Living at the Edge: A large deviations approach to the MIMO Outage Capacity

Aris Moustakas
P. Kazakopoulos, P. Mertikopoulos (Univ. Athens)
&
G. Caire (USC)

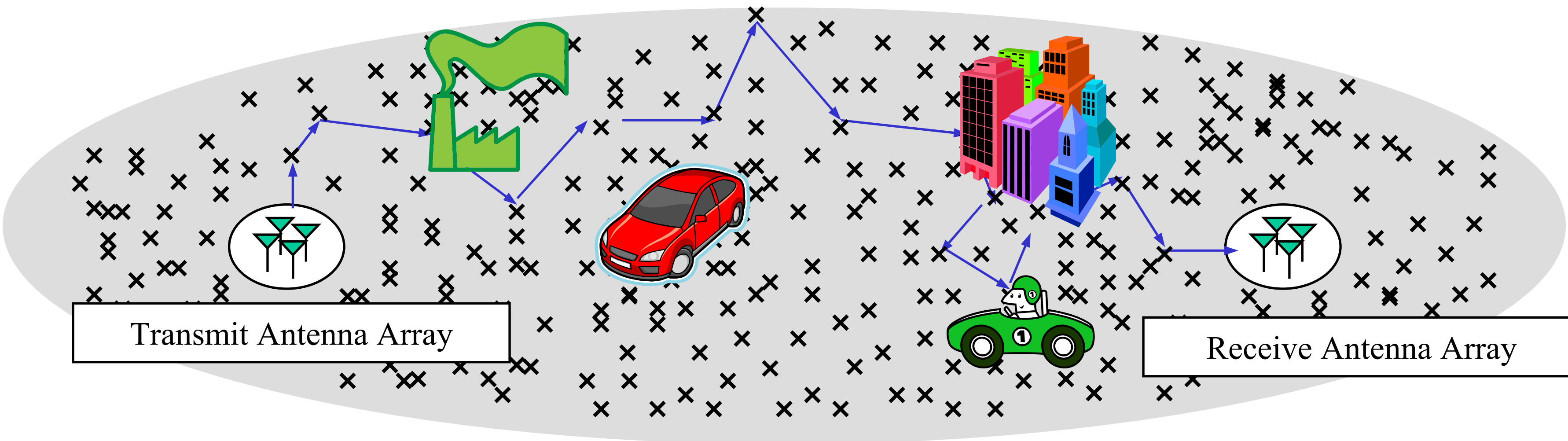
Introduction

- 4G Wireless Communications: High rates, very mobile, network-centric

- Multi-antenna arrays are standard on both terminals and base-stations
  - How much information can you send?
  - Depends on many things…
    - Randomness
    - Mobility
    - Coding
  - Many regimes have analyzed analytically, numerically
    - Tails of distribution?
    - Important for real applications and intuition

- Use of Coulomb gas method to analyze tails
  - (Majumdar *et al)*

Wireless Setting



$$r_\alpha = \sum_i G_{\alpha i}\, t_i + \eta_a \qquad \mathbf{r} = \mathbf{G}\,\mathbf{t} + \boldsymbol{\eta}$$

- Noise η: uncorrelated ~CN(0,1)
- Scattering creates spatiotemporal fluctuations
- **G:** random N x M channel matrix: i.i.d. ~ CN(0,1/N)
  - Assume β=M/N>=1

Mutual Information

$$I_N = \log\det\left[\mathbf{I} + \rho\mathbf{G}\mathbf{G}^+\right] = \sum_{k=1}^{N}\log\left(1 + \rho\lambda_k\right)$$

- I_N Mutual Information also random
- Strategy for fast fading channel **G** (e.g. racing car):
  - Transmit message with ergodic rate

$$R_{erg} = E\left[\log\det\left(\mathbf{I} + \rho\mathbf{G}\mathbf{G}^+\right)\right]_G$$

  - "heroic coding" (message lasts long enough to ride over all G-waves)
- Strategy for slowly fading channel **G** :
  - Transmit with rate R and hope for the best (sometimes you loose)
  - **Define outage criterion for given rate R**

$$p_{out}(R) = \Pr(I_N < R) = \int_0^R dx E\left[\delta\left(x - I_N\right)\right]_G$$

- **Calculate Statistics:**
  - What is mean, variance etc?

Two parameters involved

- Large N :
  - Make large antenna arrays

- Large S :
  - Increase ρ

Asymptotics in N

- Calculate Mean (Ergodic) Mutual Information
  - All you need for fast fading.
  - For large N capacity per antenna I_N / N becomes a deterministic quantity
    - Randomness subleading in N
  - Underlying Idea:
    - Empirical distribution of eigenvalues hardens to deterministic function

$$p_{MP}(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi x} \qquad a,b = \left(\sqrt{\beta} \pm 1\right)^2$$

    - Ergodic capacity = average over MP distribution

$$R_{erg} = E\left[I_N\right] = N\int_a^b dx P_{MP}(x)\log(1+\rho x)$$

- Shannon Transform

$$r(\{\mathbf{X}\}) = E\left[\frac{1}{N} Tr\left(\log\{\mathbf{I} + \rho\mathbf{X}\}\right)\right]$$

  - For given "channel" distribution produces the (normalized) mutual information

Asymptotics in N

- However…
- I_N does not harden
  - Fluctuations of O(1) important, especially for finite N
- "Central Limit Theorem"
  - For N large (but finite) and R "close" to the ergodic MI
    - Actually for |R-Rerg| fixed and large N

$$P(R) = \frac{e^{-\frac{(R-R_{erg})^2}{2\sigma(\rho)^2}}}{\sqrt{2\pi}\sigma(\rho)}$$

  - Many ways to skin this cat (all essentially moment based)
  - Calculation of σ(ρ) has to do with fluctuations of eigenvalues "close" to the mean
  - Higher moments vanish O(1/N)

Asymptotics in ρ

- But…

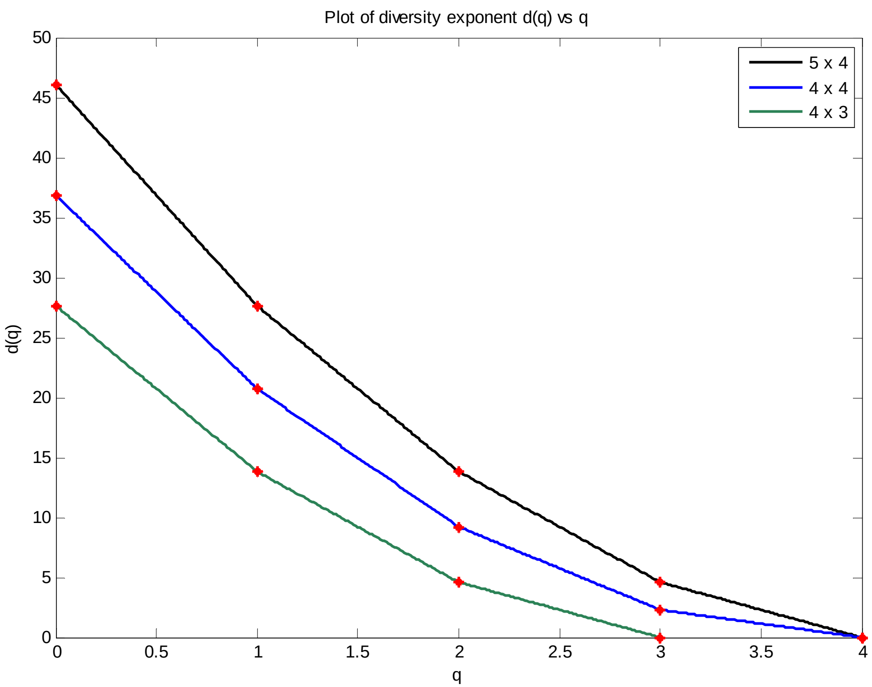
- In practice need tails of mutual information distribution
  - Low outage probabilities for better fidelity
  - Need to address tail distribution
  - One way: Large SNR
    - Large ρ focuses on the tails of the eigenvalue distribution (low outage)
  - Scale quantities properly: $R = q \log \rho$ with $q \leq N$
  - Surprising result: As $\rho \to \infty$
    - Outage Probability be comes

$$\lim_{\rho \to \infty} \frac{\log P_{out}(q \log \rho)}{\log \rho} = -d(q)$$

  - where d(q) piecewise linear function of rate variable q with endpoints

$$d_k = (M-k)(N-k)$$
$$k = 0, \ldots, N$$

Asymptotics in ρ



$$\lim_{\rho \to \infty} \frac{\log P_{out}(q \log \rho)}{\log \rho} = -d(q)$$

$$d_k = (M-k)(N-k)$$
$$k = 0, \ldots, N$$

- Example 5x4, 4x4, 4x3
  - Piecewise linear behavior of the diversity exponent
- **Pros**:
  - Very intuitive result
  - Diversity – Multiplexing Tradeoff
  - Prompted design of codes that have this exact structure (for large ρ)
  - Applicable to a huge number of applications
- **Cons:**
  - Scale of y-axis (exp(20) = 100dB !!)
  - Extremely small outages
  - Not clear how large SNR necessary
    - Original Zheng-Tse paper never mentioned any scales
  - No prediction for $O((\log \rho)^0)$ term
  - Case of $R > R_{erg}$? CCDF(R)?

Comparison

- How do Gaussian and large ρ approaches compare?
  - Look in regions where both are valid (large ρ and large N)
- Case M=N (β=1)
  - Gaussian approximation:

$$E[I_N] = R_{erg} \approx N \log \rho$$

$$Var[I_N] \approx \log \sqrt{\rho}$$

    - Thus

$$\log P_{out,Gaussian} \approx -\frac{(N \log \rho - R)^2}{\log \rho}$$

  - DMT (large ρ) approximation:

$$\log P_{out} \approx -\log \rho \left(N - \frac{R}{\log \rho} \right)^2 = -\frac{(N \log \rho - R)^2}{\log \rho}$$

  - So far so good…

Comparison

- How do Gaussian and large ρ approaches compare?
  - Look in regions where both are valid (large ρ and large N)
- Case M>N (β>1)
  - Gaussian approximation:

$$E[I_N] = R_{erg} \approx N \log \rho$$

$$Var[I_N] \approx \log\left(\frac{\beta}{\beta - 1}\right)$$

- Thus

$$\log P_{out,Gaussian} \approx -\frac{\left(N \log \rho - R\right)^2}{2 \log\left(\frac{\beta}{\beta - 1}\right)}$$

  - DMT (large ρ) approximation:

$$\log P_{out} \approx -\frac{\left(N \log \rho - R\right)\left(M \log \rho - R\right)}{\log \rho}$$

Comparison

- $\lim_{N\to\infty}\lim_{\rho\to\infty} \neq \lim_{\rho\to\infty}\lim_{N\to\infty}$
- For large SNR Gaussian approximation VERY wrong
- Why?
  - Gaussian approximation assumes freezing of eigenvalues of GG'
  - For large SNR the outage probability focuses on probability of eigenvalues close to zero (1/ρ)
- Can we get a unified behavior (working well for small and large SNR)?
- Good news:
  - Both asymptotics have identical scaling with N

$$\log P_{out} \approx -N^2 \log\rho\left(\beta - \frac{R}{N\log\rho}\right)\left(1-\frac{R}{N\log\rho}\right)$$

$$\log P_{out,Gaussian} \approx -\frac{\left(N\log\rho - R\right)^2}{2\left|\log\left(1-\beta^{-1}\right)\right|} = O\left(N^2\right)$$

    - Extend large N method to capture rare events

Coulomb – Gas Analogy

- Joint probability distribution of eigenvalues of GG'

$$P(\lambda_1, \lambda_2, \ldots, \lambda_N) \propto \exp\left[-N \sum_k \lambda_k + (M-N) \sum_k \log \lambda_k + 2 \sum_{k>m} \log|\lambda_k - \lambda_m| \right]$$

- Exponent is energy of point charges repelling logarithmically in the presence of external field
  - Most probable configuration of λ's corresponds to minimum energy

$$P \propto e^{-N^2 S[p]}$$

- Extend argument to large number of charges (Use of Dyson conjecture)
  - N plays role of temperature

$$S = \int dx \; p(x) V_{eff}(x) + \int \int dxdy \; p(y) p(x) \log|x-y|$$

$$V_{eff}(x) = x + (\beta - 1) \log x$$

  - (Existence of another term to set normalization contraint)
- Minimizing S w.r.t p(x) gives the Marcenko Pastur distribution

$$p_{MP}(x) = \frac{\sqrt{(x-a)(b-x)}}{2\pi x}$$

- To calculate P(r) need to impose constraint

$$P(r) = E\left[\delta\left(Nr - \sum_k \log(1+\rho\lambda_k)\right)\right]$$

  - Fourier transform (or use large deviations arguments – Lagrange multiplier)

$$S \to S - k\int dx\ \ p(x)\left(\log(1+\rho x) - r\right)$$

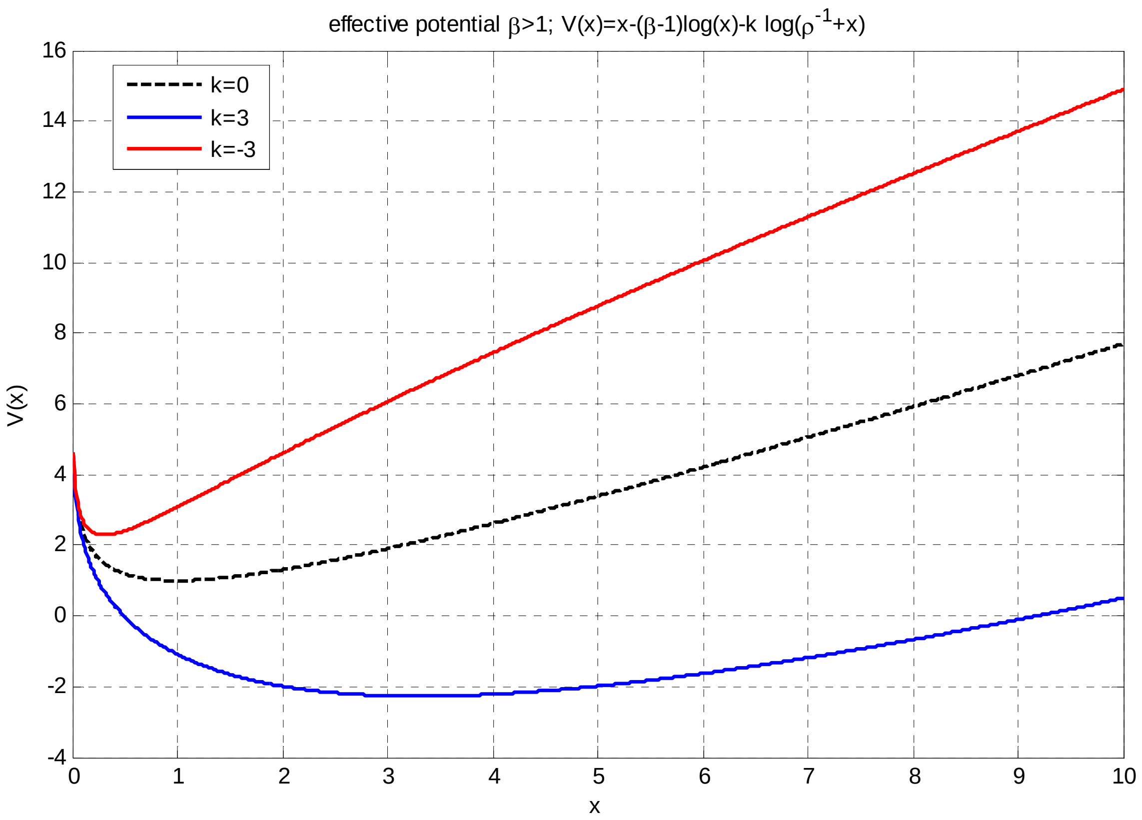
$$V_{eff}(x) \to V_{eff}(x) - k\log\left(1+\rho x\right)$$

- k plays role of strength of logarithmic attraction/repulsion at $x_0 = -\rho^{-1}$

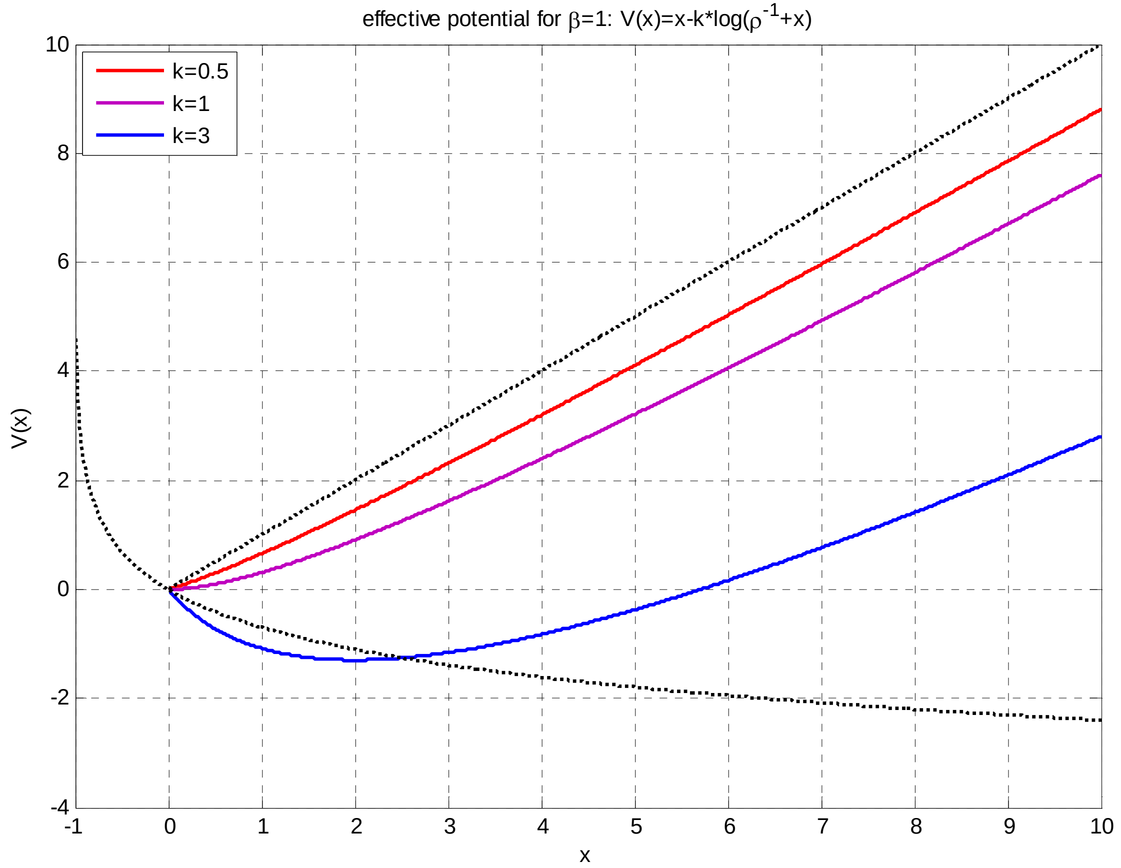
  - k>0 shifts charge density to larger values (R>Rerg), k<0 to smaller ones (R<Rerg)

- Minimizing S w.r.t p(x) gives the *generalized* Marcenko Pastur distribution

- Qualitative features of solution β>1 from V(x)
  - Minimum always at x>0, irrespective of sign of k



- Qualitative features of solution β=1
  - For k>kc>0 minimum shifts from x=0 to x>0



Generalized MP equation

- Use of Tricomi theorem to calculate p(x) and closed form expr. for energy S[p]
- Case β>1

$$p(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi x(1+\rho x)}\left(\rho x + \frac{\beta - 1}{\sqrt{ab}}\right)$$

  – a, b, k calculated from

$$p(b) = p(a) = 0$$
$$r = \int_a^b dx\, p(x) \log\left(1+\rho x\right)$$
$$1 = \int_a^b dx\, p(x)$$

- Case β = 1
  – $r > r_c(\rho) > r_{erg}(\rho)$

$$p(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi(\rho^{-1}+x)}$$

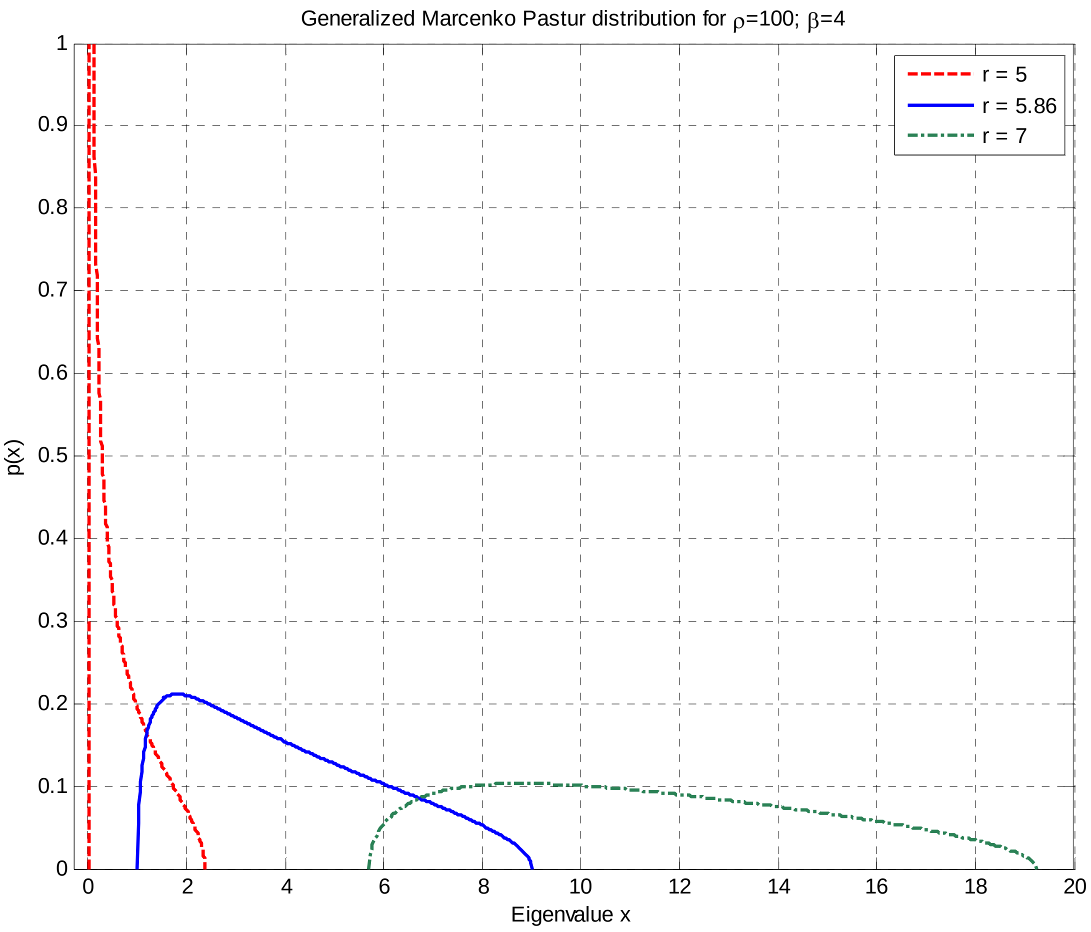
  – $r < r_c(\rho)$

$$p(x) = \frac{\sqrt{b-x}}{2\pi(1+\rho x)}\left(1+\rho x - \frac{k\rho}{\sqrt{1+\rho b}}\right)$$

  – 3rd order phase transition (Vivo, Majumdar *et al*)

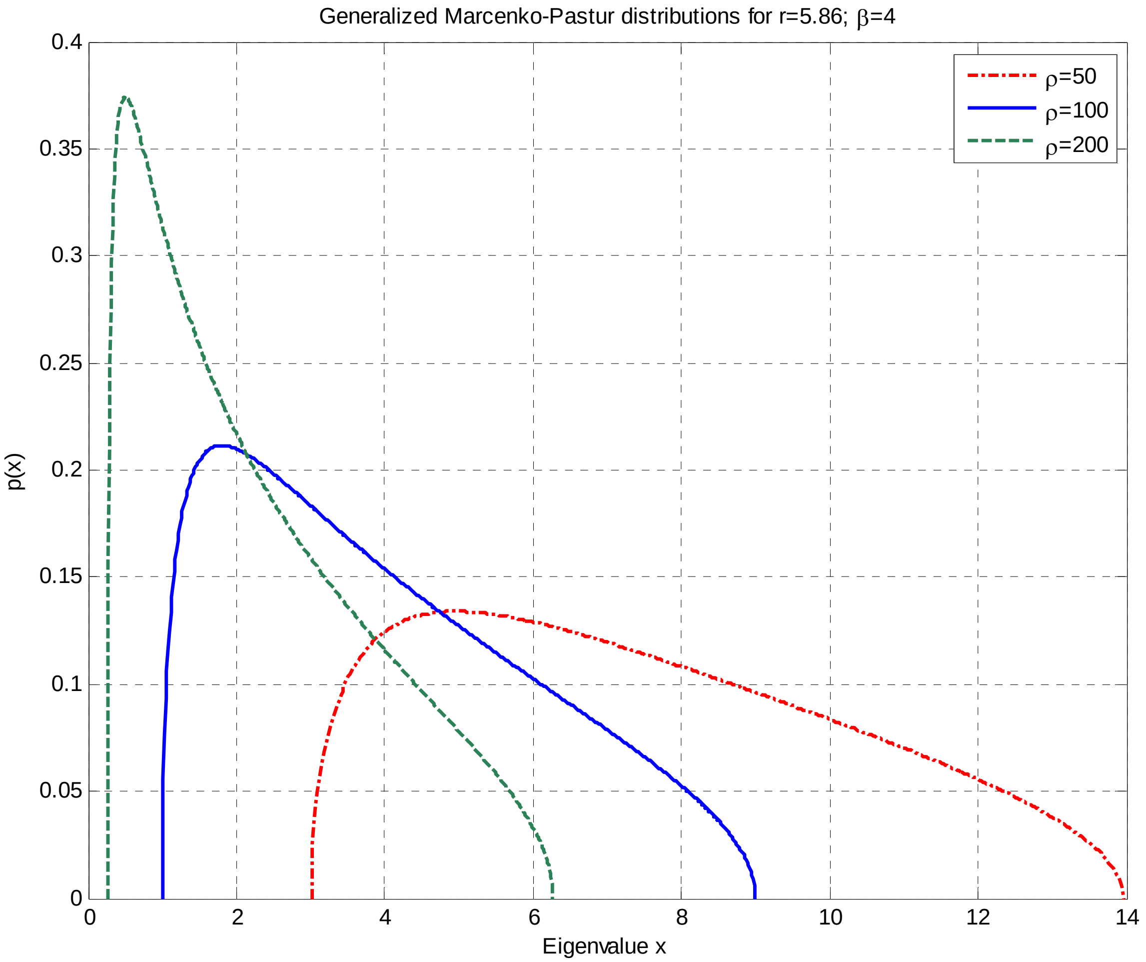
Generalized MP equation

- Examples



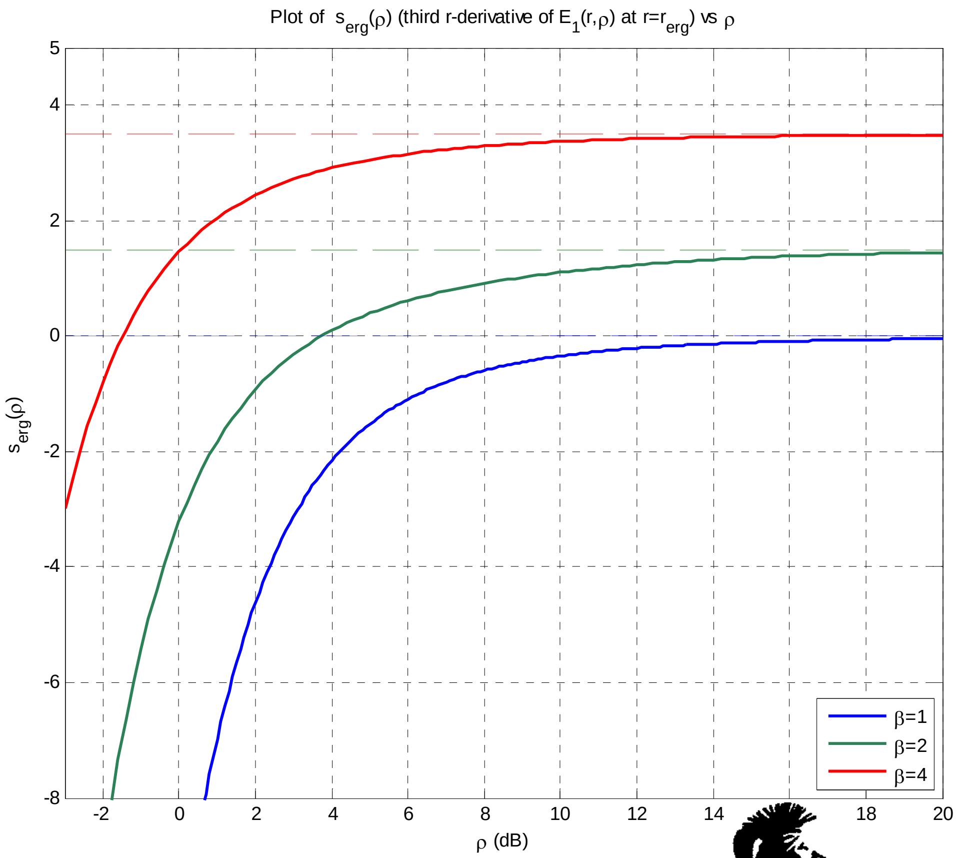
Generalized MP equation

- Examples



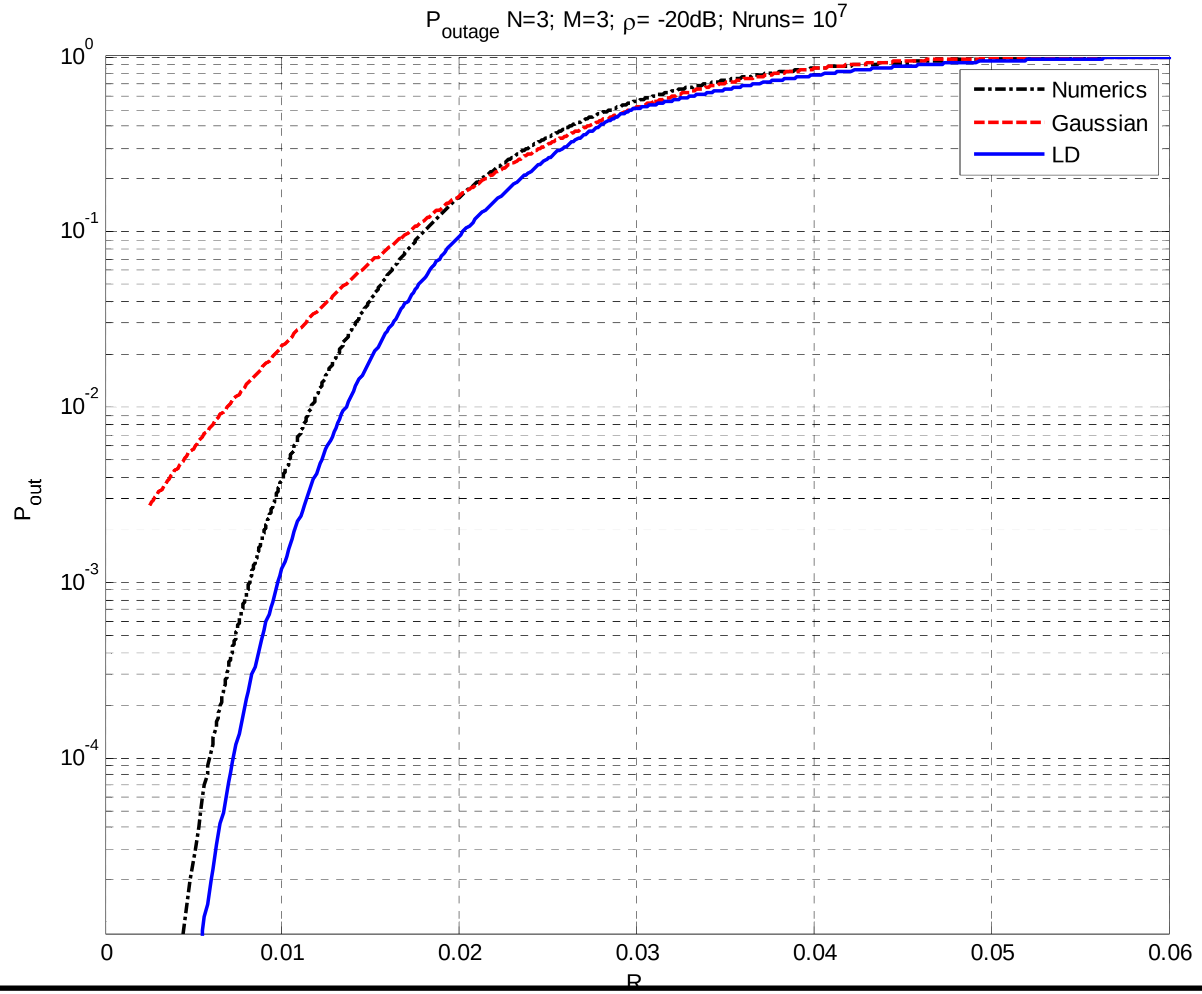
Special Cases

- Limit k=0 gives Gaussian approximation
  - Not surprising: finite k gives the MGF of mutual information
  - As usual CLT can be derived from small k-behavior of MGF
  - k **IS** the replica number
- Skewness??
  - β=1 – large ρ: "Gaussian"
  - β > 1 – large ρ: NOT Gaussian
  - Small ρ: NOT Gaussian



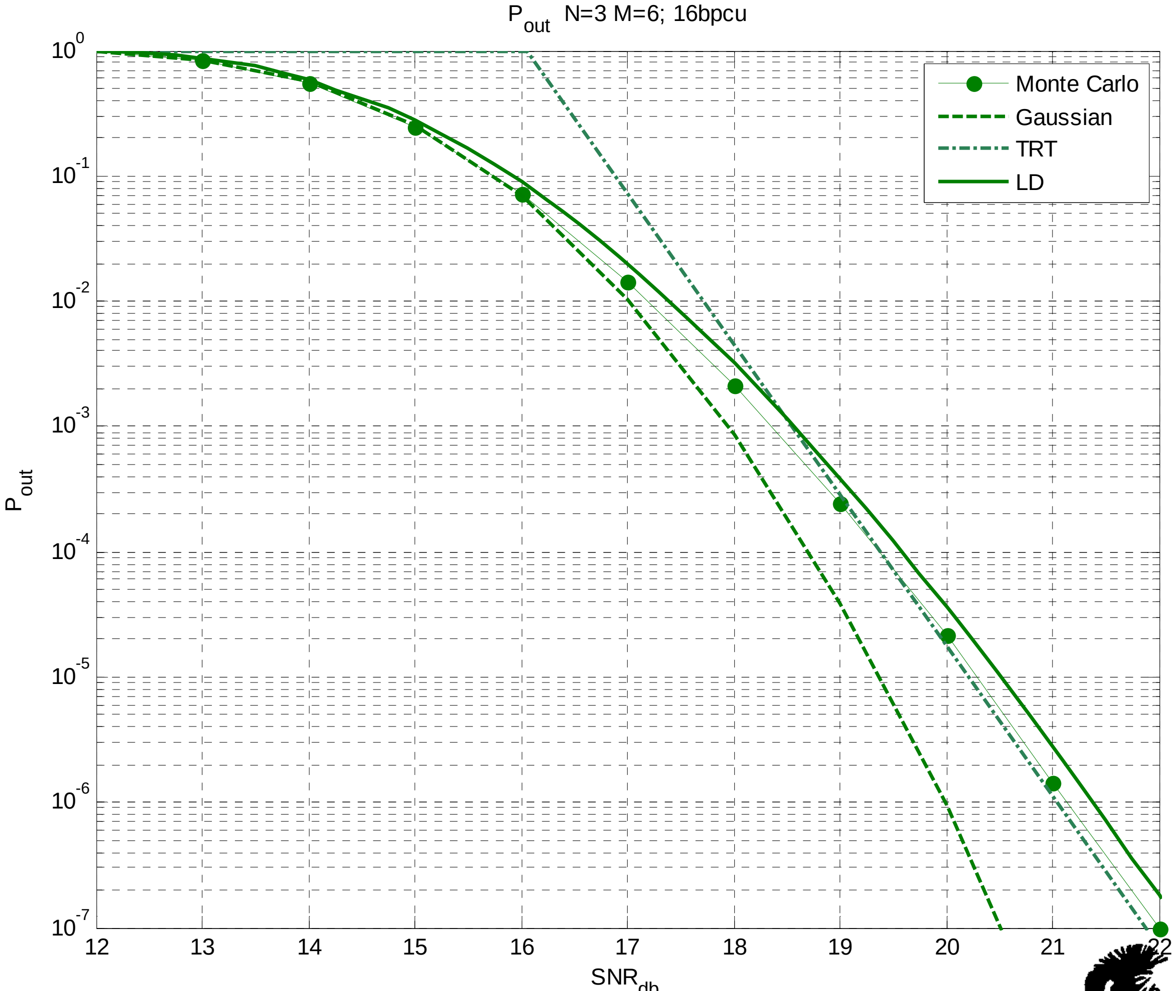
Special Cases

- Small ρ:
  - r_{erg} small so Gaussian approx cannot be true



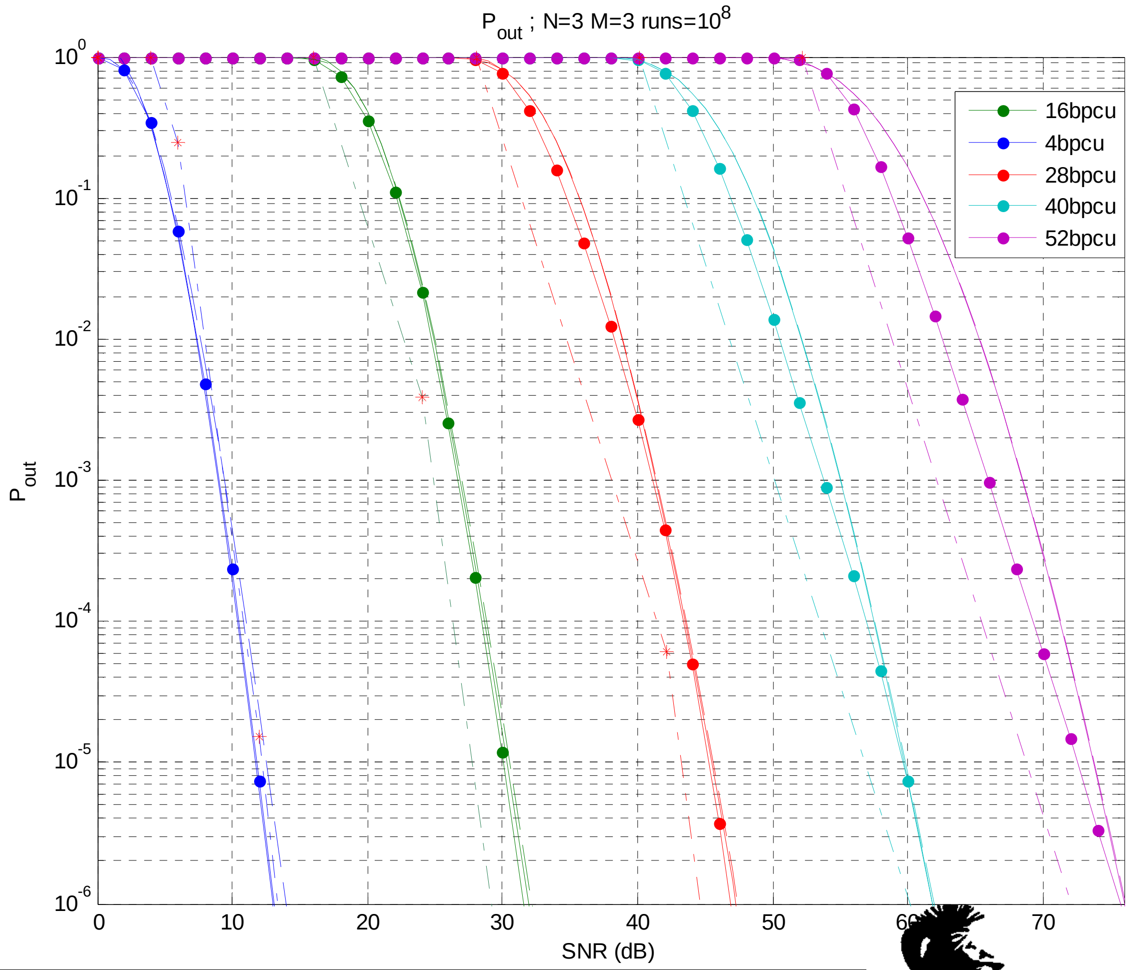
Special Cases

- Large ρ 3x6
  - Compare LD, DMT, Gaussian



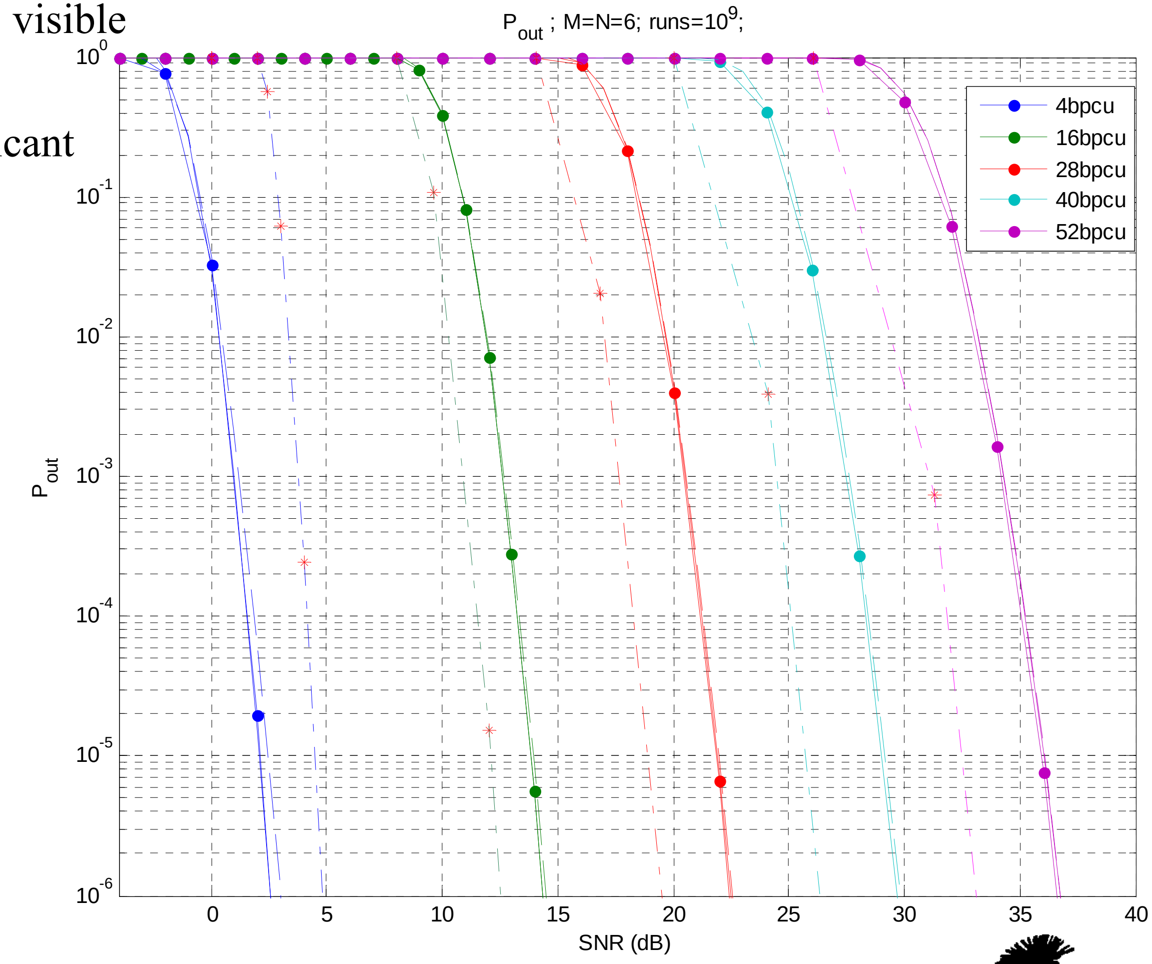
Special Cases

- Large ρ 3x3
  - Compare LD, DMT, Gaussian
  - DMT linear behavior visible
  - O(1) term in DMT significant
  - LD robust



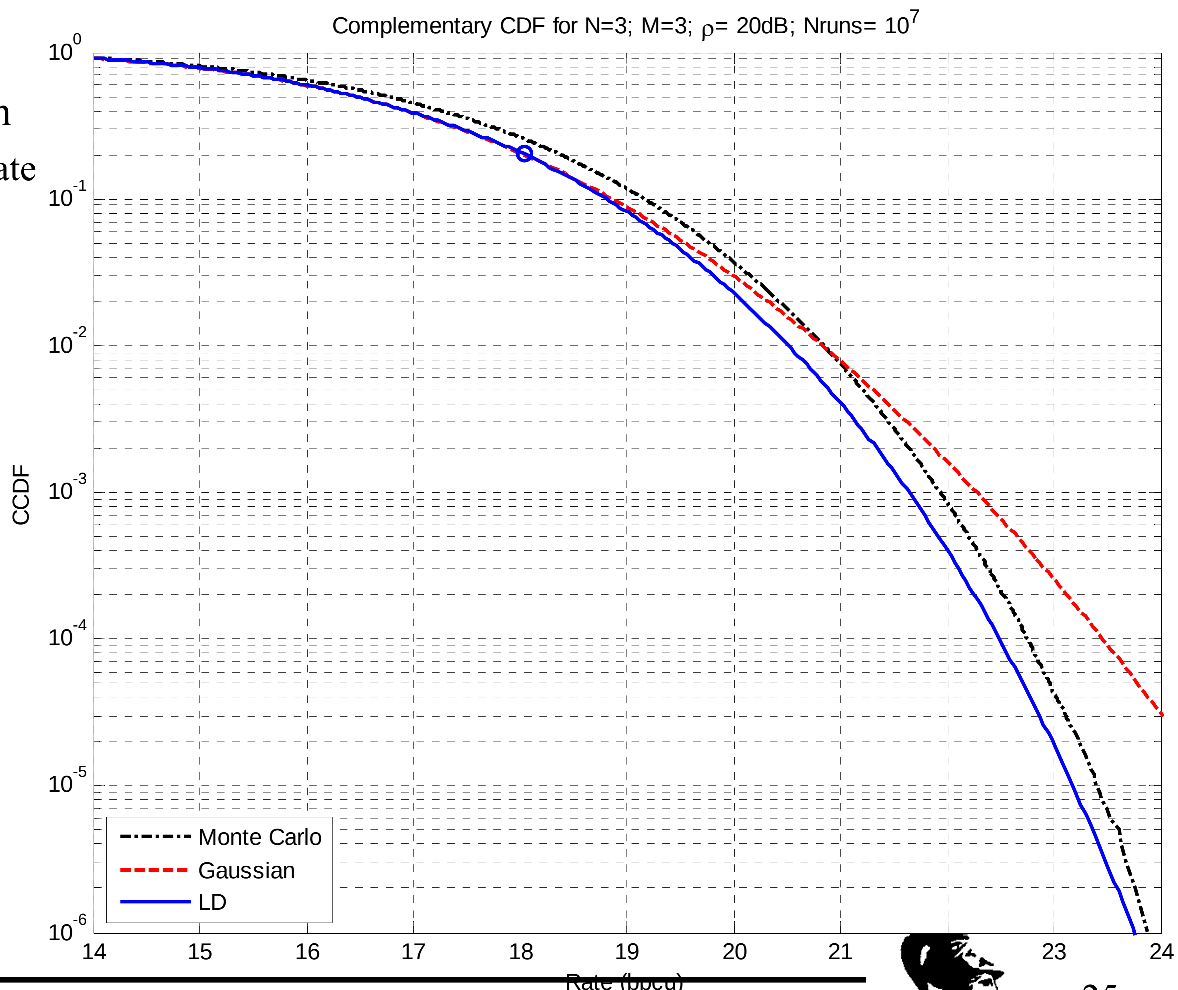
Special Cases

- Large ρ 6x6
  - Compare LD, DMT, Gaussian
  - DMT linear behavior less visible
  - O(1) term in DMT significant
  - LD robust



Special Cases

- $\beta = 1$ case
- Look at CCDF
- Relevant for scheduled transmission
  - Transmit to user with higher inst. rate
  - CCDF shows gains in trans rate
- Phase transition visible between
  - "Gaussian"
  - New phase $r > r_c(\rho) > r_{erg}(\rho)$
  - 3rd order discontinuity



Special Cases

- Large r and ρ
  - By expanding S[p] for large ρ, fixed

$$q \approx \frac{r}{\log \rho}$$

$$\log P_{out} \approx -N^2 \log \rho (\beta - q)(1 - q)$$

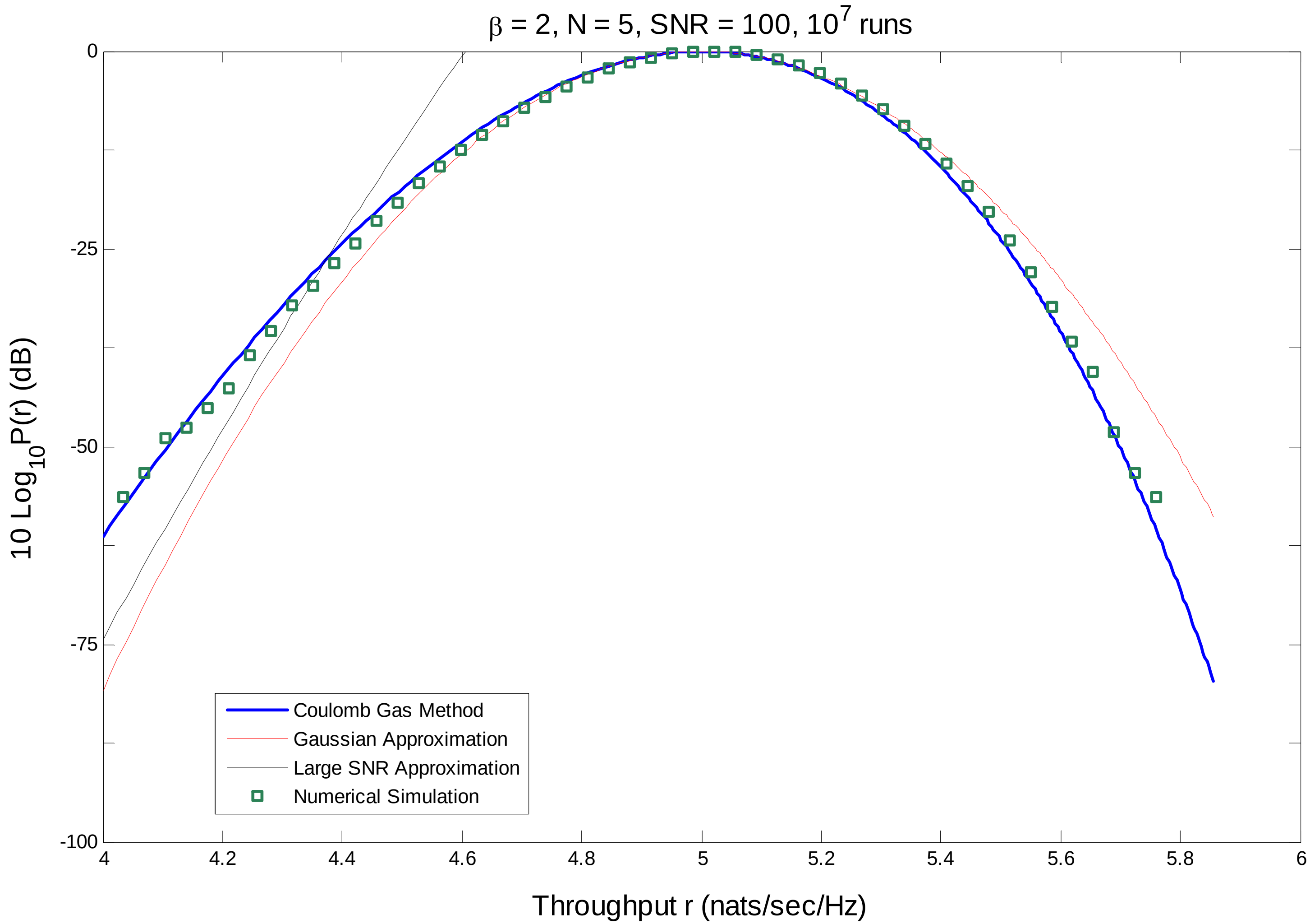
  - Correctly recover DMT outage behavior

- r=R/N<<1 gives $S = -\beta \log\left(\frac{eR}{M\rho}\right)$

$$P_{out} \approx \left(\frac{eR}{M\rho}\right)^{-MN}$$

  - Intuition: all elements of GG* have to be less than R/ρ, i.e. all MN elements of G small

Large N Validation



Summary

- Distribution of MIMO mutual information:
  - Large N, arbitrary SNR
  - Smooth transition between known approximations
  - Understanding of all interesting limiting cases

- Generalized Marcenko – Pastur distributions:
  - Provides distributions of eigenvalues at the tails of the MI distribution
  - Fully characterized by information-theoretic quantities/parameters of system (R, N, β)

- Open Questions:
  - Correlated Channels?
  - Applications to other IT problems
    - Waterfilling solution statistics
    - MMSE SINR distribution