

Rethink Possible



Industry Trends: Mo Money, Mo Problems

Graham Cormode
Principal Member of Technical Staff
AT&T Laboratories, Inc.



Driving Innovation

AT&T Labs



AT&T's patent portfolio includes

over 9,000
issued applications

Over 75%
of our researchers
hold a PhD degree

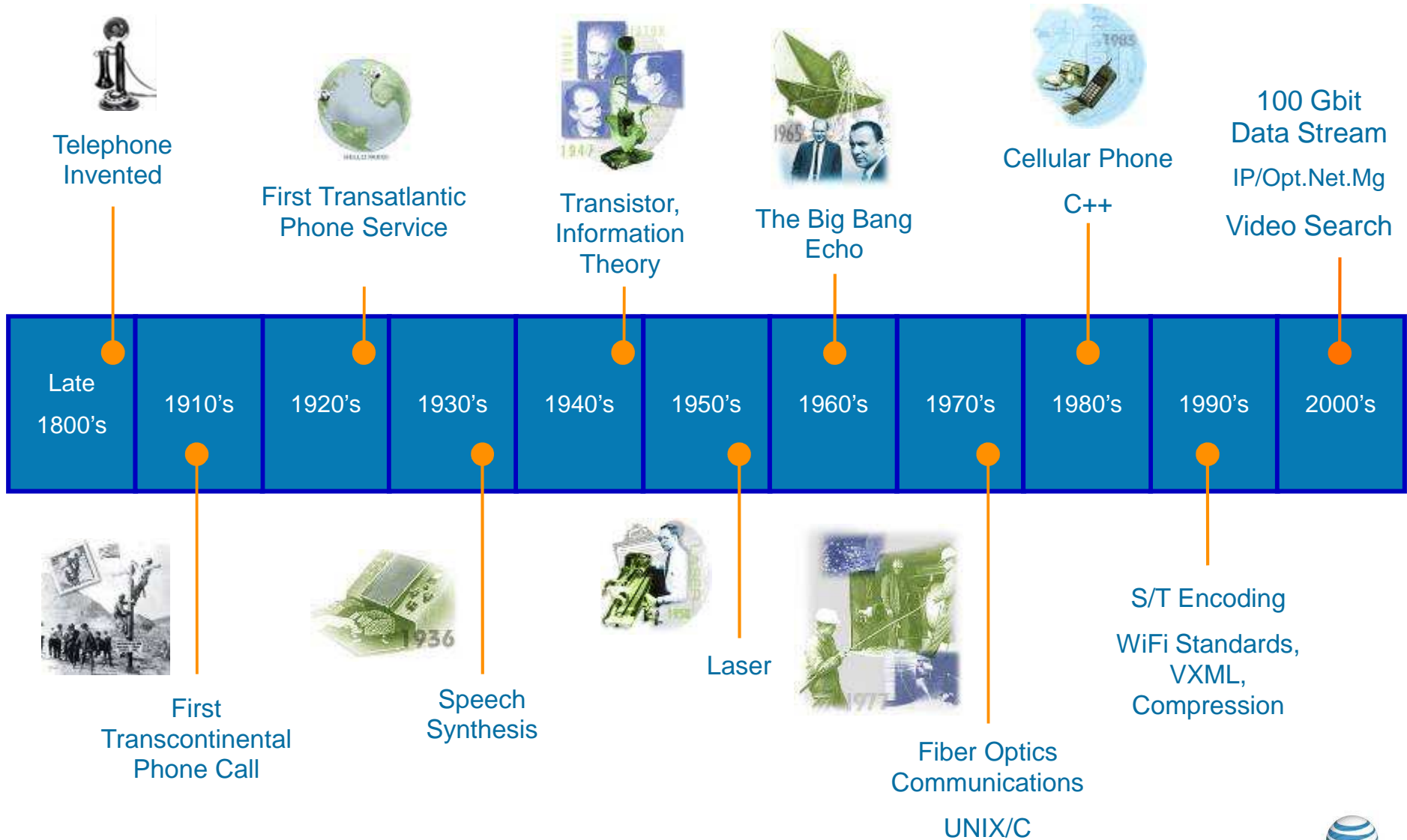
120 years
of technology breakthroughs
and product/service innovation

Over 1,300
of the world's best
scientists and
engineers

Heritage includes
9 Nobel Prizes

**The Innovation Engine Behind AT&T's
World-Class Technology**

Making History



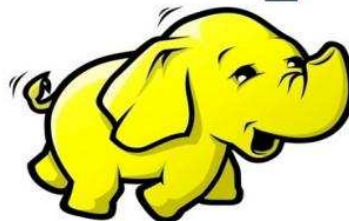
Outline

- ◆ Some trends in industrial research
 - May look very similar to trends in academic research
- ◆ Some **suggestions** for REU projects
- ◆ Some **perceived roadblocks** to doing them in academia
- ◆ Some ideas for **circumventing the roadblocks**
- ◆ Some REU experiences

Big Data Analytics

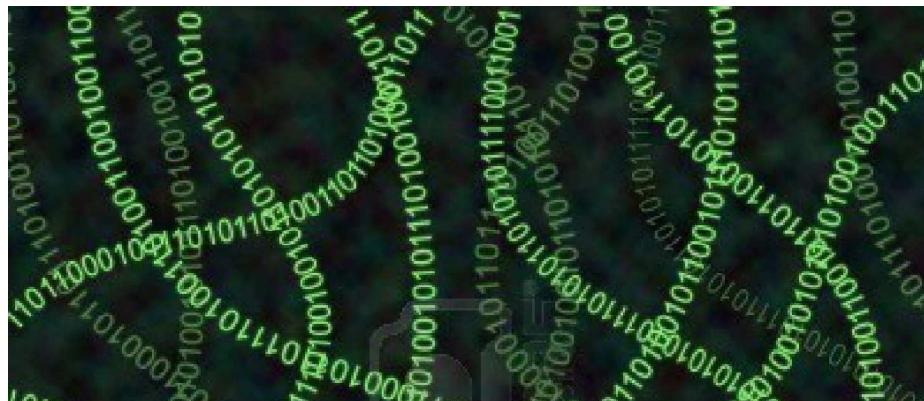
- ◆ Huge amounts of data generated by modern applications
 - All clicks on Facebook, all transactions on Amazon
 - All measurements from Large Hadron Collider
- ◆ New models of computation to process big data
 - MapReduce and **Hadoop**
- ◆ **Possible Project**: Scale your work to Hadoop
- ◆ **Problem**: Availability of large compute cluster / data center
- ◆ **Solution**: Use Amazon Web Services, or simulate on desktop

hadoop



Data Streams

- ◆ Streams of data arriving at high rates
 - Activity streams on social networks (Twitter, Facebook)
 - Traffic data on IP networks
- ◆ Need new algorithms that can keep pace with these streams
 - Keep **small summary** to deliver approximate answer
- ◆ **Possible Project**: mine streams for trends, unusual events
- ◆ **Problem**: don't have access to live data streams
- ◆ **Solution**: replay stored public stream captures



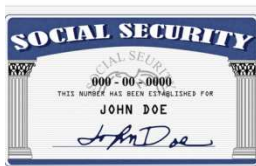
Data Science

- ◆ A new area of “data science” is emerging between CS and Stats
 - Large companies hiring ‘data scientists’ to analyze their data
 - “the sexy job in the next 10 years” -- Hal Varian, Google (2009)
- ◆ Need a mixture of stats, data handling, mining, learning
 - Make use of open source tools: [R](#), [Weka](#), Linux tools
- ◆ Possible Project: extract new intelligence from a dataset!
- ◆ Problem: Don’t have access to business data
- ◆ Solution: Study city/gov data: [data.gov](#), [data.cityofnewyork.us](#)



Privacy

- ◆ With all of this study of data, concern for individual's privacy
 - Many applications predicated on using detailed personal data
 - I don't want you knowing my finance, health, web data
- ◆ New models of privacy that transform data to mask individuals
 - **Differential privacy** adds statistical noise scaled to the application
- ◆ **Possible Project**: adapt existing algorithms to work privately
- ◆ **Problem**: don't want to get IRB approval for private data
- ◆ **Solution**: work with already anonymized data as ground truth



To	Number	Date/Time	Minutes	Rate	Total Amount
CONCORDIA UNIV	0075411000	02/05/07 14:54:30	4.5	0.006	0.027
CONCORDIA UNIV	0075411000	02/05/07 14:14:20	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 13:14:00	1.2	0.006	0.007
CONCORDIA UNIV	0075411000	02/05/07 08:12:00	2.7	0.006	0.016
CONCORDIA UNIV	0075411000	02/05/07 08:11:00	4.0	0.006	0.024
CONCORDIA UNIV	0075411000	02/05/07 08:10:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:09:00	3.5	0.006	0.021
CONCORDIA UNIV	0075411000	02/05/07 08:08:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:07:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:06:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:05:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:04:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:03:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:02:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:01:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 08:00:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:59:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:58:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:57:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:56:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:55:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:54:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:53:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:52:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:51:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:50:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:49:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:48:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:47:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:46:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:45:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:44:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:43:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:42:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:41:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:40:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:39:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:38:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:37:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:36:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:35:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:34:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:33:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:32:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:31:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:30:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:29:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:28:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:27:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:26:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:25:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:24:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:23:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:22:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:21:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:20:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:19:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:18:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:17:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:16:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:15:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:14:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:13:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:12:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:11:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:10:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:09:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:08:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:07:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:06:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:05:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:04:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:03:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:02:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:01:00	2.5	0.006	0.015
CONCORDIA UNIV	0075411000	02/05/07 07:00:00	2.5	0.006	0.015



Smartphone Apps

- ◆ Smartphones are sophisticated sensors carried by millions
 - A powerful computing platform in the pocket
 - Hundreds of apps released every day
- ◆ Can **build apps** to help people in their daily lives
 - Measure environment; learning; accessibility; communicate
- ◆ **Possible Project**: Build app for problems you meet in life
- ◆ **Problem**: can't afford to buy latest smartphone for REU
- ◆ **Solution**: use PC-based simulators to develop and test app



Social Web

- ◆ Over 1 Billion people use online social networks (OSNs)
 - General purpose social networks, but also specialist
 - Many people's communication primarily through OSNs
- ◆ All the previous questions apply to the OSN setting
 - Algorithms, Streams, Analytics, Privacy, Apps...
- ◆ Possible Project: apply one of the previous questions to OSNs
- ◆ Problem: don't know much about OSNs
- ◆ Solution: Your REU can teach you what you need to know



My REU Experiences

- ◆ Involved with DIMACS REU since 2003, worked with 8 students
- ◆ **Best experience:**
 - REU solved algorithmic problem in data streaming
 - Student went on to a top grad school
 - We worked on more problems together, 5 papers & counting...
- ◆ **Worst experience:**
 - Productive summer, generated many ideas and claims
 - Student went back to uni promising a write-up
 - The write up never materialized...
- ◆ **Possible Project:** classifier for which category a student falls into

Rethink Possible

