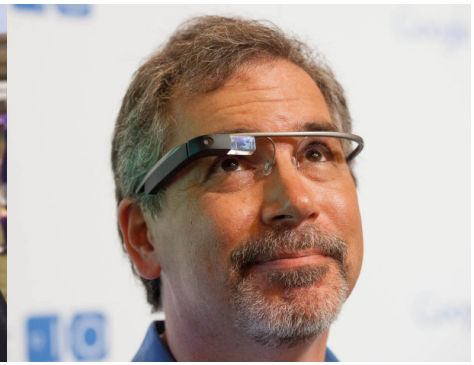
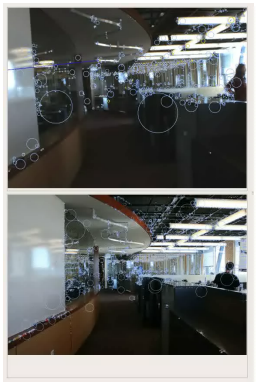
The Case for **ON**loading Continuous High-Datarate Perception to the **Phone**

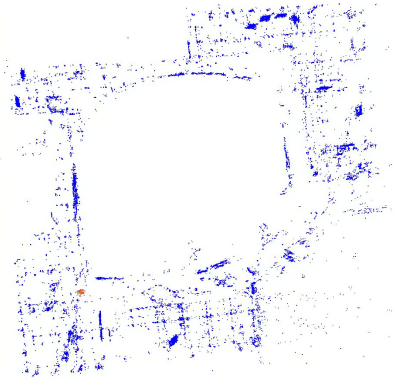
Seungyeop Han and Matthai Philipose

University of Washington and Microsoft Research





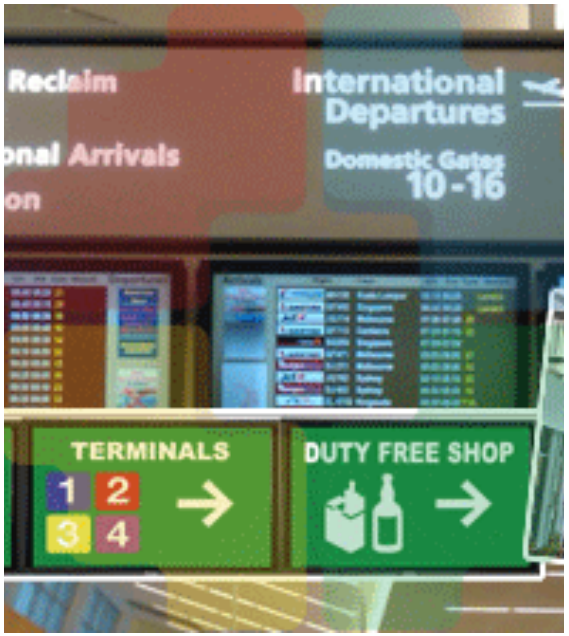




Who? **What?** **Where?**

It’s Worth It!

Change behaviors



Aid in complex tasks

Augment cognition

The Performance Gap

- Resource use of today's best systems:

	Speech	Face	Object	Location
Model Size	0.5 – 600 GB	1M/person	3-6 GB	20M/floor
Compute	16kHz @ 160% Xeon E5640	30fps @ 8-core SandyBridge	30fps @ Nvidia 580 GPU	30fps @ 2-core 2.4 GHz '09 Intel CPU

1000x CPU, 100x memory relative to today's phone

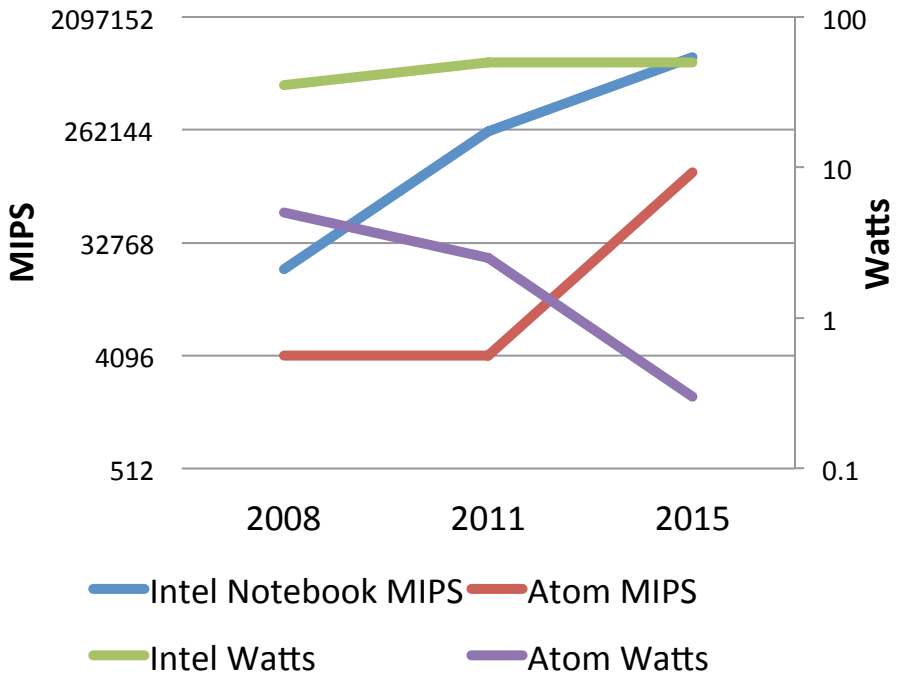
Offloading Does Not Bridge the Gap

Continuous vision requires:

- High availability
- High core network capacity
- High compute capacity
- A good privacy story

Bridging the gap

Processors getting more efficient



Gaining ~50x efficiency

Most bytes are irrelevant

Drop frames if	Fraction to process
None	100%
No voice	6%
High acceleration	2.5%
Low light	1.5%

Gaining ~50x efficiency

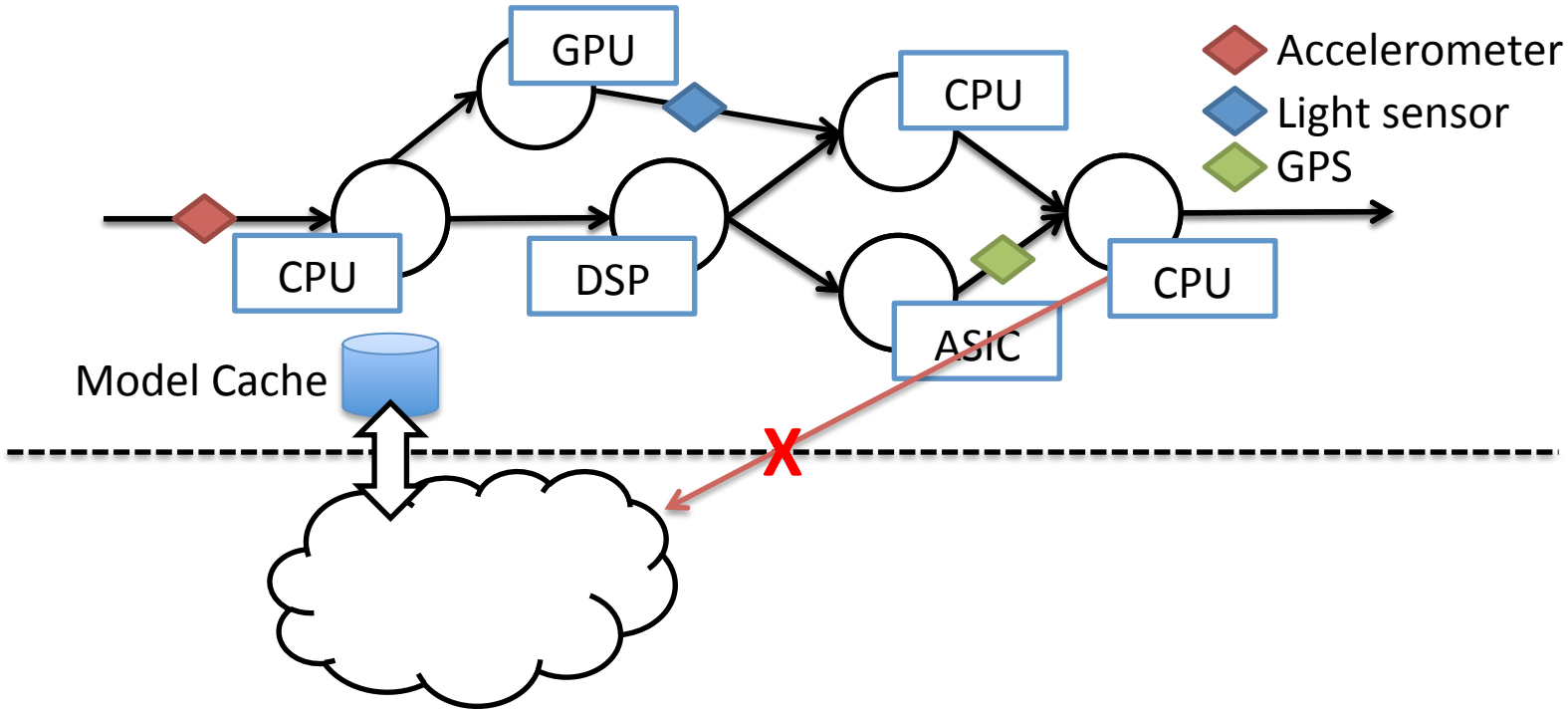
- Vision algorithms are reaching maturity
- Applications could add huge value
- Vision on the phone likely necessary + feasible

What does the OS do?

The OS Multiplexes Across Apps …

… while providing efficient access to:

- Heterogeneous resources
- Sensor gating services
- Cached cloud-scale models
- Privacy primitives



- Vision algorithms are reaching maturity
- Applications could add huge value
- Vision on the phone likely necessary + feasible
- OS provides efficient, private cross-app access:
  - processors, models, low-power sensors

THANKS!