Can cellular networks handle 1000x the data?

Jeffrey G. Andrews

Director, Wireless Networking and Comm. Group
Department of Electrical and Computer Engineering
The University of Texas at Austin

Seminar, UT Austin

Oct. 14, 2011

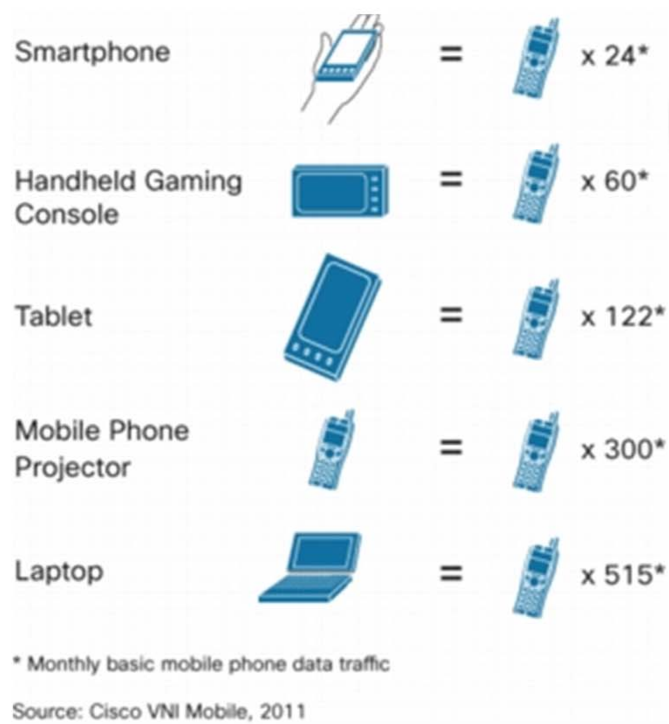
A Brief Sampling of Recent Cellular History

- 2000
  - Height of the Internet Bubble: Global Internet traffic doubling annually
  - Increasingly miniature cell phones were the rage
  - Mobile traffic metrics: “subscribers” and “minutes”
  - SMS popular internationally but not (yet) in US

- 2005
  - Not that much had fundamentally changed
  - “Device of the year” was Motorola Razr (stylish with nice form-factor, but not much new, really)
  - Mobile sector still dreaming of killer data apps
  - 3G rollout commencing slowly and cautiously
  - Mobile WiMAX standardized at very end of the year (802.16e), began to spur interest in 4G (LTE), but mostly from a “defensive” standpoint

Fast forward to 2010, BAM!

- The dream now a nightmare come true?
- Wireless networks in major cities suddenly at point of failure during peak hours
- Global mobile traffic increasing at well over 100% a year, no sign of relenting, in fact may still be accelerating



The Cisco Feb 1, 2011 Report
http://ow.ly/3S58j

- Global mobile data traffic grew 2.6-fold in 2010, nearly tripling for the third year in a row
- 2010's mobile data traffic was three times the size of the entire global Internet in 2000
- Mobile-connected tablets will generate as much traffic in 2015 as the entire global mobile network in 2010
- There will be 788 million mobile-only Internet users by 2015. (Up 56-fold from 14 million today).
- The mobile network will break the electricity barrier in more than 4 major regions by 2015.
  - By 2015, 40 countries (inc. India, Indonesia, Nigeria) will have more people with mobile network access than with access to electricity at home.

What does this mean to wireless companies and engineers?

Summary of Basic Conflict

- Over 100%/year growth in data traffic set to continue indefinitely
  - At least a 1000x increase over a decade (from say 2007 to 2017)
- Revenues will increase, but much more slowly
  - Business models for mobile data and especially video remain fuzzy
- Revenues and traffic have suddenly decoupled, compared to the cellular voice model that worked wonders for nearly two decades

Consequences to Industry

1. Must meter or restrict data usage, charging in proportion to the bits consumed (which will be very unpopular) OR
2. Decrease $/bit cost **exponentially** (how?) OR
3. Lose money and/or watch network collapse (which is the default)

Good News First: Cooper's "Law"

- Engineers have been exponentially increasing the achievable wireless rate for a very long time
- Cooper's Law [Martin Cooper, paraphrased]:

"The data rate available to a wireless device doubles roughly every 30 months"

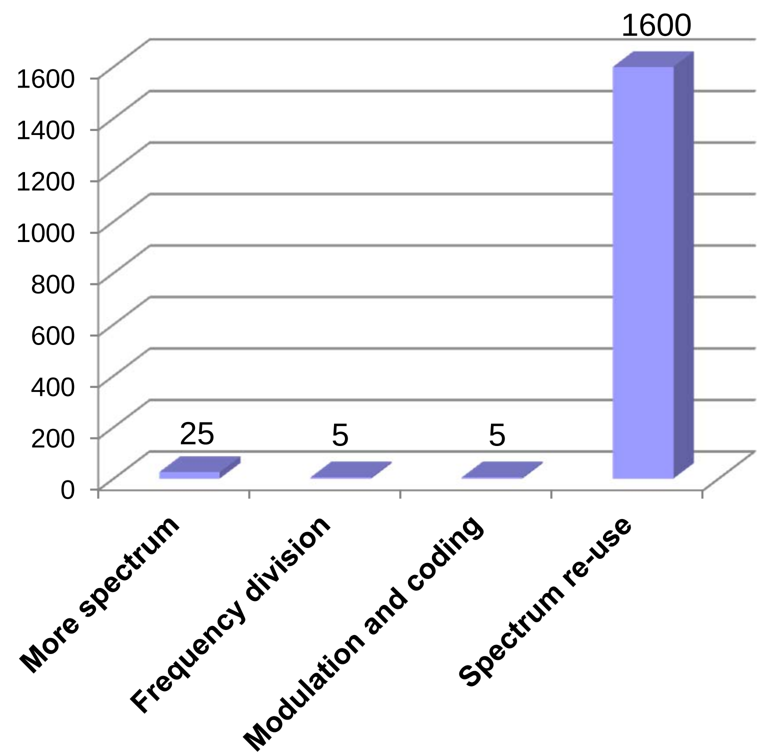
- This has held, more or less, for over 100 years (since Marconi)
- Unfortunately, the time constant (30 months) is far too slow to cope with the present demand surge

Incremental approaches are less interesting than ever before

Cooper's Law in the Modern Era

- 1,000,000x increase since ~1955
- Mostly driven by smaller cells
  - More base stations has been the key
- Good per link communication engineering (e.g. new standards) has had surprisingly little impact
  - 4G is not the answer, despite the relentless hype
  - Neither is 5G or 6G
  - Why?

Capacity multiplier since 1955 categorized:



The Shannon Limit

- The Shannon Limit is the maximum bit rate that a given communication channel can support
- Under certain conditions, in bits per second, it is:

$$R \leq nW \log_2(1 + \frac{S}{I+N})$$

# of independent symbol dimensions. Practical limit is small.

Bandwidth (Hz). Expensive.

SINR: Signal power divided by interference + noise power. Has to overcome a log() term.

- This is further divided among the number of users per cell
  - <u>This is a key but subtle point</u>: We assume the offered rate by base station n must be divided by the number of users K_n to get the per user rate.
  - Decreasing the number of users K_n is therefore equivalent to increasing rate: this is the main basis behind the 1600x small cell gain since 1950
  - Multiuser information theory (esp. with MIMO) argues that even better scaling is possible, but this may not have practical value.

The Limits of a “Good” Protocol

- Summary of <u>best case</u> for next decade:
  - ***n***: Factor of 4-8 from MIMO over next decade, including SDMA
  - ***W***: Factor of 2-3 increase in available “traditional” bandwidth
  - log(1+SINR): Factor of 1.5-2 rate gain from interference management and cancellation, inc. base station cooperation
    - Qualcomm, Vodafone, NTT, Motorola all have recently announced an aggregate gain of just ~10% via base station cooperation (CoMP)

$$R \leq nW \log_2(1 + \frac{S}{I+N})$$

- My estimates project a <u>very optimistic</u> gain of 12-48x in cellular capacity over next decade from improved radio technology
  - 4G, 5G, 15G… Not going to solve this problem with better standards.
  - For a lively treatment along these lines see:
    - M. Dohler et al, “Is the PHY Layer Dead?”, *IEEE Comm. Mag*. April 2011.

What to do? Make Cells Yet Smaller.

Networks must grow where data is demanded, not just use brute force. They will have:

1. Tower-mounted traditional base stations.

   Expensive (over $100K, plus high OpEx), 40W EIRP, medium to long-range (1-10 km), fast dedicated backhaul, mainly for guaranteeing universal basic coverage.

2. Picocells

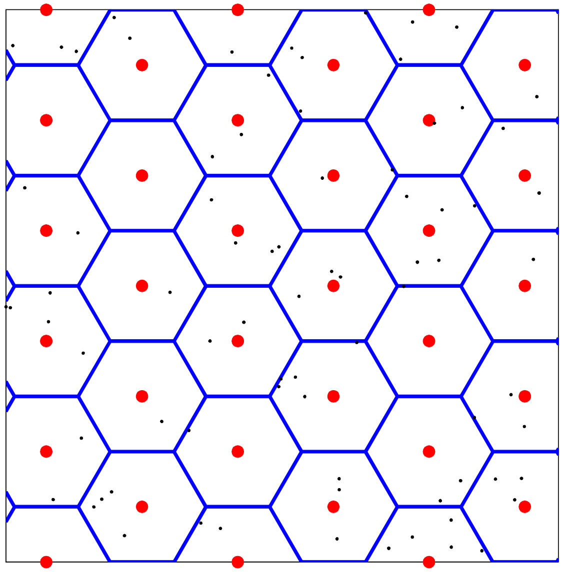
   Small, short-range (~100m), 2W EIRP, low-cost ($15-40K, small OpEx), deployed, maintained and backhauled (perhaps wirelessly) by service provider; typically targeting traffic “hotspots”

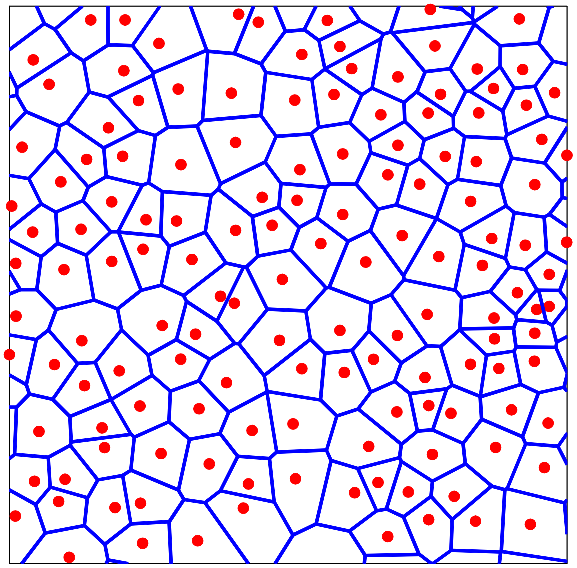
3. Femtocells

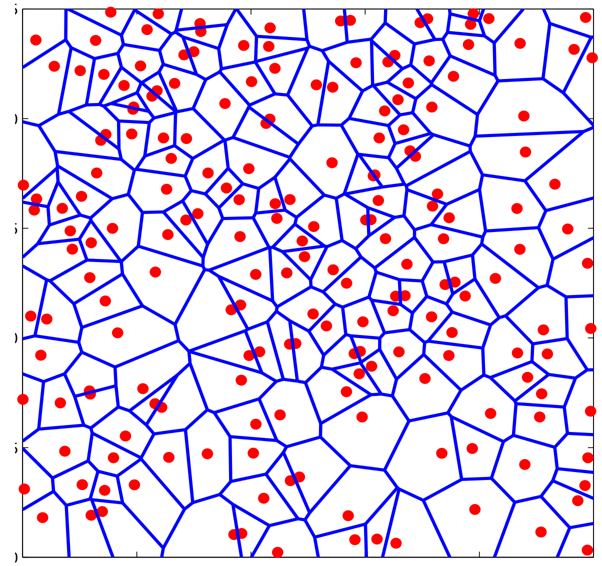
   Wi-Fi-esque range, power (200mW), cost ($100), and backhaul (IP, e.g. DSL). Licensed spectrum, cellular protocols, must inter-operate with cellular network with minimal coordination.

Is heterogeneity the way to go?

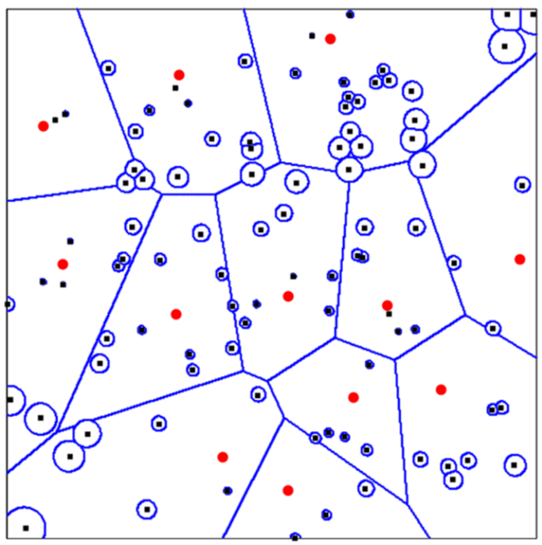
Modeling a Heterogeneous Cellular Network (HCN)

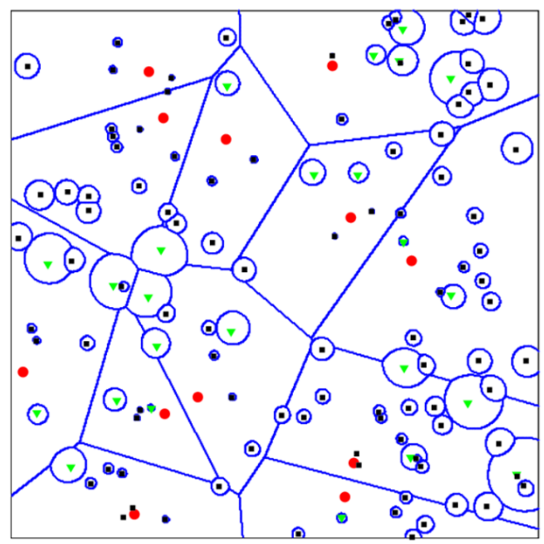

Traditional grid model


Actual 4G macrocells today


Completely random BSs

Zoom w/ femtocells

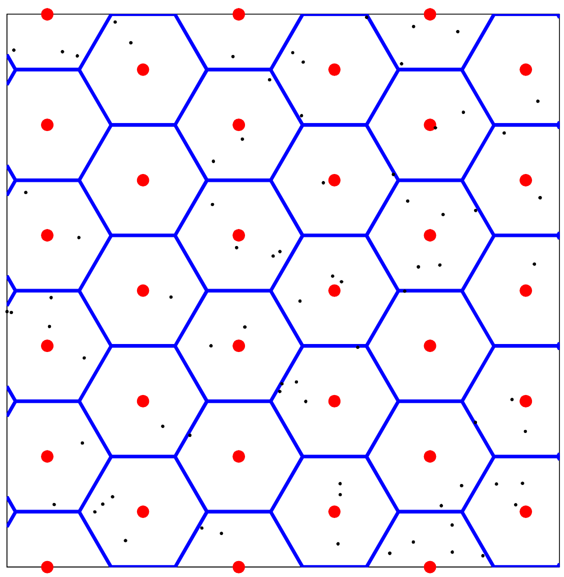




Zoom w/ picocells too

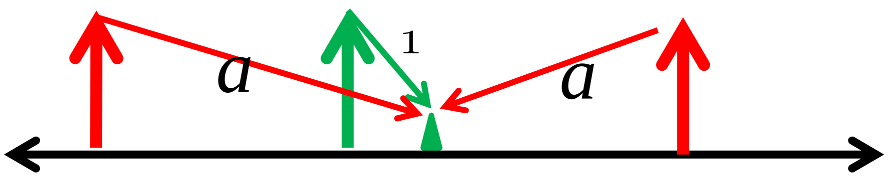
State of the Art in Cellular Models

- Industry and Simulation
  - Still predominantly use hexagonal base station model, esp. for "top tier" (macro BSs)



  - How to scale to an HCN?

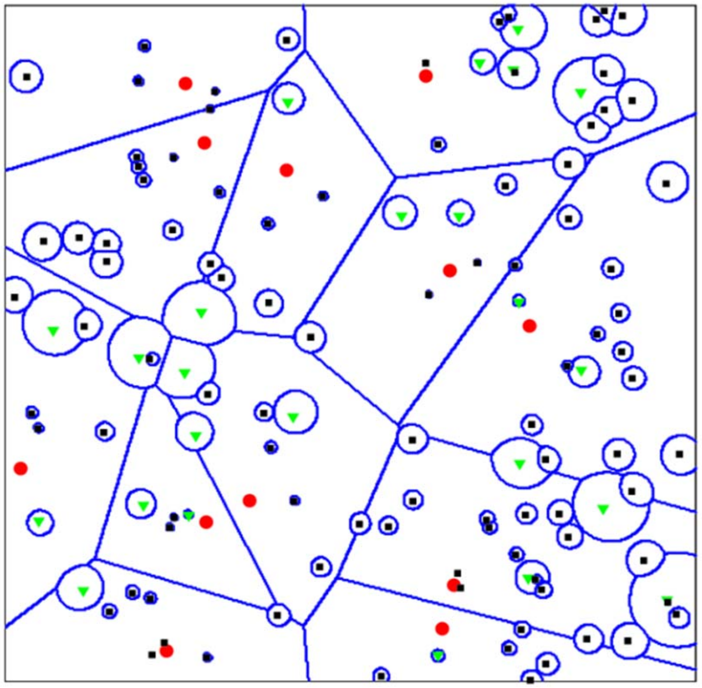
- Academia and Analysis
  - "Wyner model" formalized in early 1990's, highly idealized



  - State of the art in academia to this day, for analysis
    - Fixing SIR to be constant allows analysis, but seems pretty unrealistic
  - Considered by industry to be beneath contempt, even without heterogeneity
    - Actually not too bad a model for CDMA uplink, but about worthless otherwise

The Need for Random Spatial Models

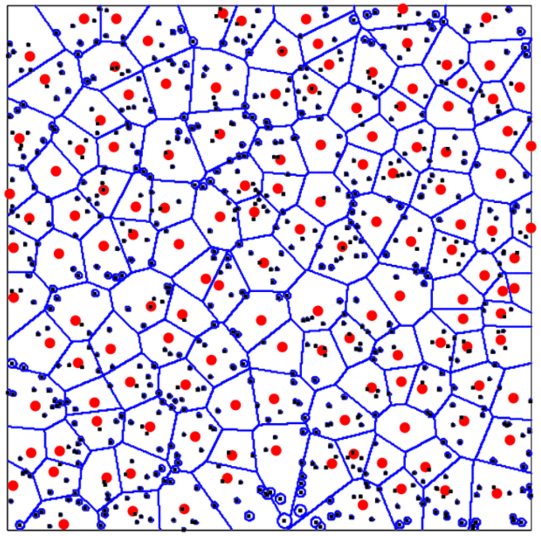
- To determine the rate, coverage, reliability of HCNs, we need to be able to model them
  - A fixed BS model is fairly absurd
  - Need statistical models
- Analogous to fading channels
  - Wireless channels are routinely modeled statistically
    - Rayleigh for small-scale variations
    - Lognormal for medium-scale
  - Idealized, but they capture the essential, promote understanding and innovation
  - More accurate and empirical random channel models are used as needed (e.g. in standards bodies)



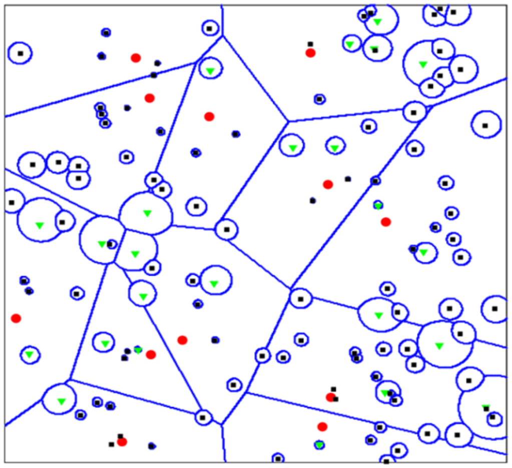
What is the outage/coverage probability or average rate in a network that looks like this?

Proposed Random Spatial Model

- K-tier network, each tier has BS locations taken from independent Poisson Point Processes (PPP)
  - Base Station Density: λ_j BS/m^2
  - Transmit Power: P_j Watts
  - SINR Target β_j
  - Path Loss Exponent: α_j
- Common Reactions to Model
  - Tier 1 BS's (macrocells) are not really "random", they are carefully planned!
  - Picocells typically clustered, not iid either, but maybe this is OK as 1st cut
  - Seems "about right" for femtocells, which are truly scattered

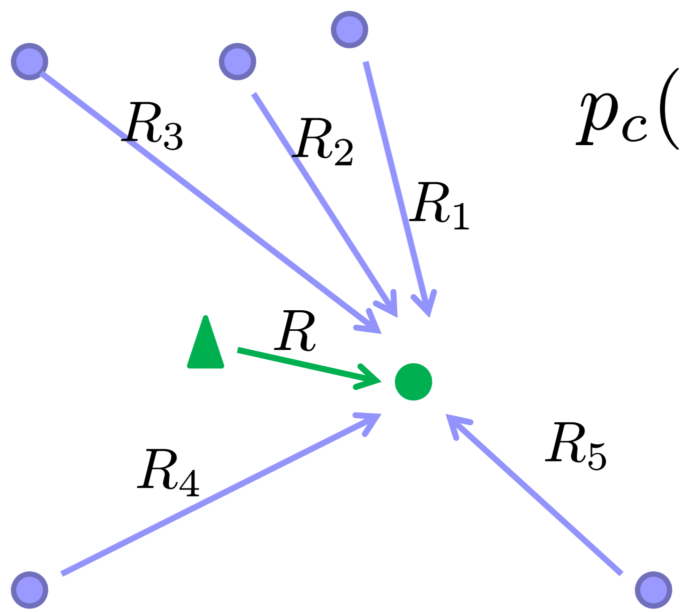


Actual tier 1 BSs with PPP femtos



3-tier zoom in showing max SINR coverage areas

Simplest Example: One-tier Downlink



$$p_c(T,\lambda,\alpha) = \mathbb{P}(\mathrm{SINR} > T)$$
$$= \mathbb{P}\left(\frac{hR^{-\alpha}}{\sigma^2 + I} > T\right)$$

Random channel effects (mean 1/μ accounts for transmit power)

Standard power law path loss

WLOG, aggregate interference can be quantified for MS at the origin as

$$I = \sum_{i \in \Phi_b \setminus b_o} g_i R_i^{-\alpha}$$

SINR can be found in very compact form

Theorem 1 [Andrews, Baccelli, Ganti'10]: When the fading power between any two nodes is exponentially distributed with mean μ^{-1}, the coverage probability is

$$p_c(T, \lambda, \alpha) = \pi\lambda \int_0^\infty e^{-\pi\lambda v(1+\rho(T,\alpha))-\mu T\sigma^2 v^{\alpha/2}} \mathrm{d}v,$$

where

$$\rho(T, \alpha) = T^{2/\alpha} \int_{T^{-2/\alpha}}^\infty \frac{1}{1+u^{\alpha/2}} \mathrm{d}u.$$

T = SINR threshold; λ = BS density; α = PL exponent; σ^2 = noise variance

Simplified coverage probability

1. Coverage probability in quasi-closed form for $\alpha = 4$

$$p_c(T,\lambda,4) = \frac{\pi^{\frac{3}{2}}\lambda}{\sqrt{T/\mathrm{SNR}}} \exp\left(\frac{(\lambda\pi\beta(T,4))^2}{4T/\mathrm{SNR}}\right) Q\left(\frac{\lambda\pi\beta(T,4)}{\sqrt{2T/\mathrm{SNR}}}\right).$$

where $\mathrm{SNR} = (\mu\sigma^2)^{-1}.$

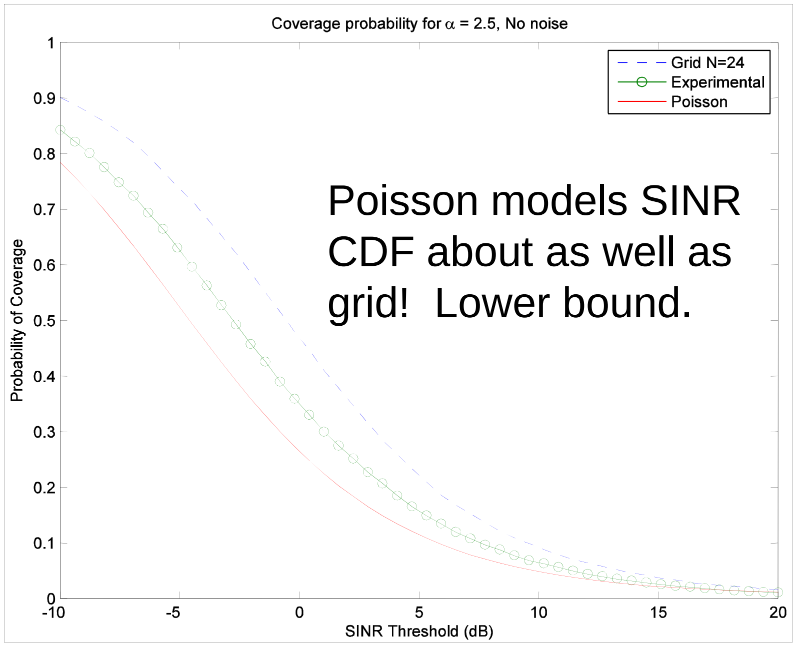
$$\beta(T,4) = 1 + \sqrt{T}(\pi/2 - \arctan(1/\sqrt{T})).$$

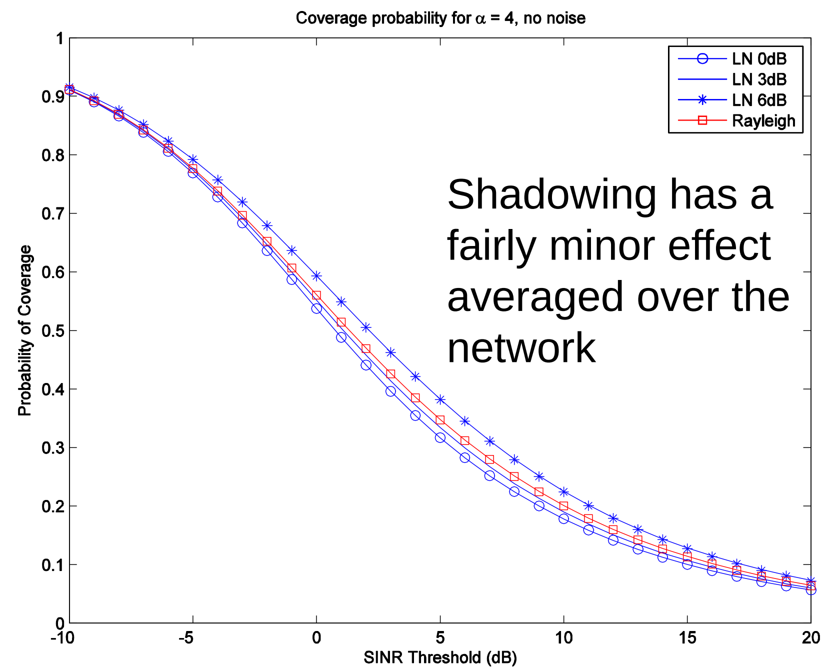
2. Coverage probability with no noise (any α):

$$p_c(T,\lambda,\alpha) = \frac{1}{\beta(T,\alpha)} = \frac{1}{1+\rho(T,\alpha)}.$$

Incredibly simple expressions. Most find surprising the weak dependence on SNR (power) and # of BSs.

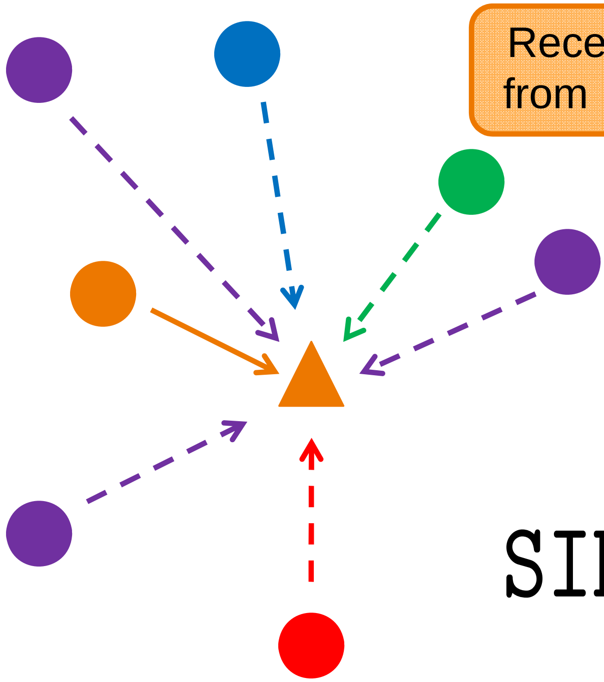
All nice, but is this model any good?





- The model is quite accurate and robust, even for the traditional planned cellular network
- Just as good as intractable hexagons (upper bound)

Extending to the K-Tier Downlink



Received Power from BS of i^{th} tier

$$P_r = P_i h_{x_i} \|x_i\|^{-\alpha}$$

Tx Power

Channel Gain

Standard Power Law Path Loss

$$\mathtt{SIR}(x_i) = \frac{P_i h_{x_i} \|x_i\|^{-\alpha}}{\sum_{j=1}^{K} \sum_{x \in \Phi_j \setminus x_i} P_j h_x \|x\|^{-\alpha}}$$

WLOG, aggregate interference can be quantified for mobile user at the origin as

$$I = \sum_{j=1}^{K} \sum_{x \in \Phi_j \setminus x_i} P_j h_x \|x\|^{-\alpha}$$

Main Result: Coverage Probability

Theorem 2 [Dhillon, Ganti, Baccelli, A'11]**:** The coverage probability for a typical mobile user connecting to the strongest BS, neglecting noise and assuming Rayleigh fading is

$$p_c(\{\lambda_i\}, \{T_i\}, \{P_i\}) = \frac{\pi}{C(\alpha)} \frac{\sum_{i=1}^{K} \lambda_i P_i^{2/\alpha} T_i^{-2/\alpha}}{\sum_{i=1}^{K} \lambda_i P_i^{2/\alpha}}, \quad T_i > 1,$$

Key Assumption

where $$C(\alpha) = \frac{2\pi^2 \csc(\frac{2\pi}{\alpha})}{\alpha}$$

Proof proceeds similarly to one-tier after applying "Key Assumption", allows us to change a union over all tiers to a sum of the per-tier coverage probabilities.

A Couple Even Simpler Special Cases

For a single-tier cellular network (K=1):

$$p_c(\lambda, T, P) = \frac{\pi}{C(\alpha)T^{2/\alpha}}$$

- Only depends on SIR target and path loss, very similar to previous result for any SIR (here recall SIR > 1)

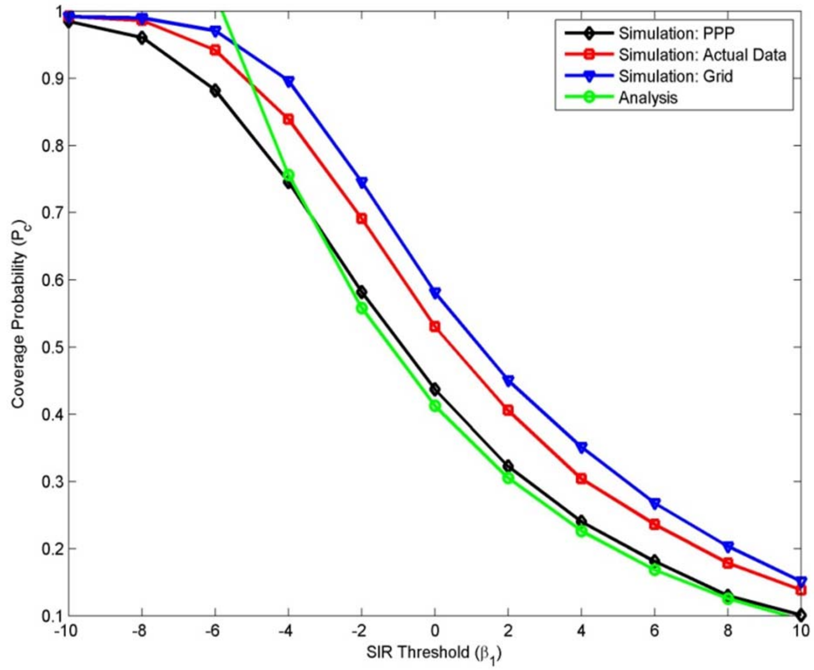
For a K-tier network with same SIR thresholds for all tiers

$$p_c(\{\lambda_i\}, T, \{P_i\}) = \frac{\pi}{C(\alpha)T^{2/\alpha}}$$

- Interestingly, same as single-tier.
- **Neither adding tiers nor base stations changes the coverage/outage in the network!**
- Therefore, the network sum rate increases linearly in principle with the number of BSs, formally as $\sum_{i=1}^{K} \lambda_i$

How Accurate is the K-Tier Model?

- Nearly as accurate as a grid
  - Grid provides upper bound (1-2 dB)
  - PPP provides lower bound (1-2 dB)
- Our results hold down to an SIR target of about -4dB
  - In practice, rarely have a target below -3 to -5 dB, so $T > 0$ dB simplification may be OK
- Recent standards contributions from Samsung and Motorola (for picocells) and TI (femtos) agree with our analytical results



$\alpha = 3$, $P_1 = 1000\ P_2$
$\lambda_2 = 2\ \lambda_1$, $\beta_2 = 1$ dB

Open Questions for HCNs

- Dynamic traffic models undermine spatial averages
  - We assumed all base stations transmit all the time
  - Untrue for small cells, changes interference/load significantly
- Biasing, cell association and load balancing
  - Network would like to push users off heavily-loaded macrocells and onto small cells, even if they take a large SINR hit
  - Macrocell users and off-loaded user both get larger resource %
- Uplink SINR Model appears much harder
  - Now must model both MS and BS locations at once
  - Same analysis approach does not work
  - This has important implications for cell association rules
- Interference management, scheduling, MIMO, power control, mobility management… Pretty much everything we "know" about cellular networks must be re-thought!

Can we get to 1000x by 2017?

- Assume in the next 7 years that communication technology (MIMO, scheduling, interference management) and increased bandwidth provides 10x
  - We need 100x from HCNs
- This means we need 100x more effective base stations
  - That is, these new small BSs must carry an equivalent load to a current base station to count as a full effective BS
  - Biasing, intelligent off-loading, and load-balancing will be huge
- This is a lot of infrastructure and backhaul, but within reach at projected cost of picos ($20K) and femtos (~$200)
  - Sizeable chunk also likely to be provided by improved Wi-Fi offloading

Summary

- Femtocells, picocells and other small-cell architectures are going to be ubiquitous very soon
  - There is no other plausible way to meet the 1000x challenge
  - We have shown that such infrastructure can in principle be added to the network without limit, at arbitrary power, and at arbitrary locations, without hurting coverage
- Random spatial models are essential mathematical tools for modeling the new cellular paradigm
  - Paradoxically, the organic capacity-centric networks of the future may be easier to analyze than the structured coverage-centric ones of the past
  - Powerful analysis and optimization tools are available and under further development, including for more complex point processes
    - See M. Haenggi et al, JSAC 2009 for a tutorial introduction.

Sparse Bibliography

[AndBacGan10] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A Tractable Approach to Coverage and Rate in Cellular Networks", submitted to *IEEE Trans. on Communications*, Sept. 2010. http://arxiv.org/abs/1009.0516

[ChaAndGat08] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey", *IEEE Communications Magazine*, Sept. 2008.

[DhiGan11] H. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and Analysis of K-Tier Downlink Heterogeneous Cellular Networks", to appear, *IEEE Journal on Sel. Areas in Comm.*, April 2012. http://arxiv.org/abs/1103.2177

[Hae09] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic Geometry and Random Graphs for the Analysis and Design of Wireless Networks", *IEEE Journal on Sel. Areas in Comm*, Sept. 2009.

Special Thanks

- Texas Instruments, Motorola, and National Science Foundation for Direct Funding
- DARPA for Indirect Funding (development of the stochastic geometry tools and models for ad hoc networks, with M. Haenggi, S. Weber, and N. Jindal)
- Collaborators on this Work
  - Vikram Chandrasekhar (former student, now at TI)
  - Radha K. Ganti and Han Shin Jo (current postdocs)
  - Ping Xia and Harpreet Dhillon (current PhD students)
  - Prof. Francois Baccelli (ENS, Paris)