologies can be
  - Intra-site – contained within a single rack
  - Inter-site – spanning multiple racks over the L2 networks
- VLAN tag remapping when tags do not match across sites
  - Using RENCI-owned resources at RENCI and StarLight
  - OSCARS does it for some networks (w/ MPLS)
  - Work with NLR to integrate this into Sherpa

Connectivity details (continued)

3. Optionally rack has a connection into the campus OpenFlow network
   - Potentially bursty, high traffic demand originating from experiment slices
   - Needs negotiation with campuses for security and performance

Remote Management/Site Logistics

- Racks will be assembled/tested by IBM at the manufacturing or integration facility
  - Software pre-installed
  - Shipped directly to the site
- Extensive remote management/low remote hands/eyes requirements
  - Secure management network linking the racks
  - IBM MediaKey
  - IPMI 2.0
  - Remote power on/off

Uses

- Resources delegated via ORCA to various resource pools
  - GENI – for all GENI users
  - Local use – for local users
  - Others*
    - Sites are encouraged to purchase/add own compliant hardware to expand existing resource pools or create new ones to serve other domain science research.