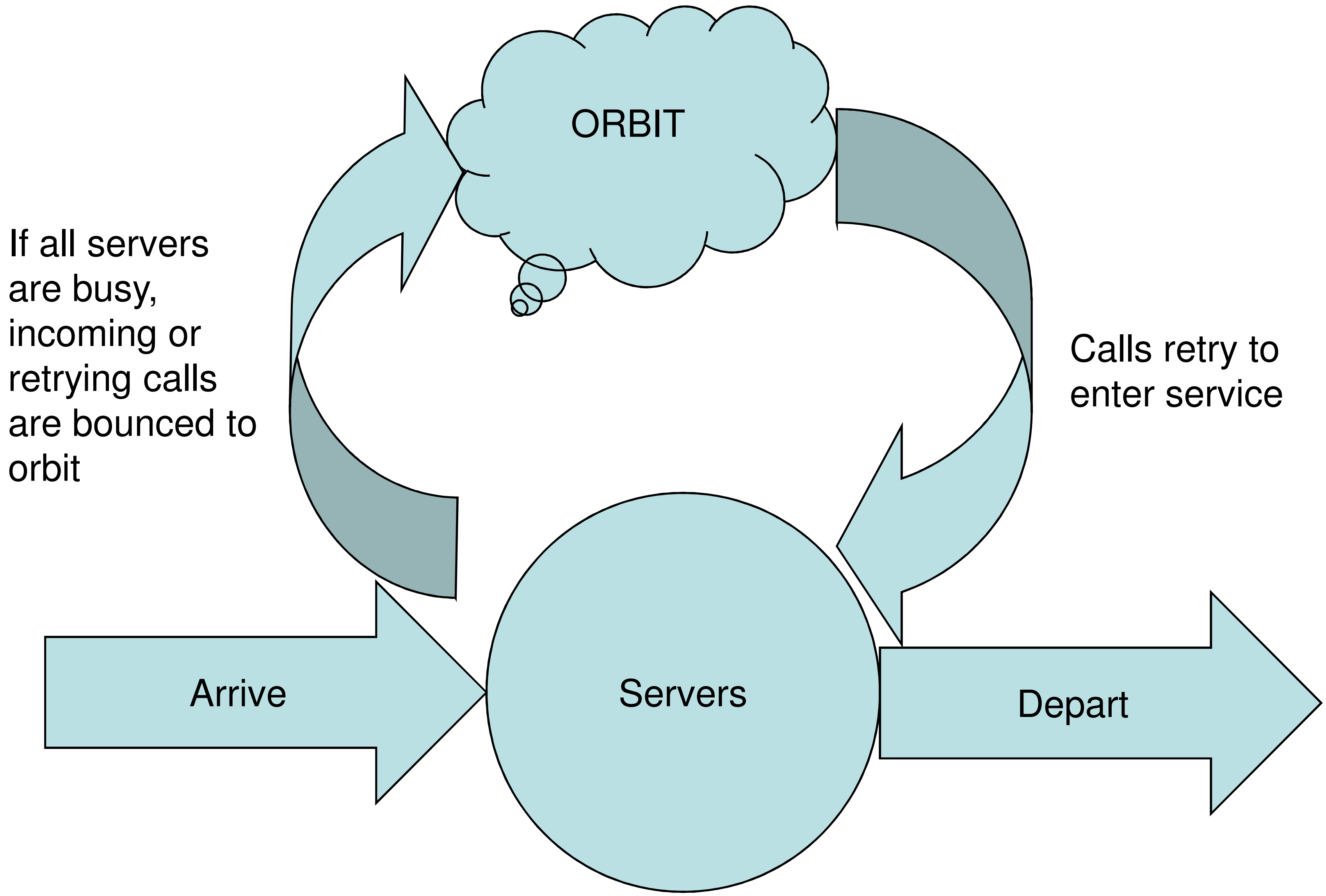
# Low-Variance Retrial Times in a Multiserver Queueing Model

Prof. Andrew Ross, David Lubke,
Andrew Livingston, Katherine Ballentine

Eastern Michigan University,
Ypsilanti, Michigan

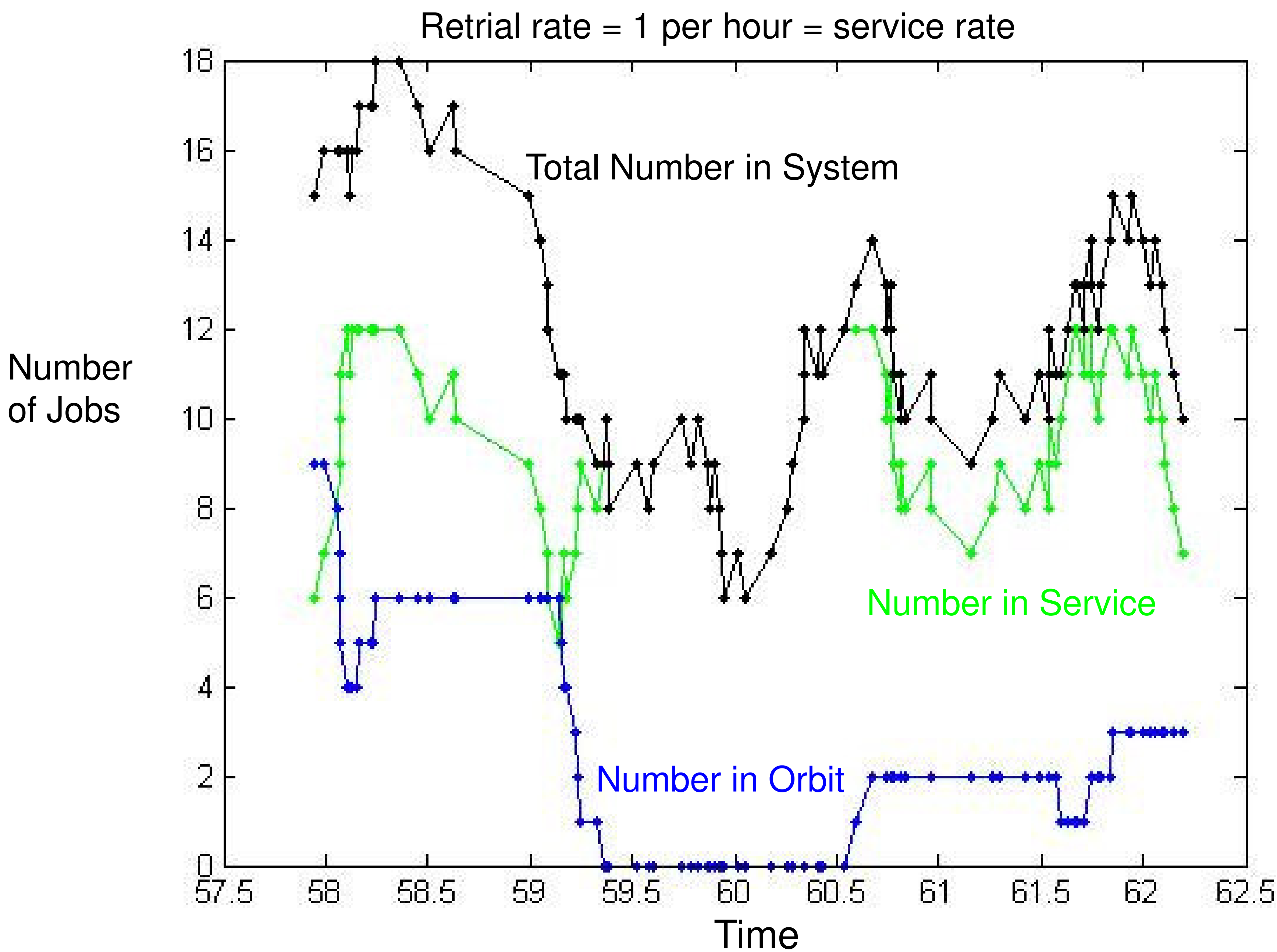
CanQueue 2009, Univ. of Windsor


ORBIT
If all servers are busy, incoming or retrying calls are bounced to orbit
Calls retry to enter service
Arrive
Servers
Depart

# Multiserver Retrial Queues

- Mobile phone: link to tower
- Satellite phone/data: link to satellite
- Dial-up internet
- Credit card verification

All have non-exponential retrials


Retrial rate = 1 per hour = service rate
Number of Jobs
Total Number in System
Number in Service
Number in Orbit
Time
0
2
4
6
8
10
12
14
16
18
57.5
58
58.5
59
59.5
60
60.5
61
61.5
62
62.5

# Single-Server Systems: Distribution matters!

- Ethernet and WiFi deliberately avoid using deterministic retrial distributions
- They are single-server systems, though
- Multi-server systems generally act differently for measures like probability of delay.

# Our Main Question

- When must you take the retrial distribution into account?

- Methods:
  - Discrete-event simulation
  - Markov-chain computation

# M/M/c/0 + G-retrials

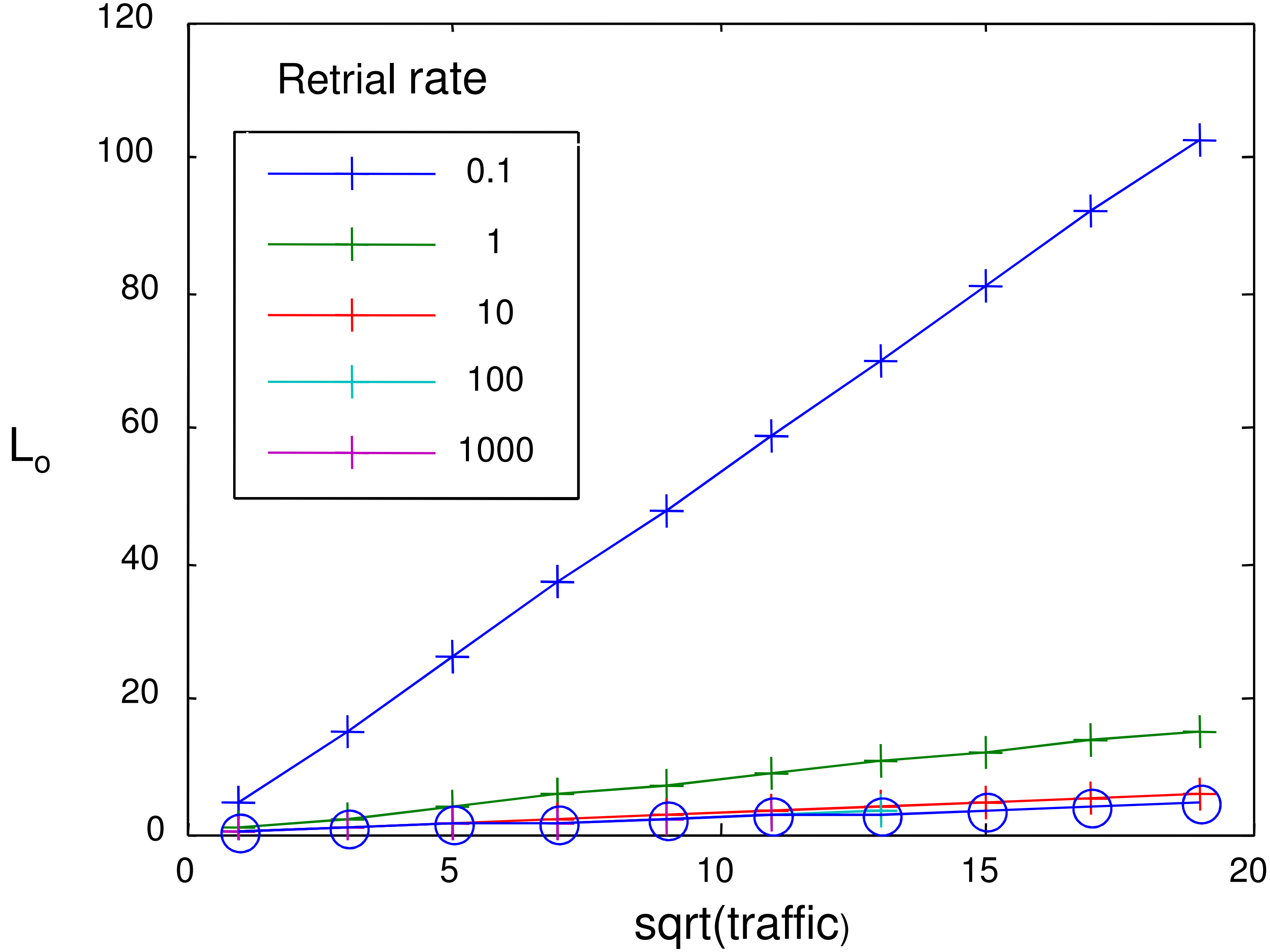
- Poisson arrivals with constant rate
- Exponential service
- No organized buffer
- Everyone in orbit retries
  - not just one person
- Customers never give up

# Square-Root Staffing

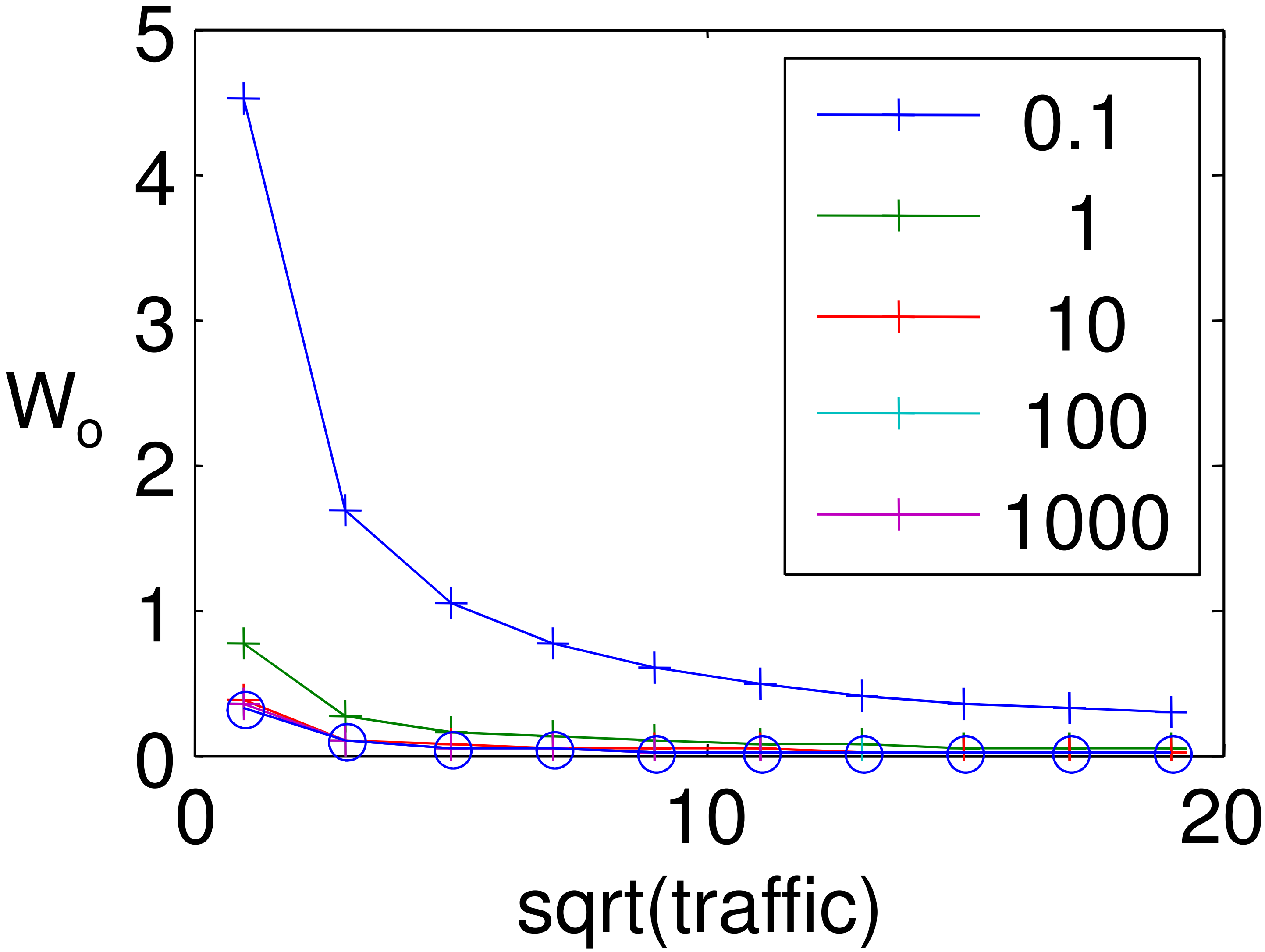
- Servers = traffic + 1*sqrt(traffic)

- QED: Quality and Efficiency Domain

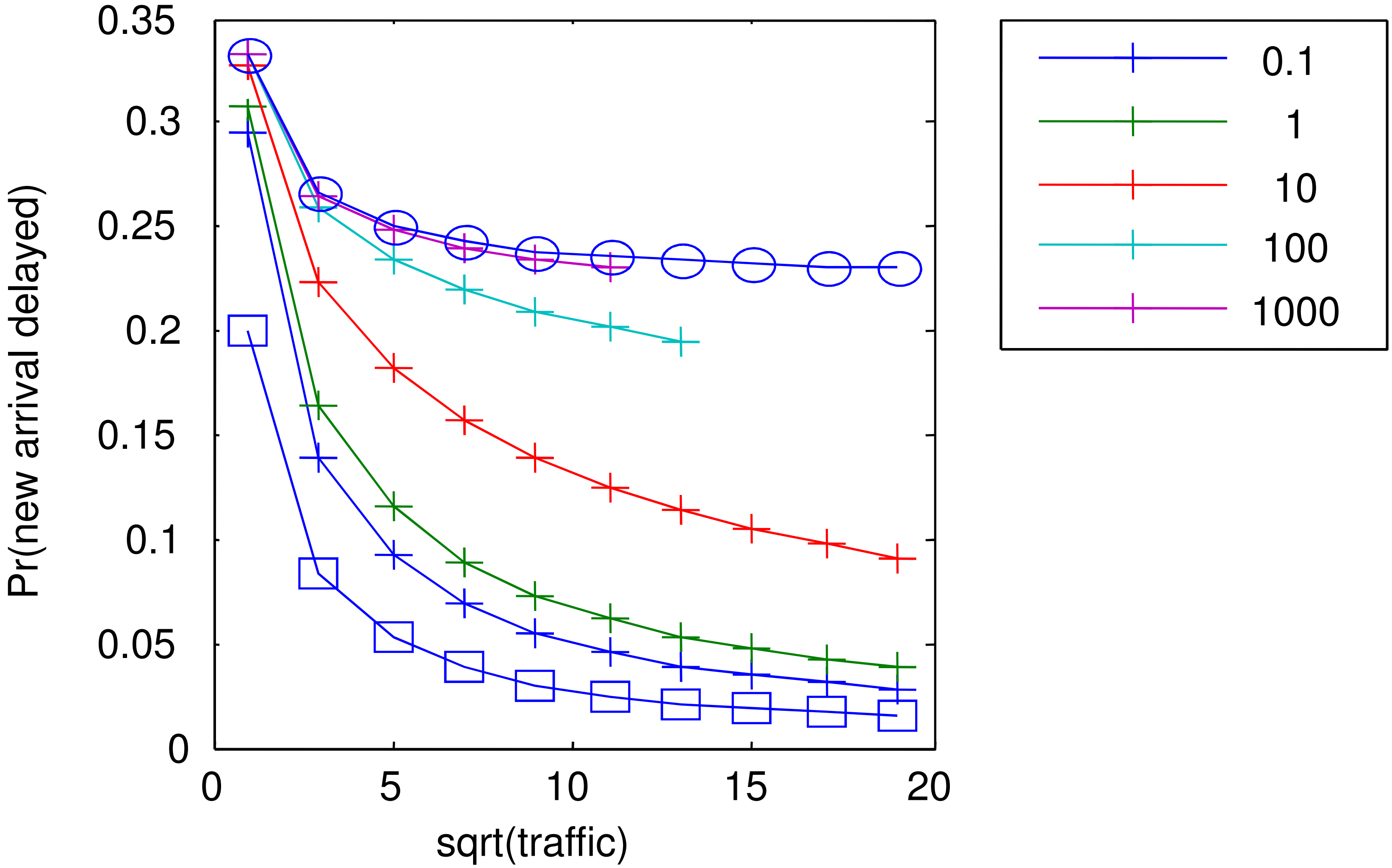
# $L_o$ as system grows



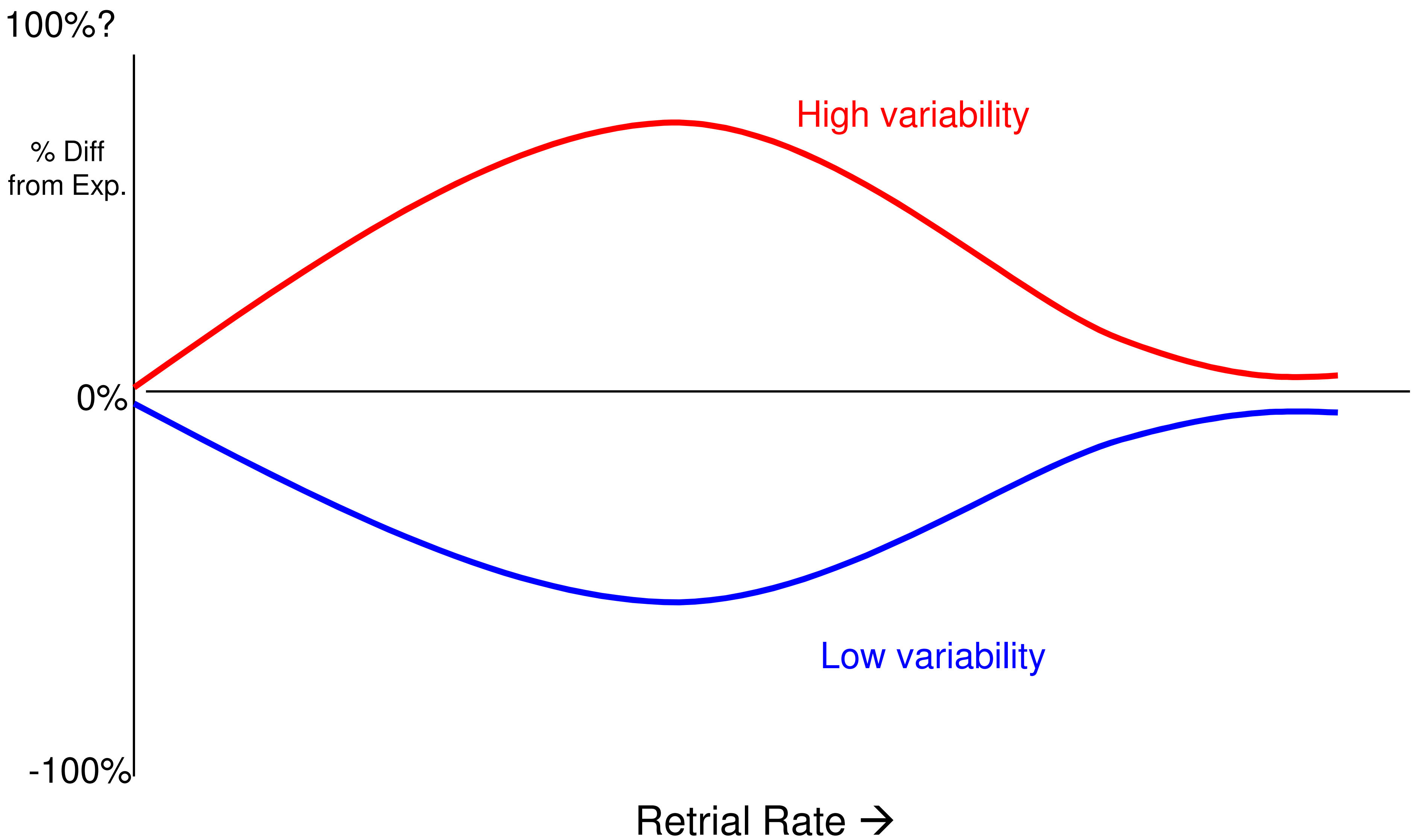
# $W_o$ as system grows



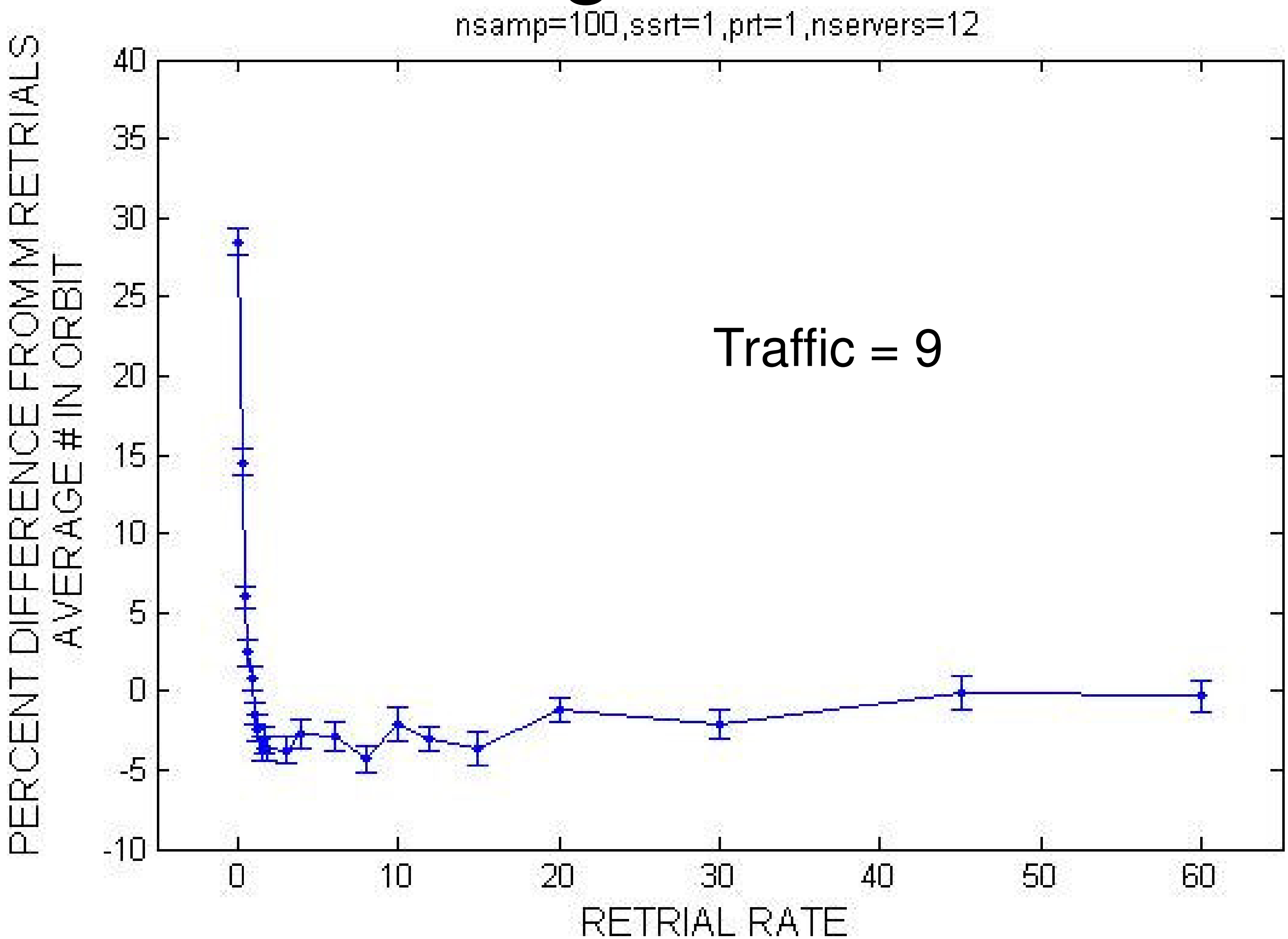
# Pr(new arrival delayed)



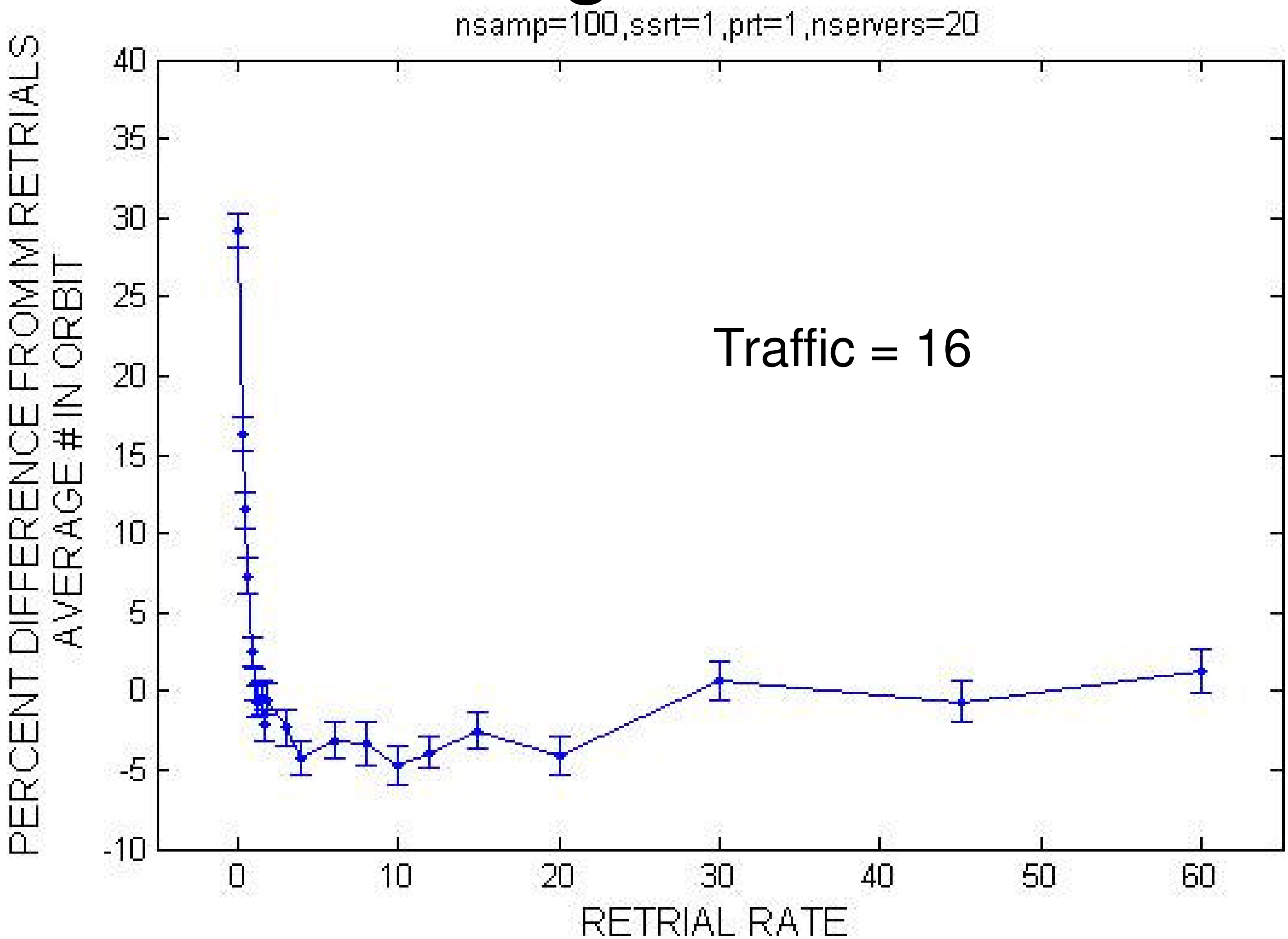
# What We Expected to See



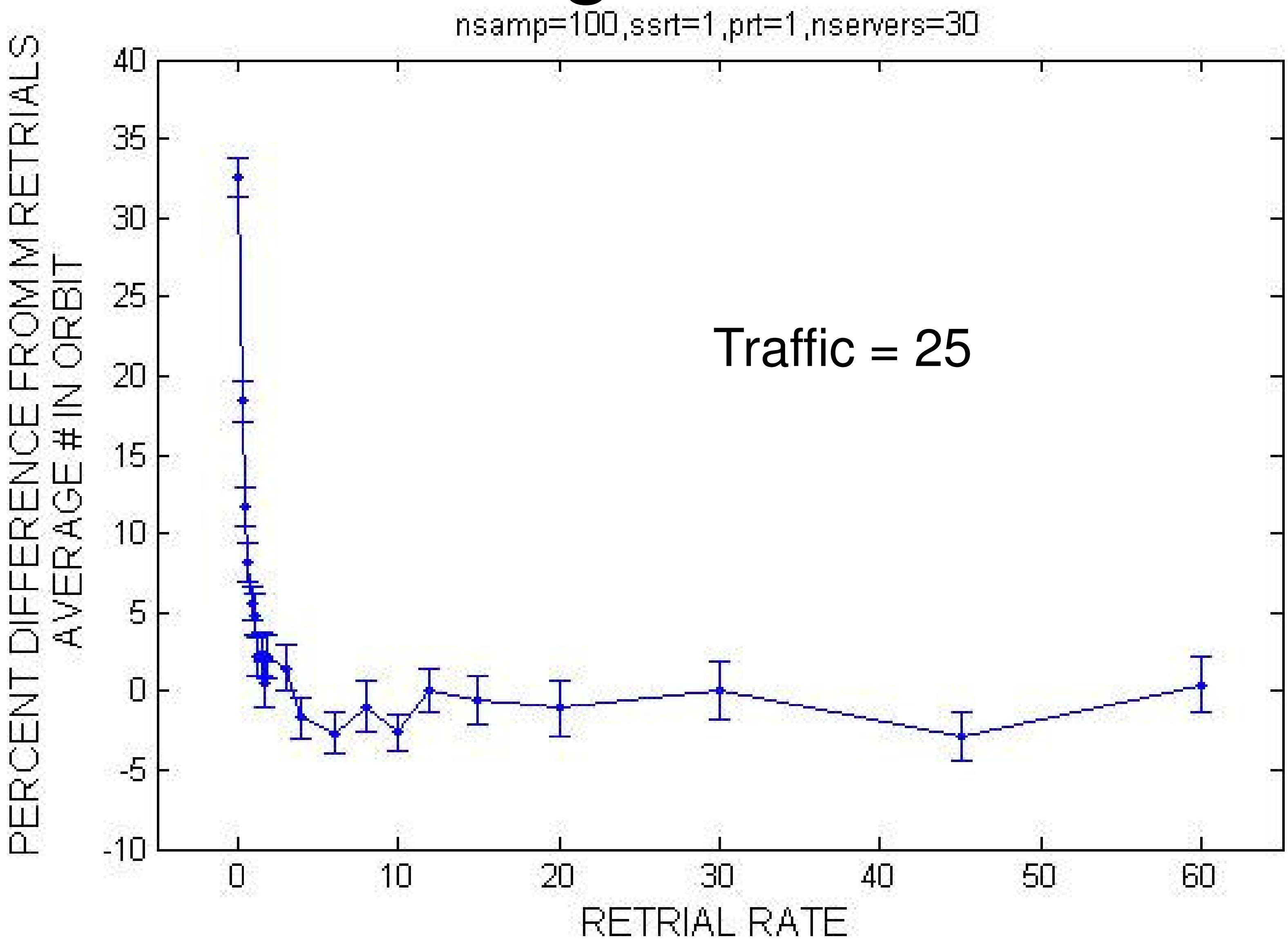
# % Diff: Average Number in Orbit



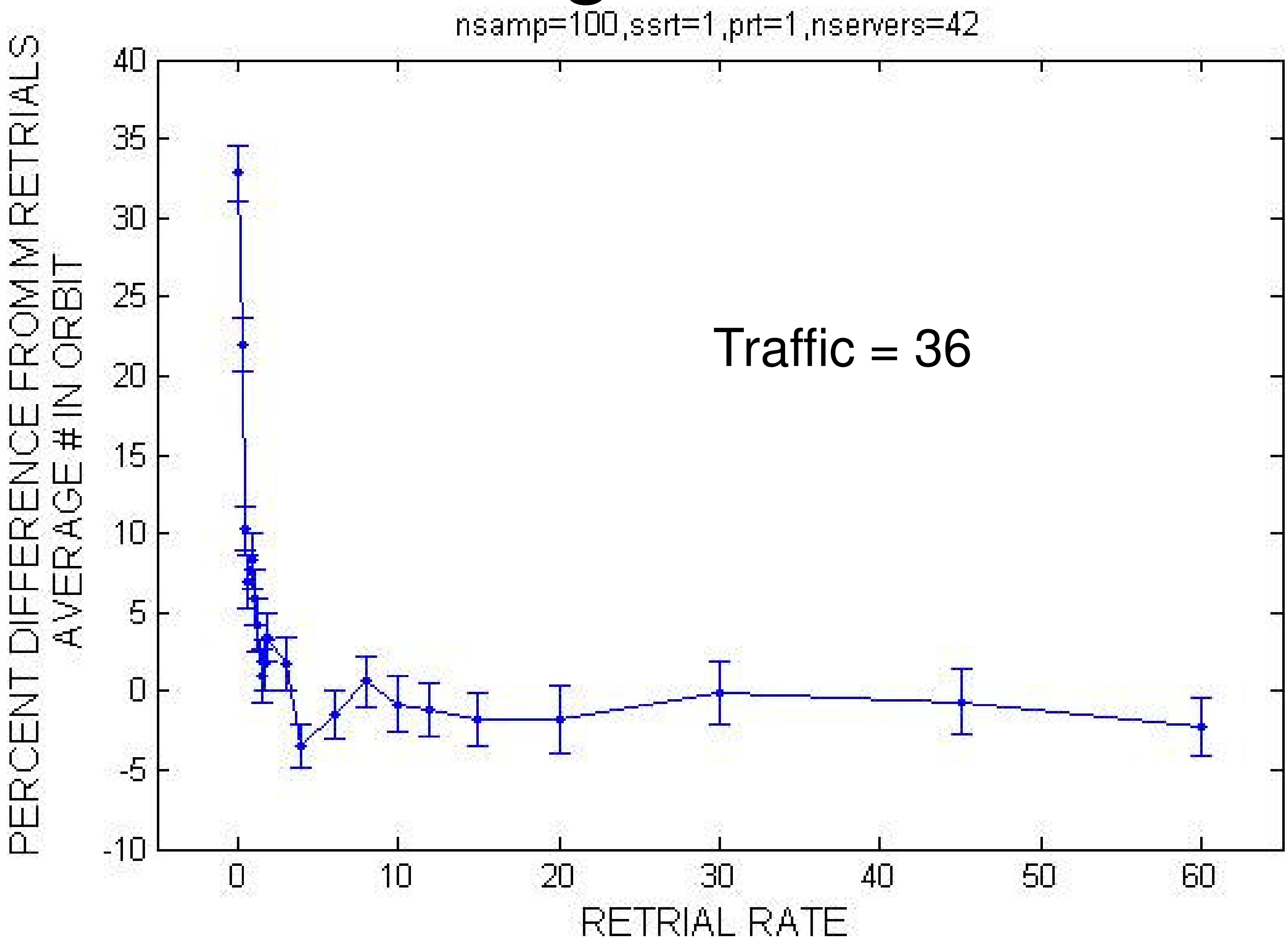
# % Diff: Average Number in Orbit



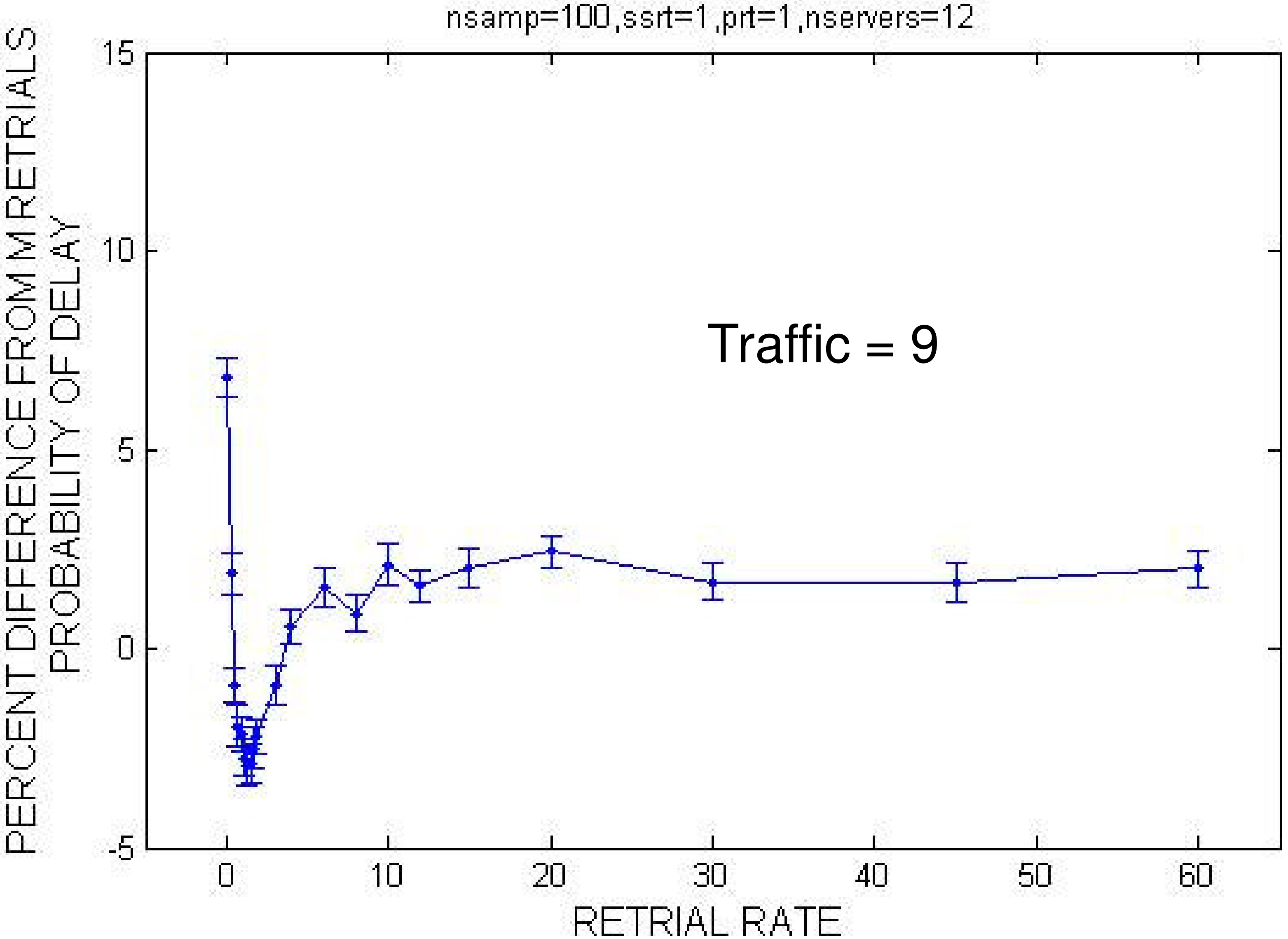
# % Diff: Average Number in Orbit



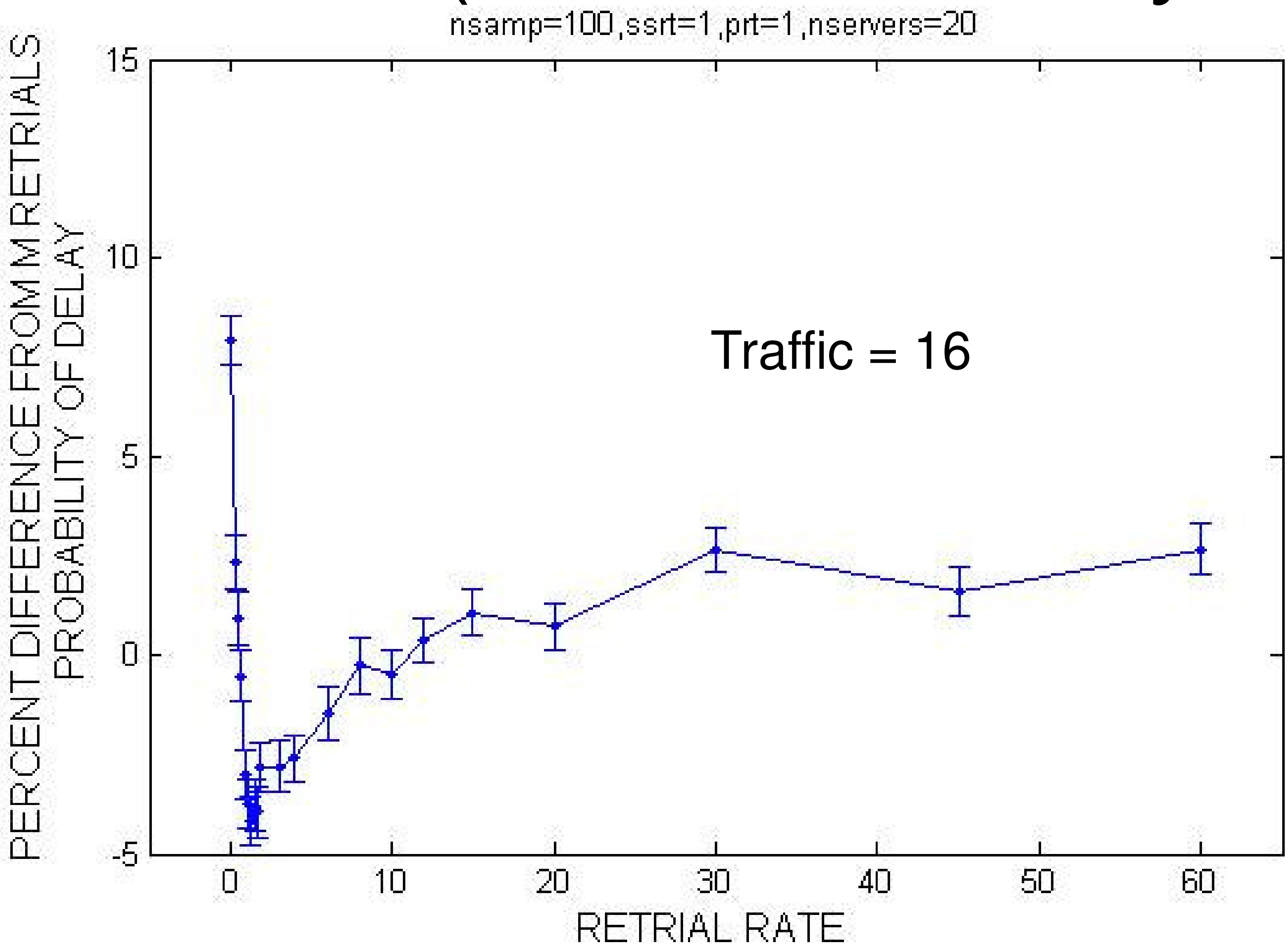
# % Diff: Average Number in Orbit



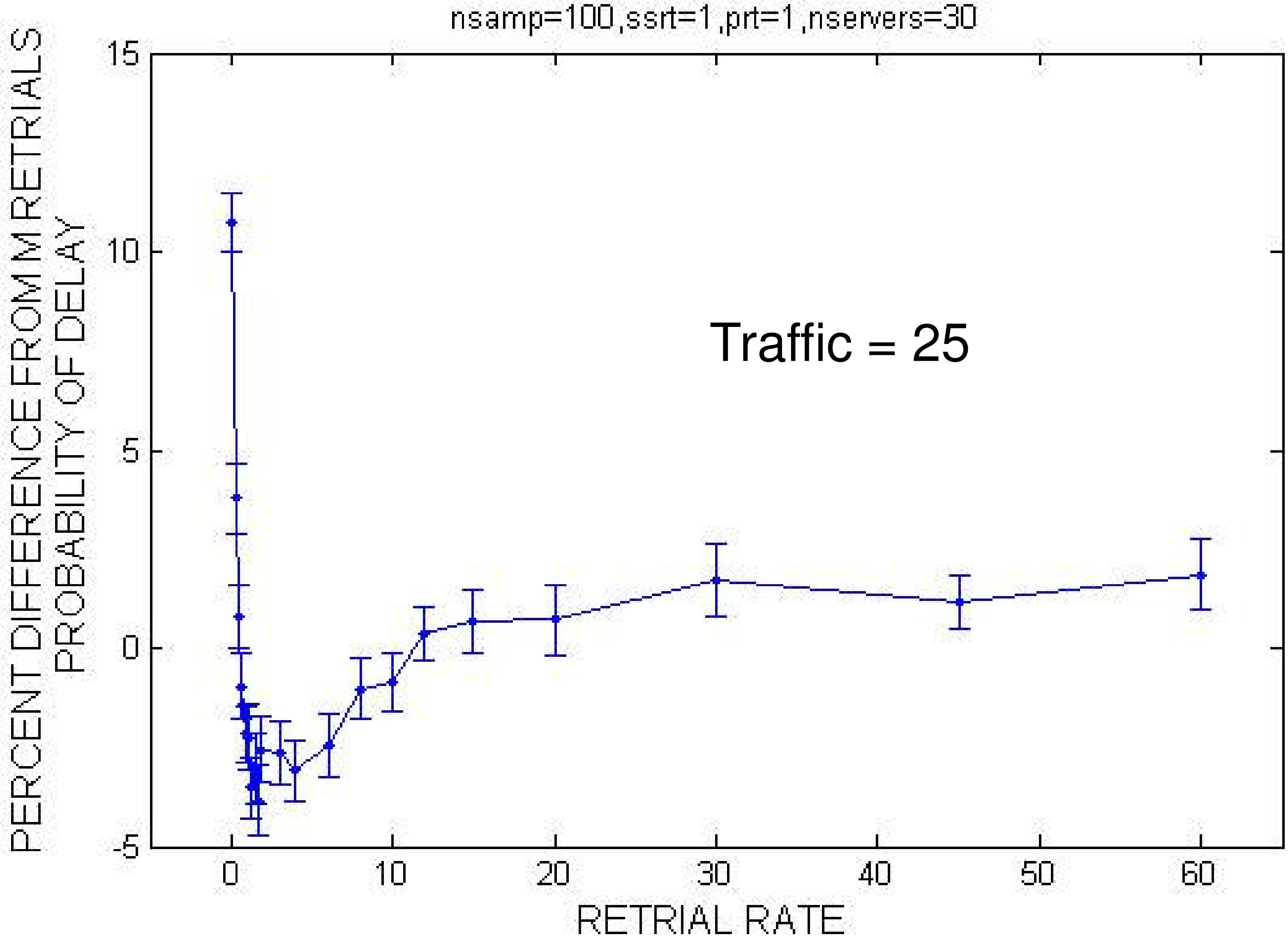
# % Diff: Pr(new arrival delayed)



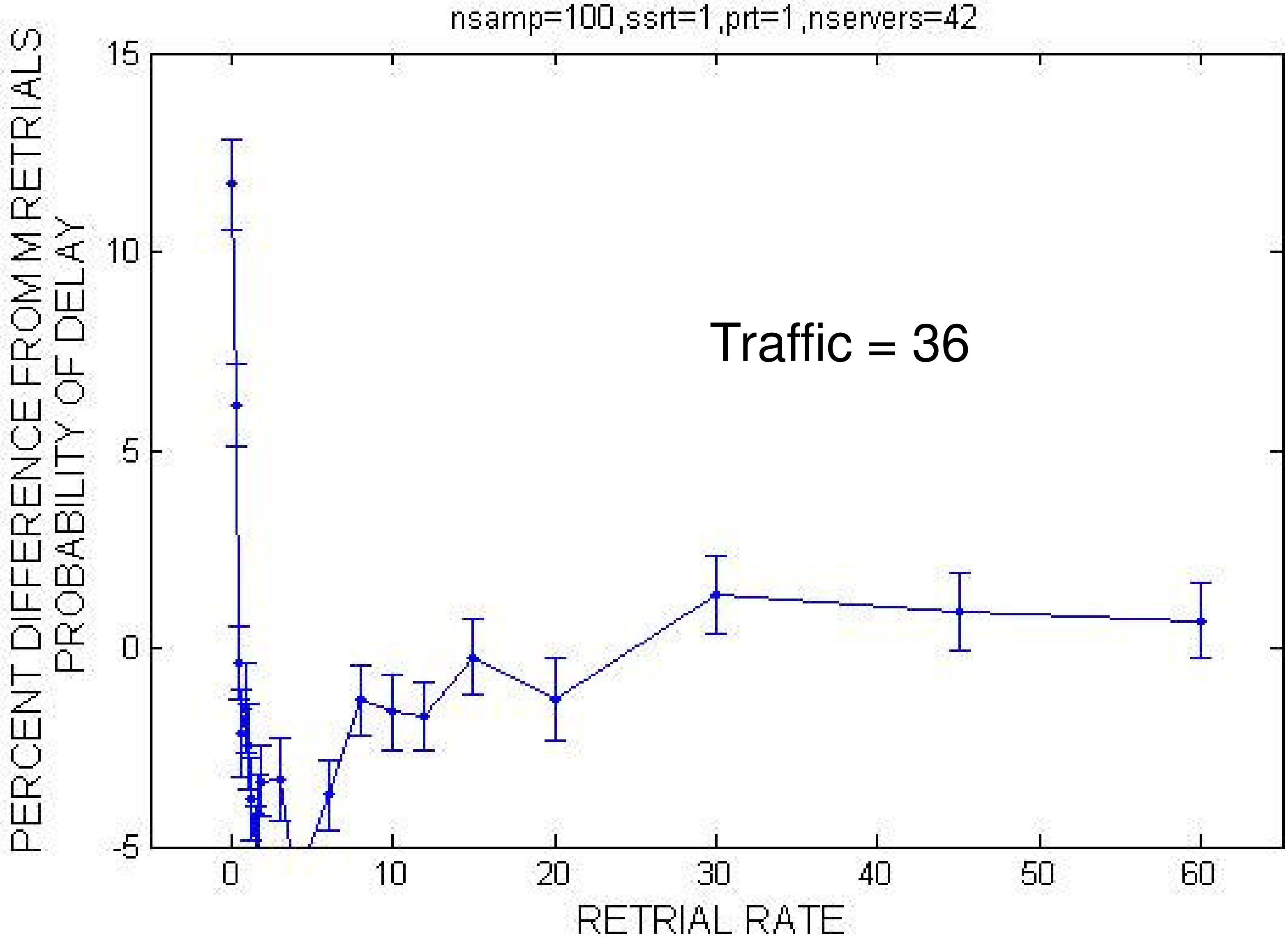
# % Diff: Pr(new arrival delayed)



# % Diff: Pr(new arrival delayed)



# % Diff: Pr(new arrival delayed)



WHY?

# Exponential Retrials

| | Retry 1 | Retry 2 | Retry 3 | Retry 4 |
|---|---|---|---|---|
| Job 1 | 10 | 27 | 4 | 22 |
| Job 2 | 19 | 11 | 23 | 5 |
| Job 3 | 7 | 51 | 13 | 17 |

# Personal Retrial Times (PRT)

| | Retry 1 | Retry 2 | Retry 3 | Retry 4 |
|---|---|---|---|---|
| Job 1 | 10 | 10 | 10 | 10 |
| Job 2 | 19 | 19 | 19 | 19 |
| Job 3 | 7 | 7 | 7 | 7 |

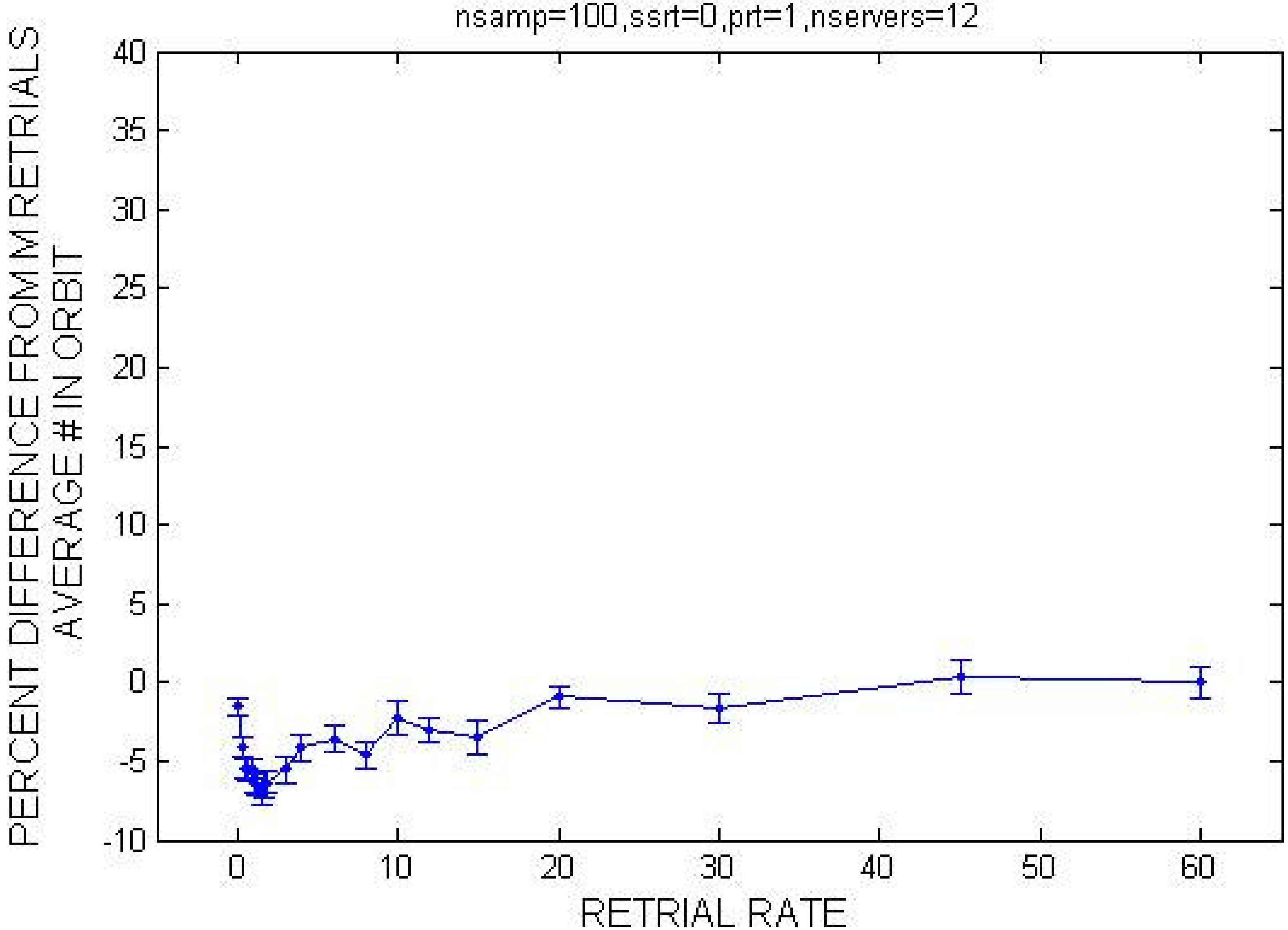
# Shared Sequence of Retrial Times (SSRT)

| | Retry 1 | Retry 2 | Retry 3 | Retry 4 |
|---|---|---|---|---|
| Job 1 | 10 | 27 | 4 | 22 |
| Job 2 | 10 | 27 | 4 | 22 |
| Job 3 | 10 | 27 | 4 | 22 |

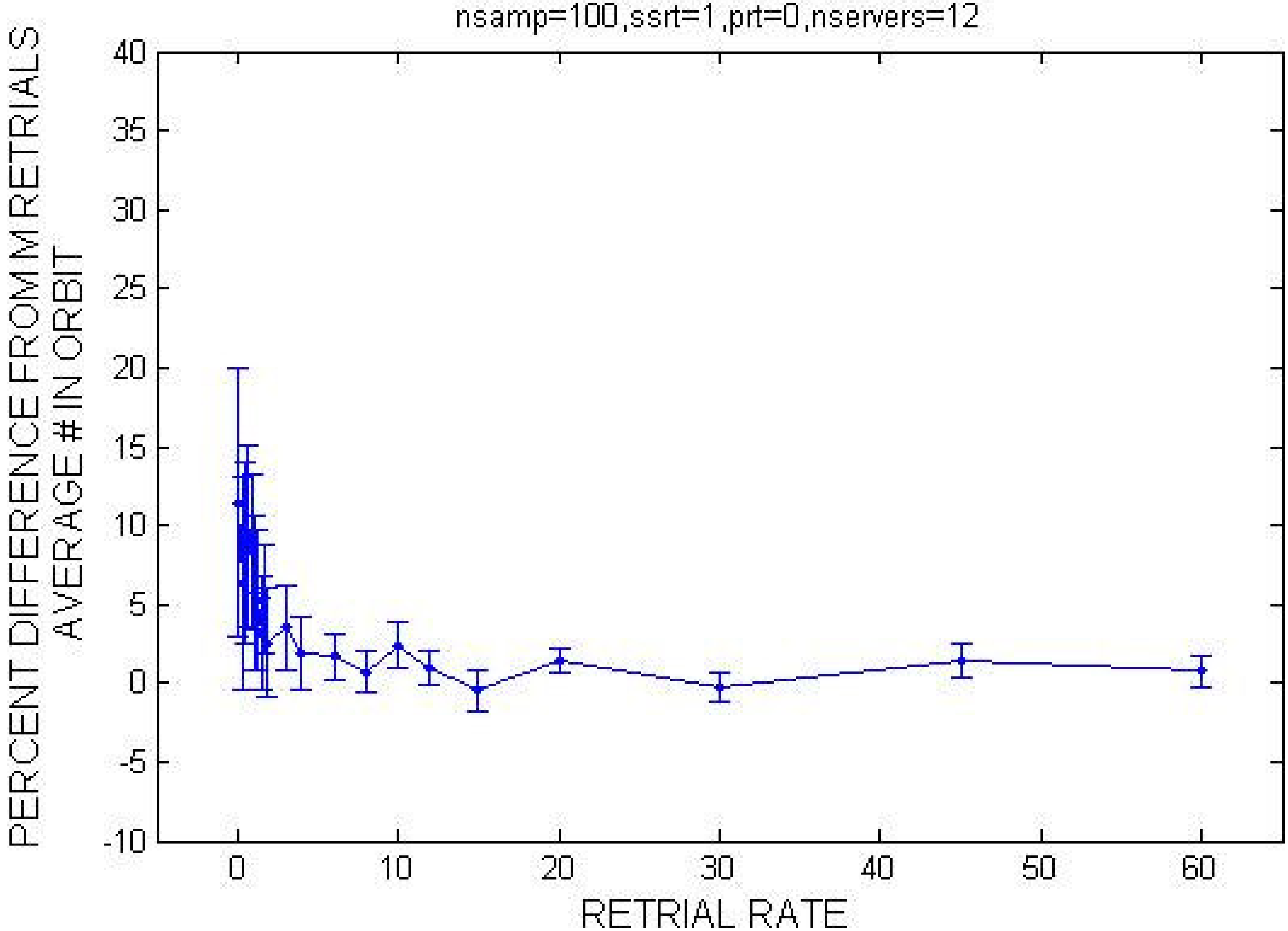
# Deterministic Retrials

| | Retry 1 | Retry 2 | Retry 3 | Retry 4 |
|---|---|---|---|---|
| Job 1 | 10 | 10 | 10 | 10 |
| Job 2 | 10 | 10 | 10 | 10 |
| Job 3 | 10 | 10 | 10 | 10 |

# % Diff in Lo: PRT vs M

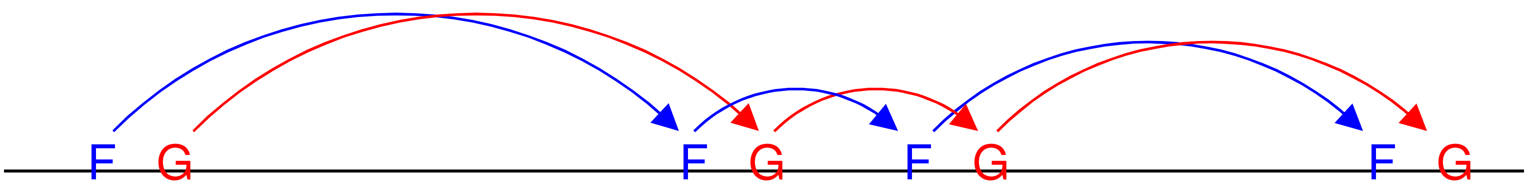


# % Diff in Lo: SSRT vs M



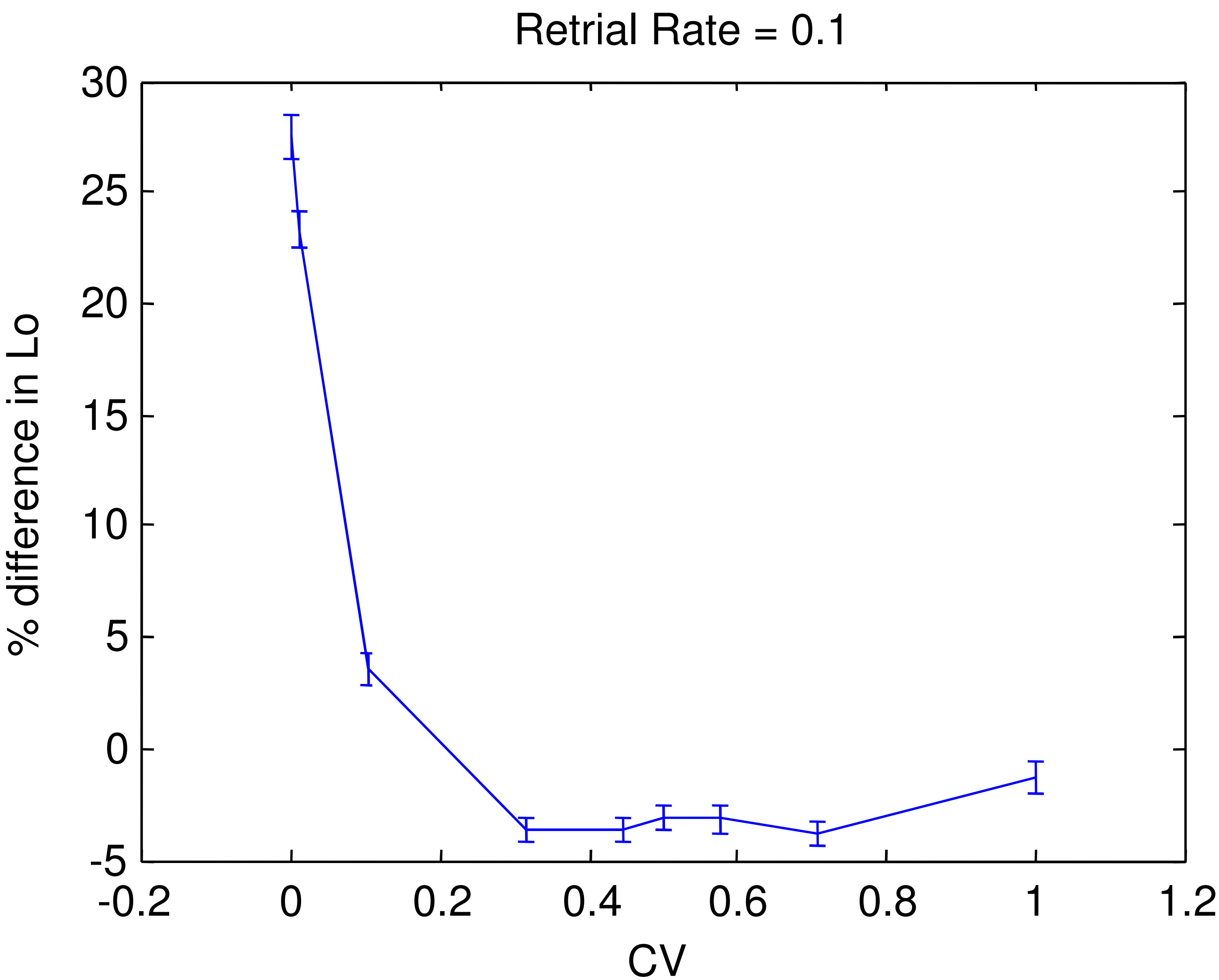
# Why? Because:

- Shared Sequence of Retrial Times is the dominant effect.



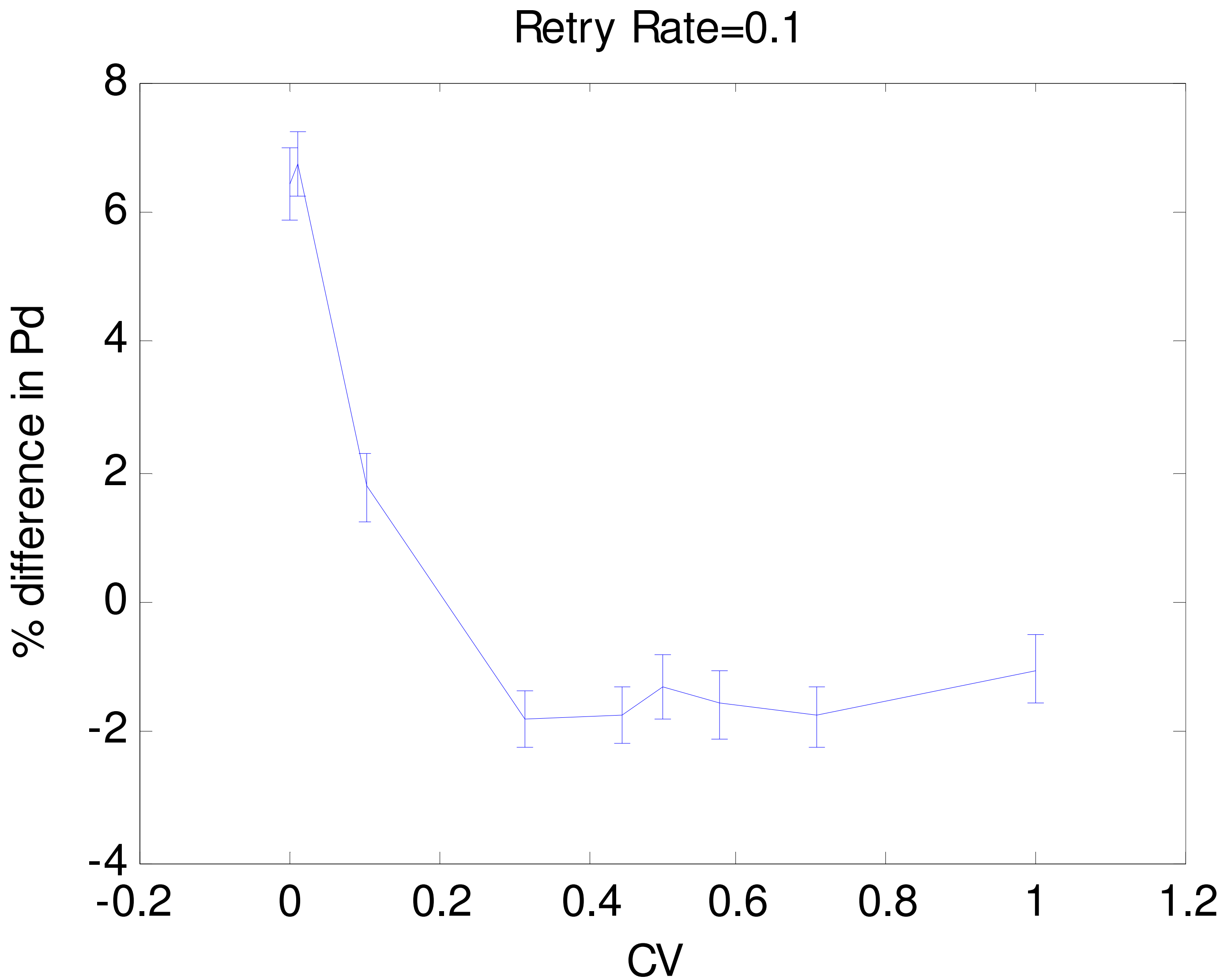

- How deterministic does it have to be?
- We will change the Coefficient of Variation (CV)

# % Diff in Lo



Retrial Rate = 0.1
30
25
20
15
10
5
0
-5
% difference in Lo
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
CV

# % Diff in Pr(delay)



# Markovian Approach

- $M/M/c/0 + PH_2$ retrials

- Lower limit on variability:
  - Two-phase Erlang has Squared Coefficient of Variation = 1/2

- Can get lower SCV using negative(!) probabilities

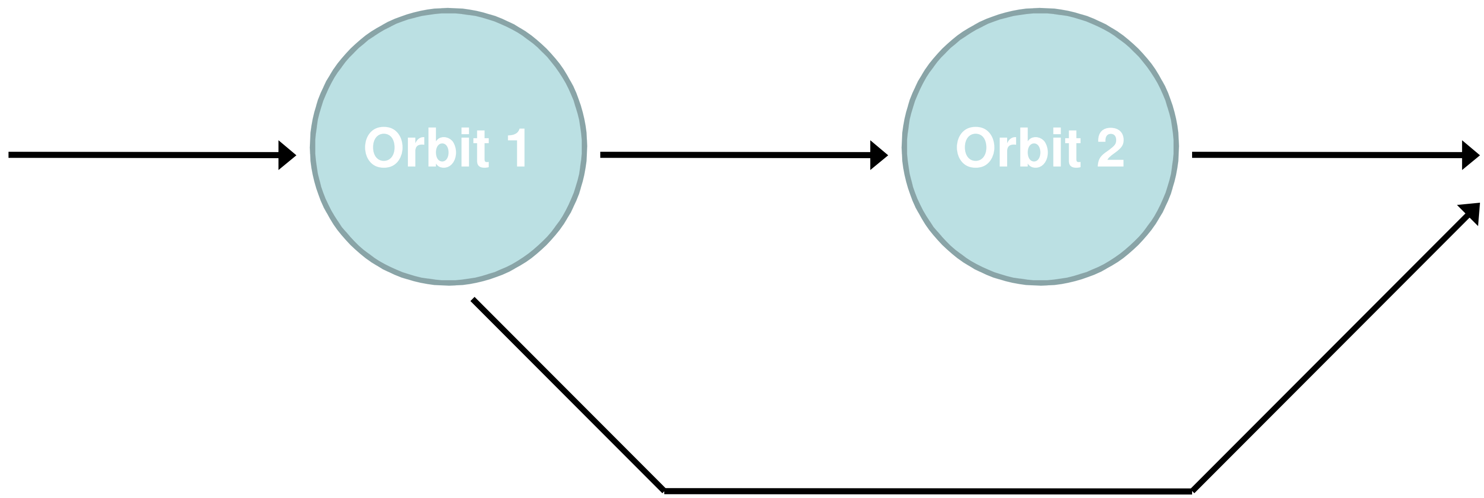
# Extended Probabilities

- 1955, Cox: Complex probabilities
- 1987, Nojo and Watanabe:

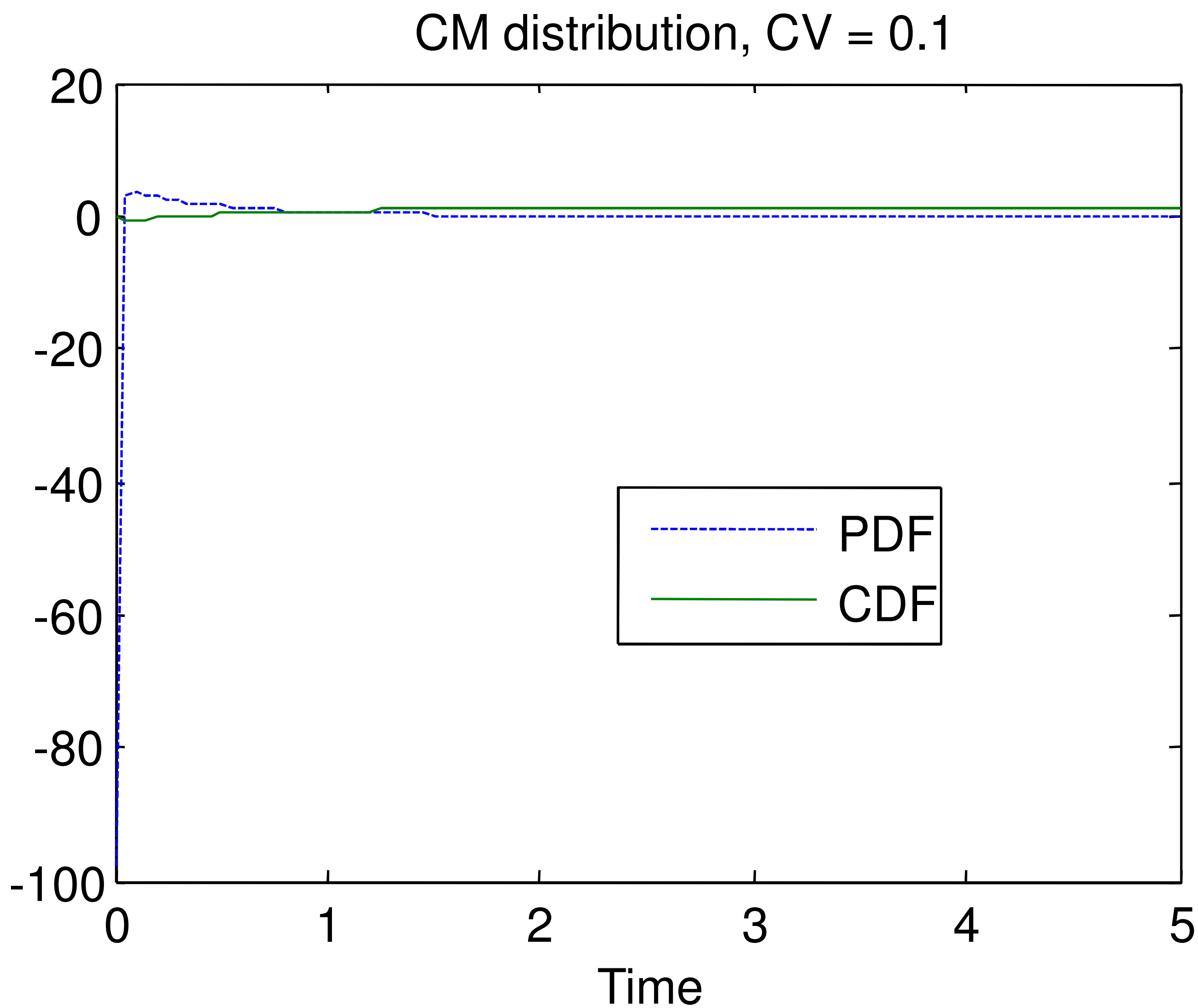
  Negative branching Probability (NP) distrib.
- 1994, Graham, Knuth, Patashnik
- 1999, Ball et al.:

  $H_2^*$ distribution
- 2007/8, Tijms: M/D/1 via $M/PH_2/1$
- Quantum physics

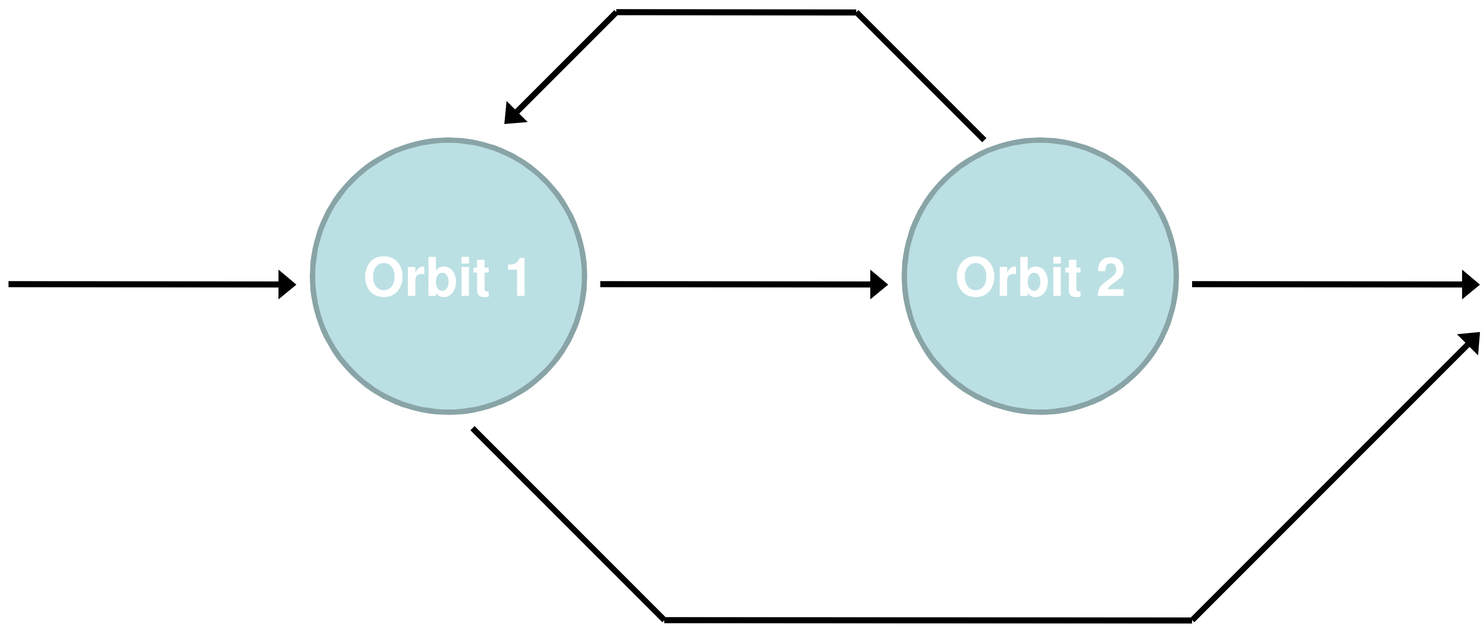
# Cox-Marie distribution



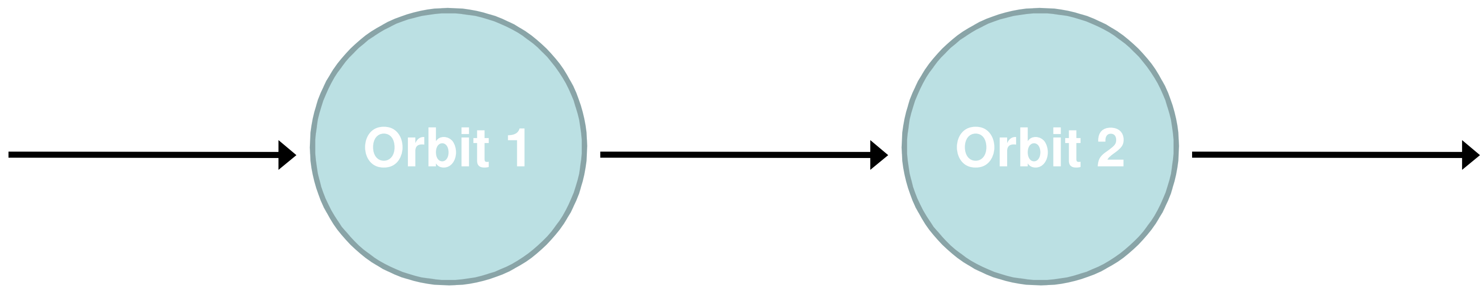
# Cox-Marie distribution



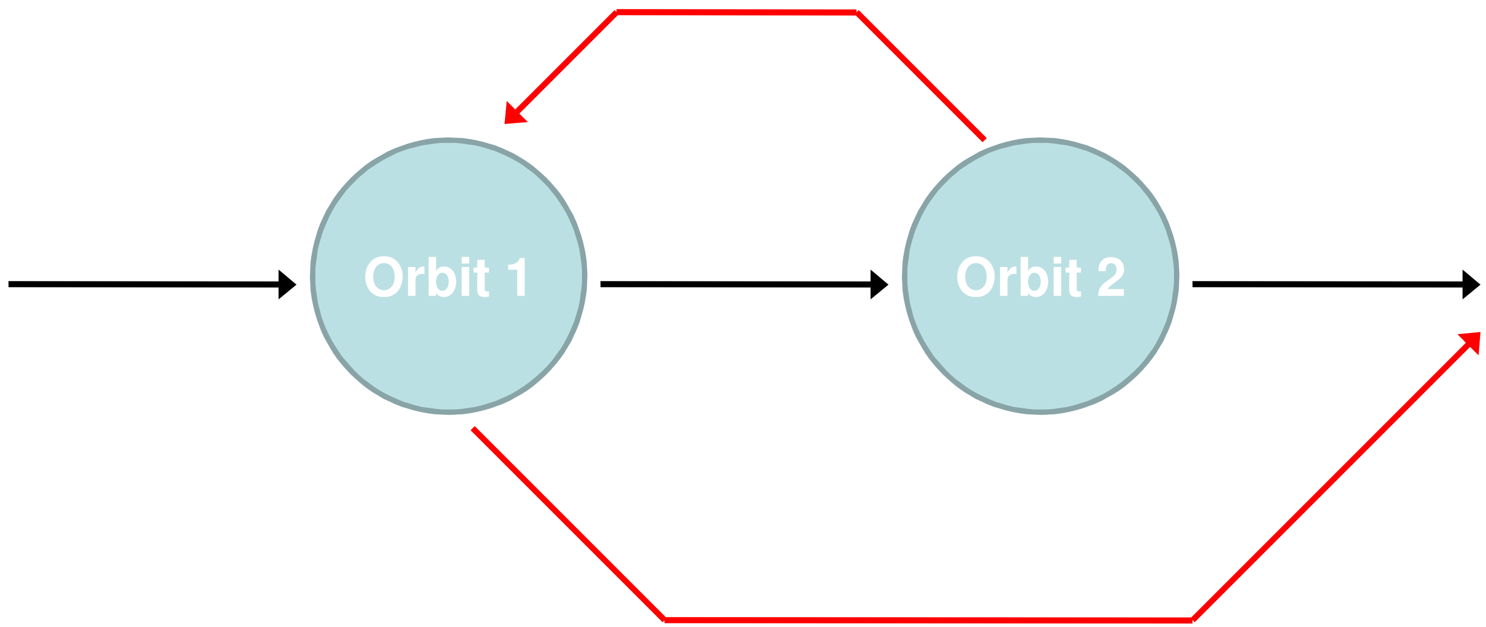
# NP distribution, SCV > 1/2



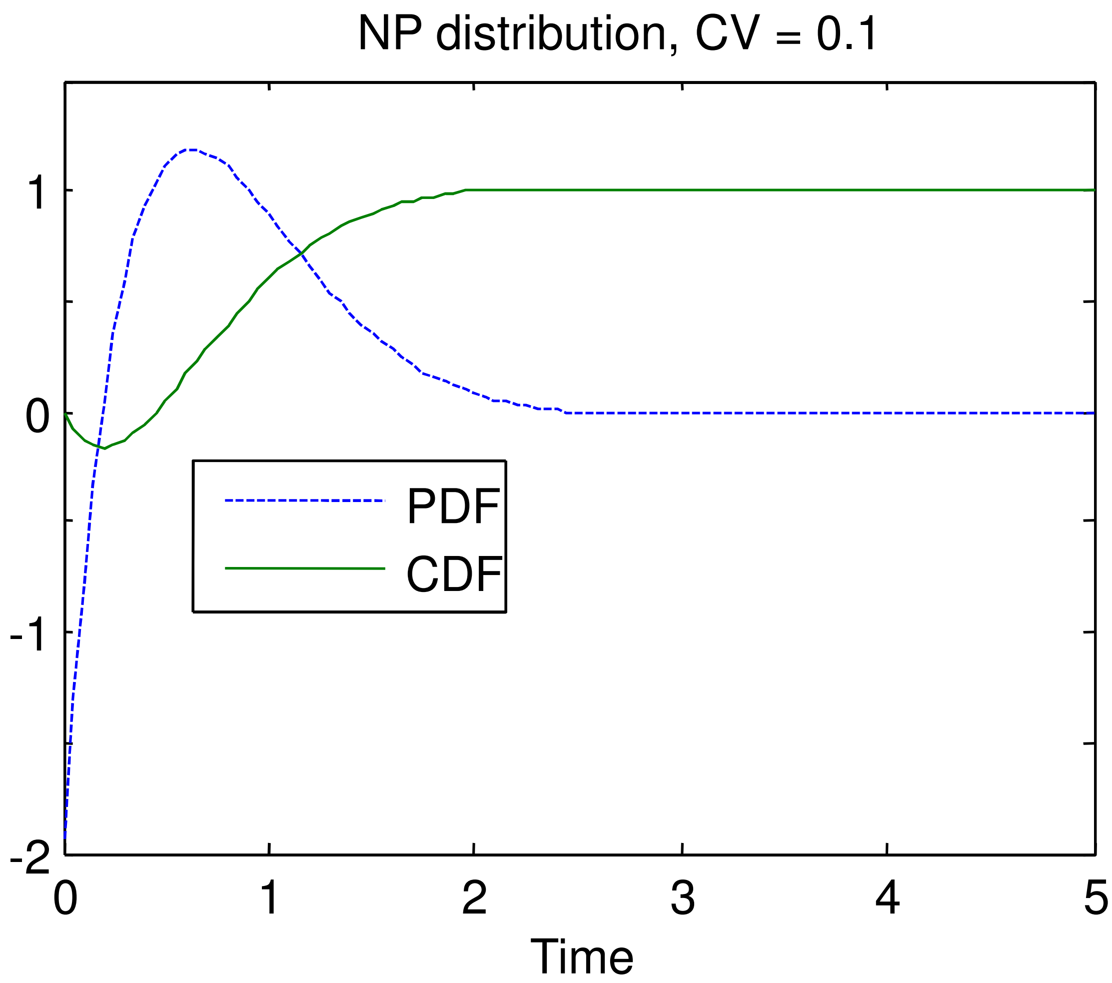
# NP distribution, SCV = 1/2



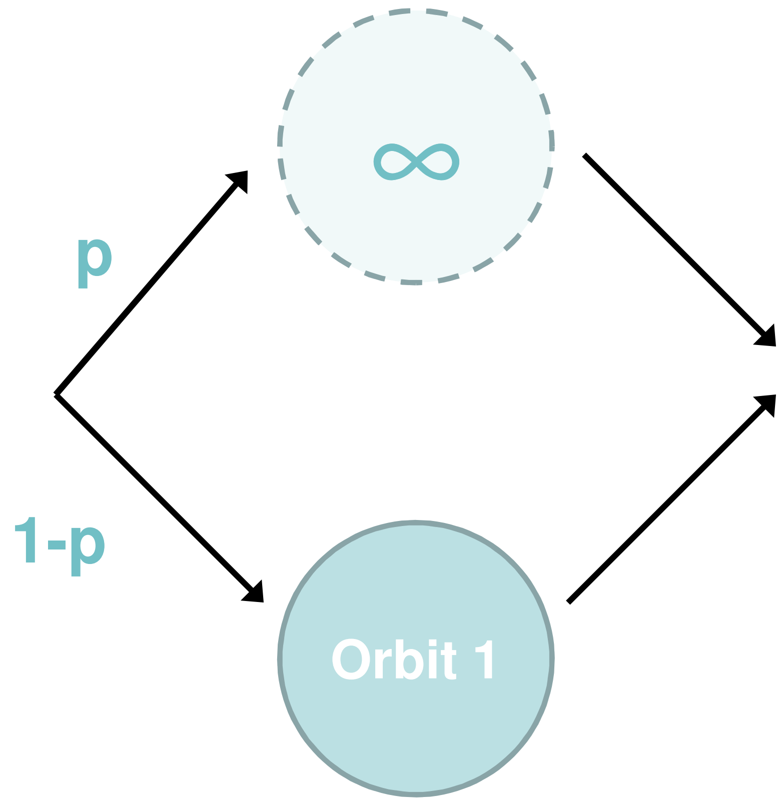
# NP distribution, SCV < 1/2



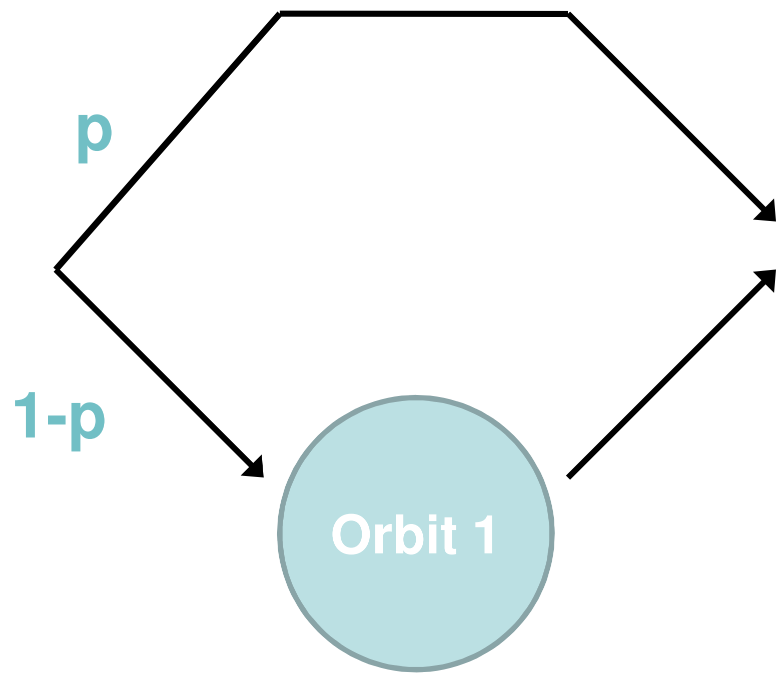
# NP Distribution



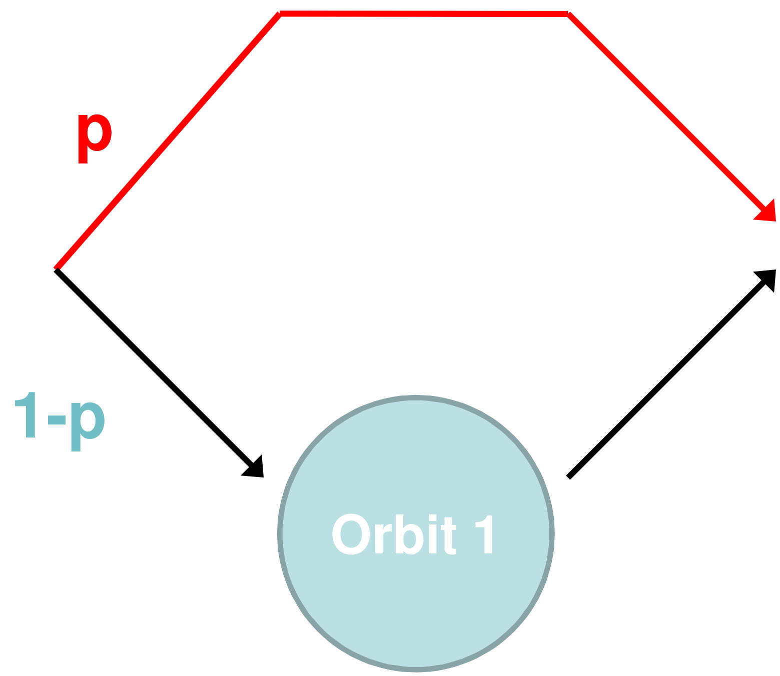
# $H_2^*$ distribution



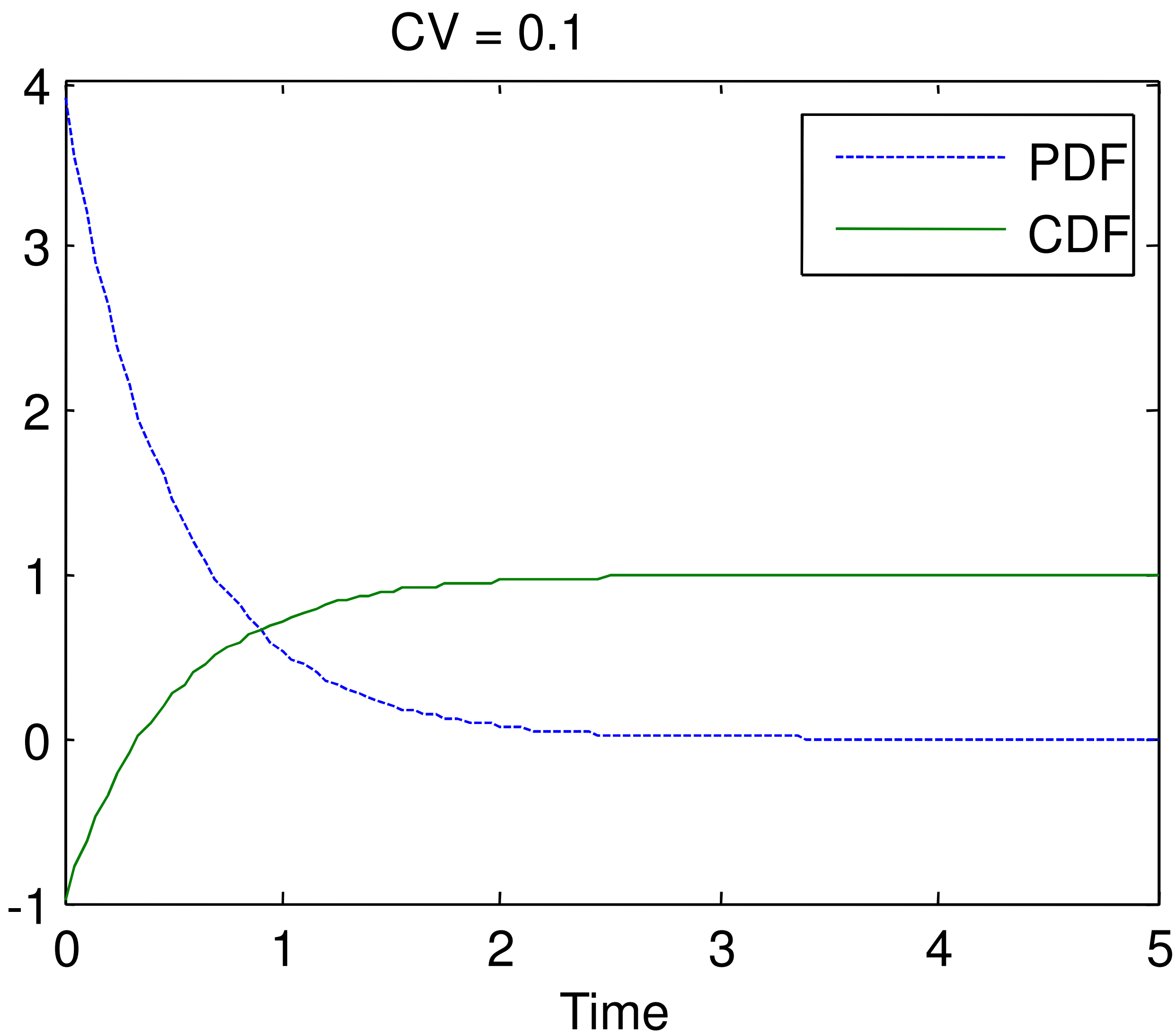
# $H_2^*$ distribution



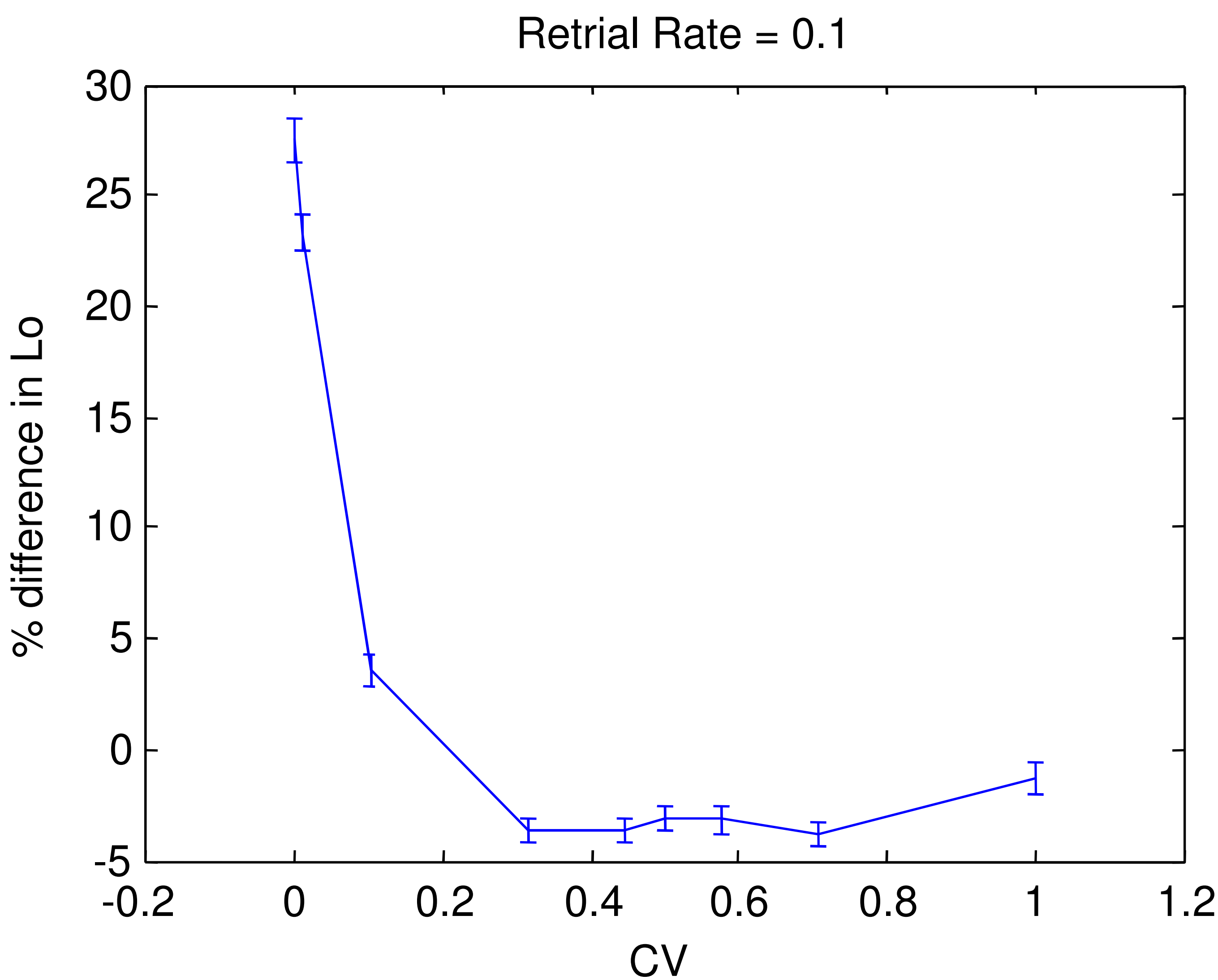
# $H_2$* distribution, SCV < 1

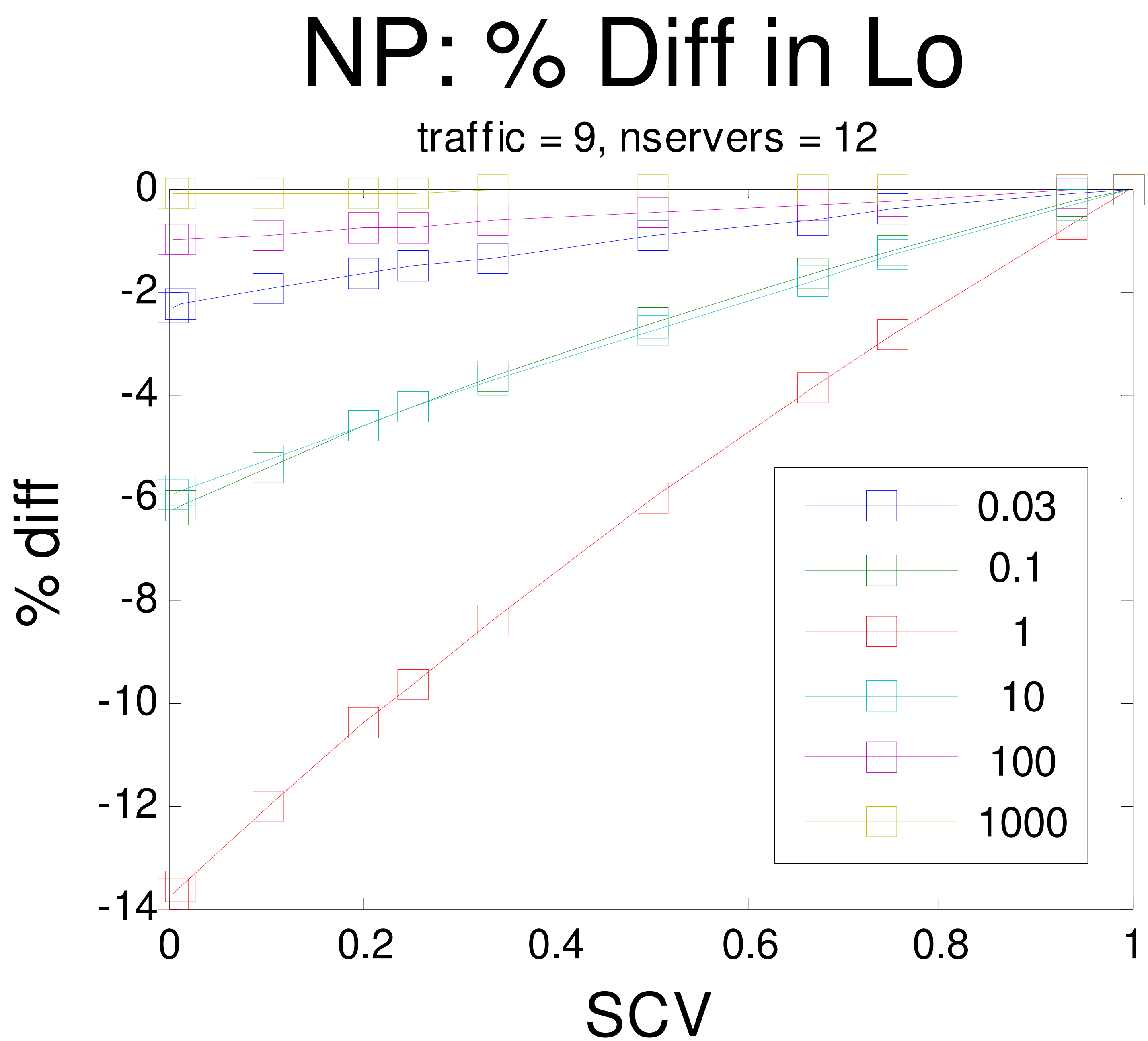


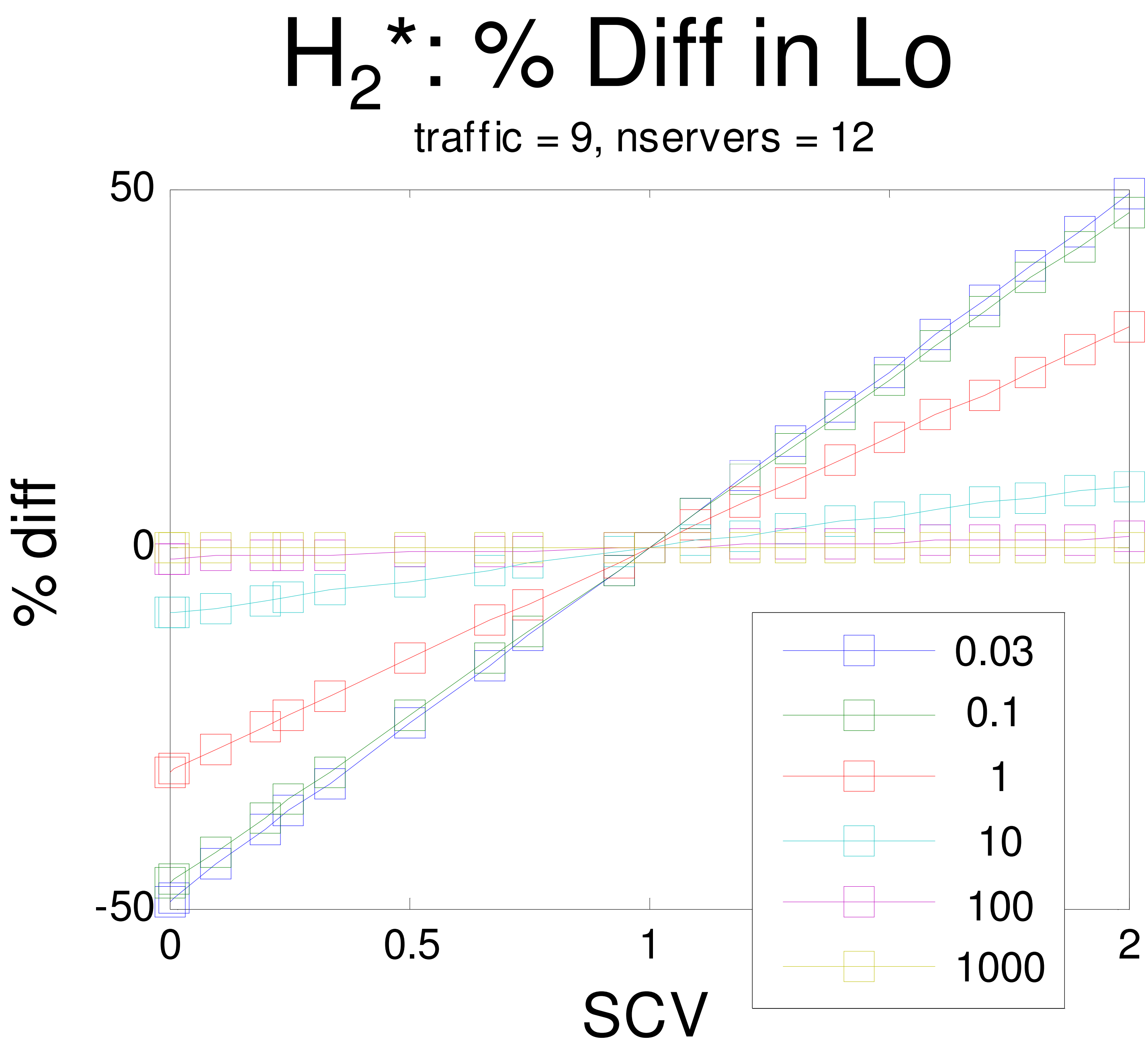
# $H_2$* distribution



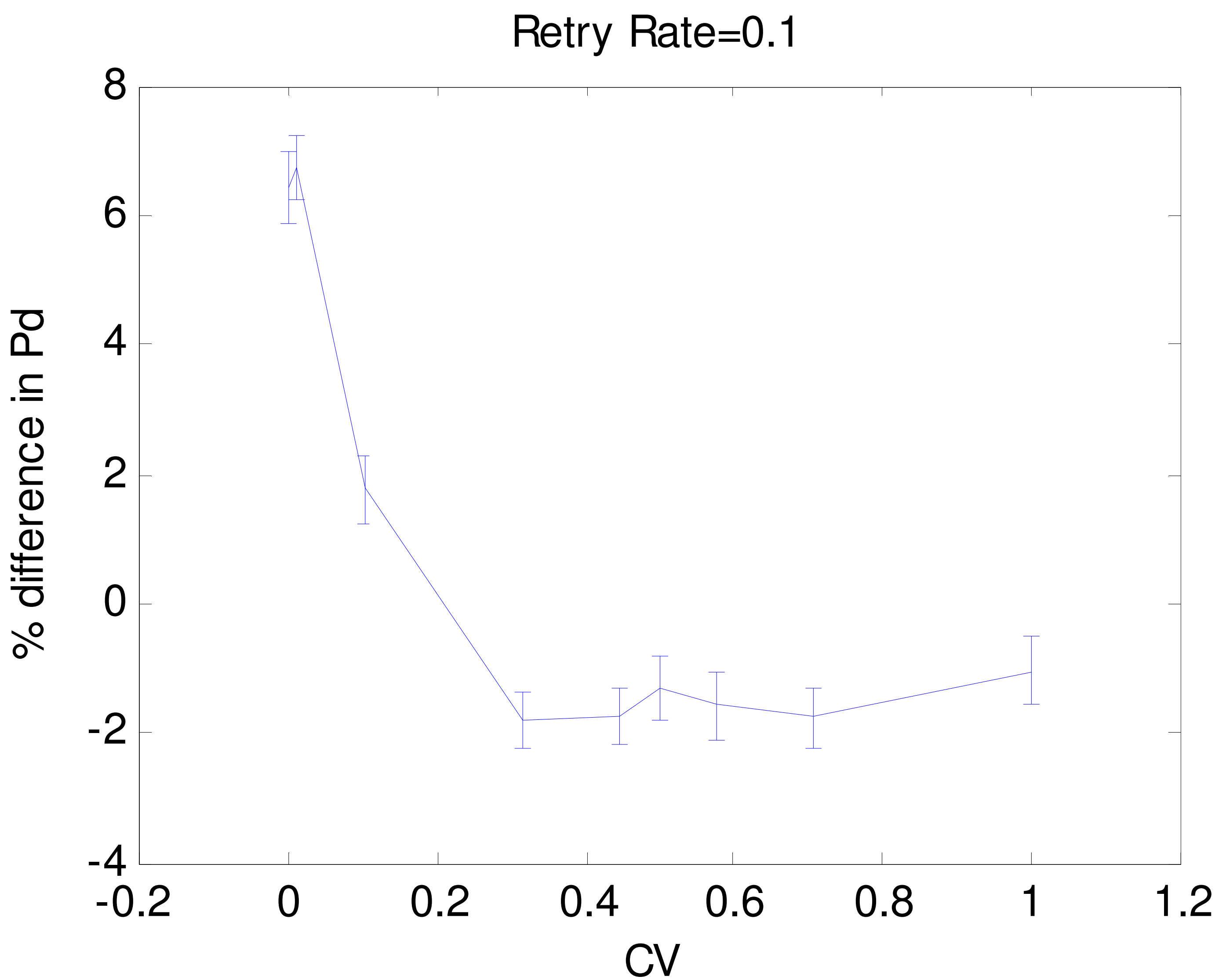
# Recall our simulations: Lo



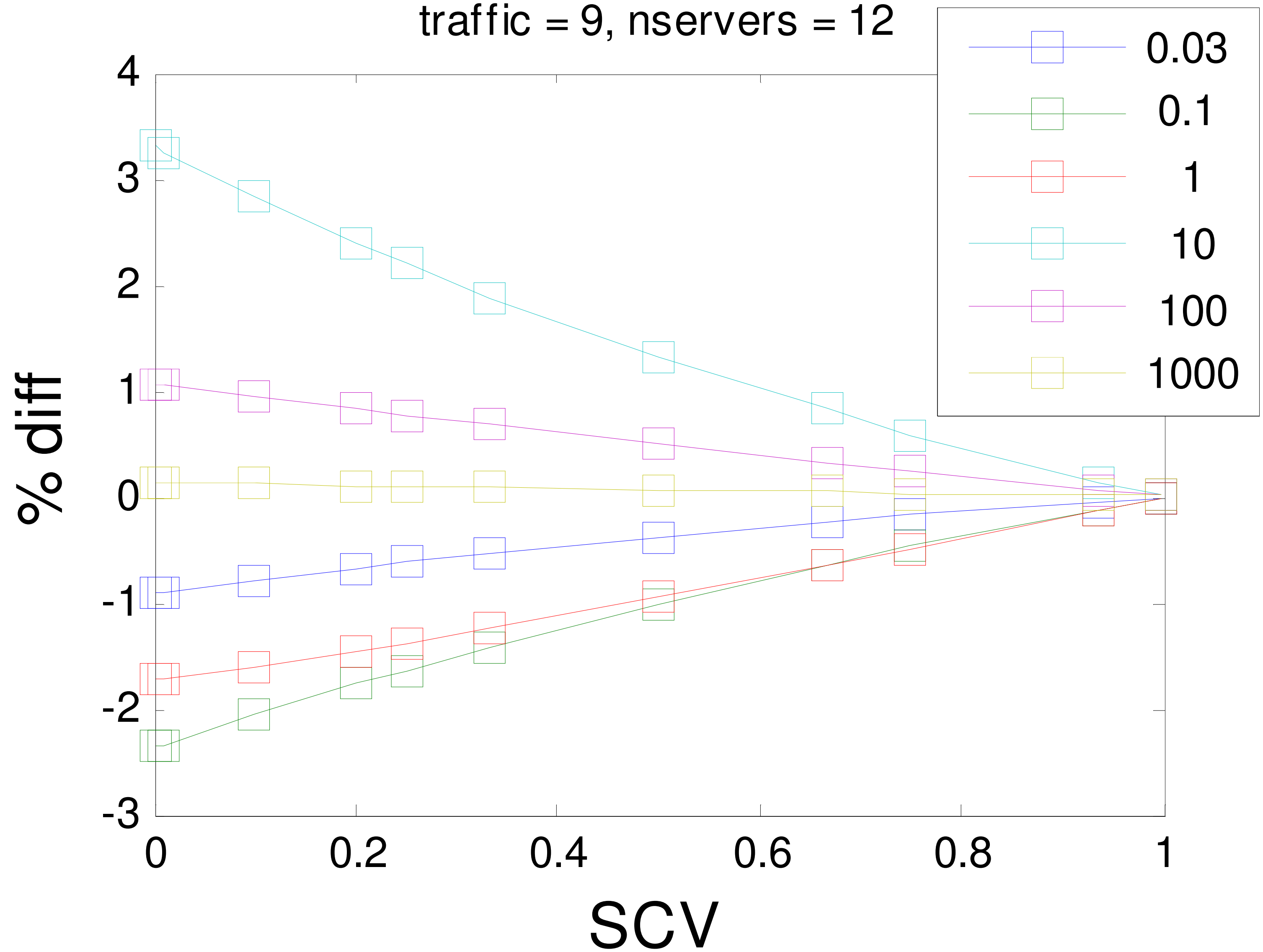

NP: % Diff in Lo
traffic = 9, nservers = 12
% diff
SCV
0
-2
-4
-6
-8
-10
-12
-14
0
0.2
0.4
0.6
0.8
1
0.03
0.1
1
10
100
1000


$H_2$*: % Diff in Lo
traffic = 9, nservers = 12
50
0
-50
% diff
0
0.5
1
2
SCV
0.03
0.1
1
10
100
1000

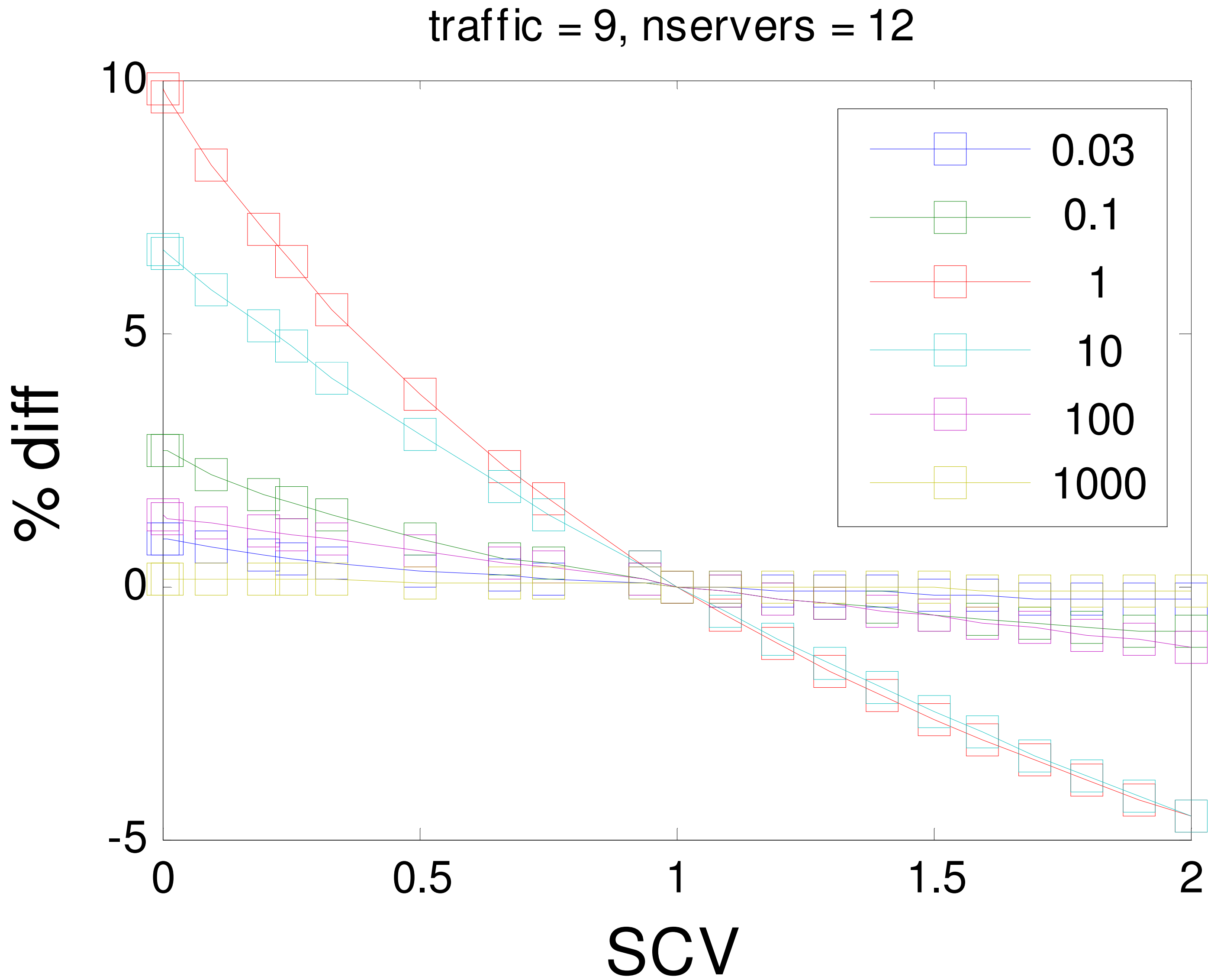
# Recall: % Diff in Pr(delay)



# NP: % Diff in Pr(delay)



# $H_2$*: % Diff in Pr(delay)



# Conclusions

- Do not use exponential retrials as an approximation to G-retrials when $CV < 0.1$ and retrial rate $<= 0.1$
- NP and $H_2^*$ distributions do not replicate simulations at low CV

# Queue-and-eh?

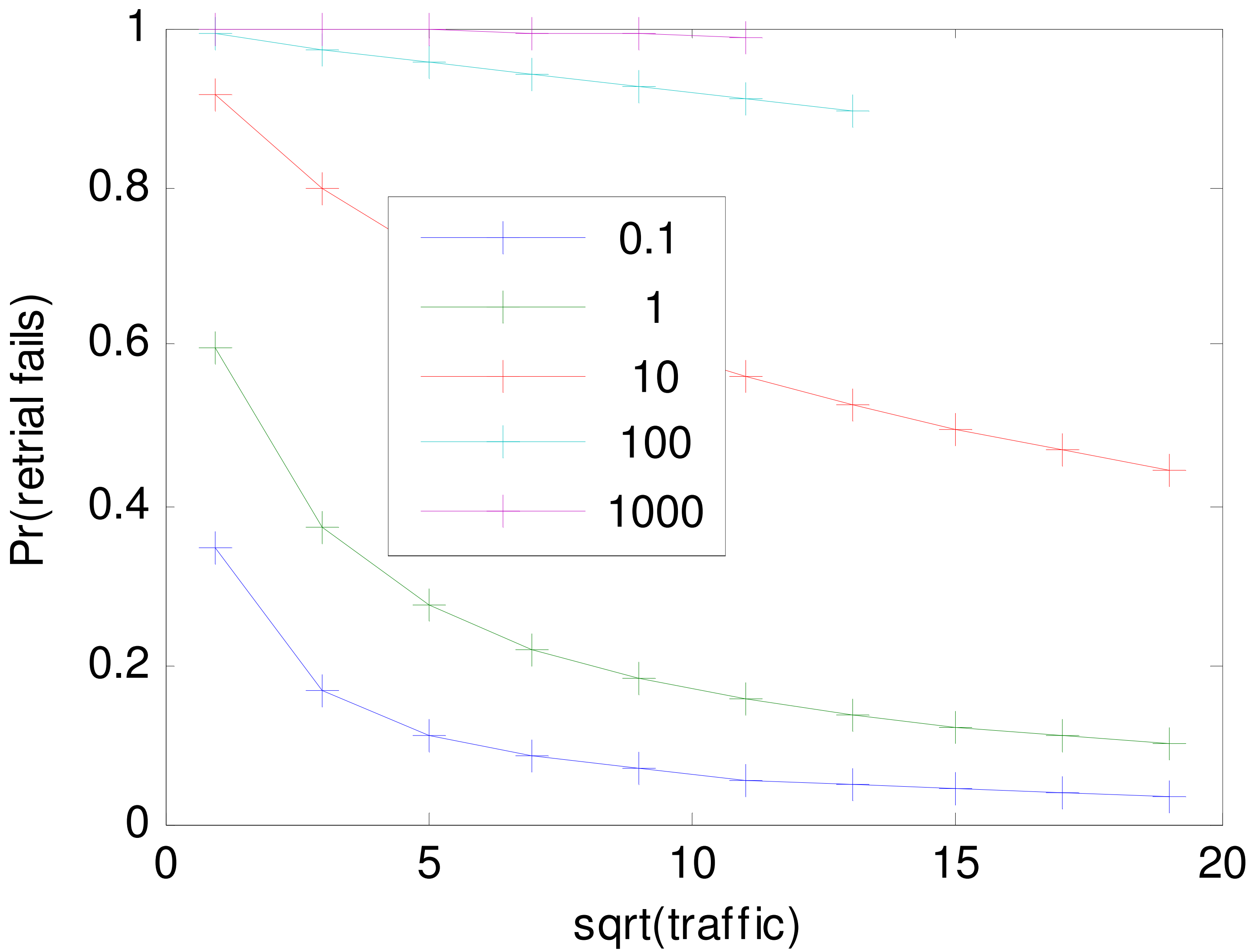

- Andrew Ross, andrew.ross@emich.edu
- David Lubke, dlubke@emich.edu
- Andrew Livingston, alivings@emich.edu
- Katie Ballentine, knballentine@gmail.com

# Appendix

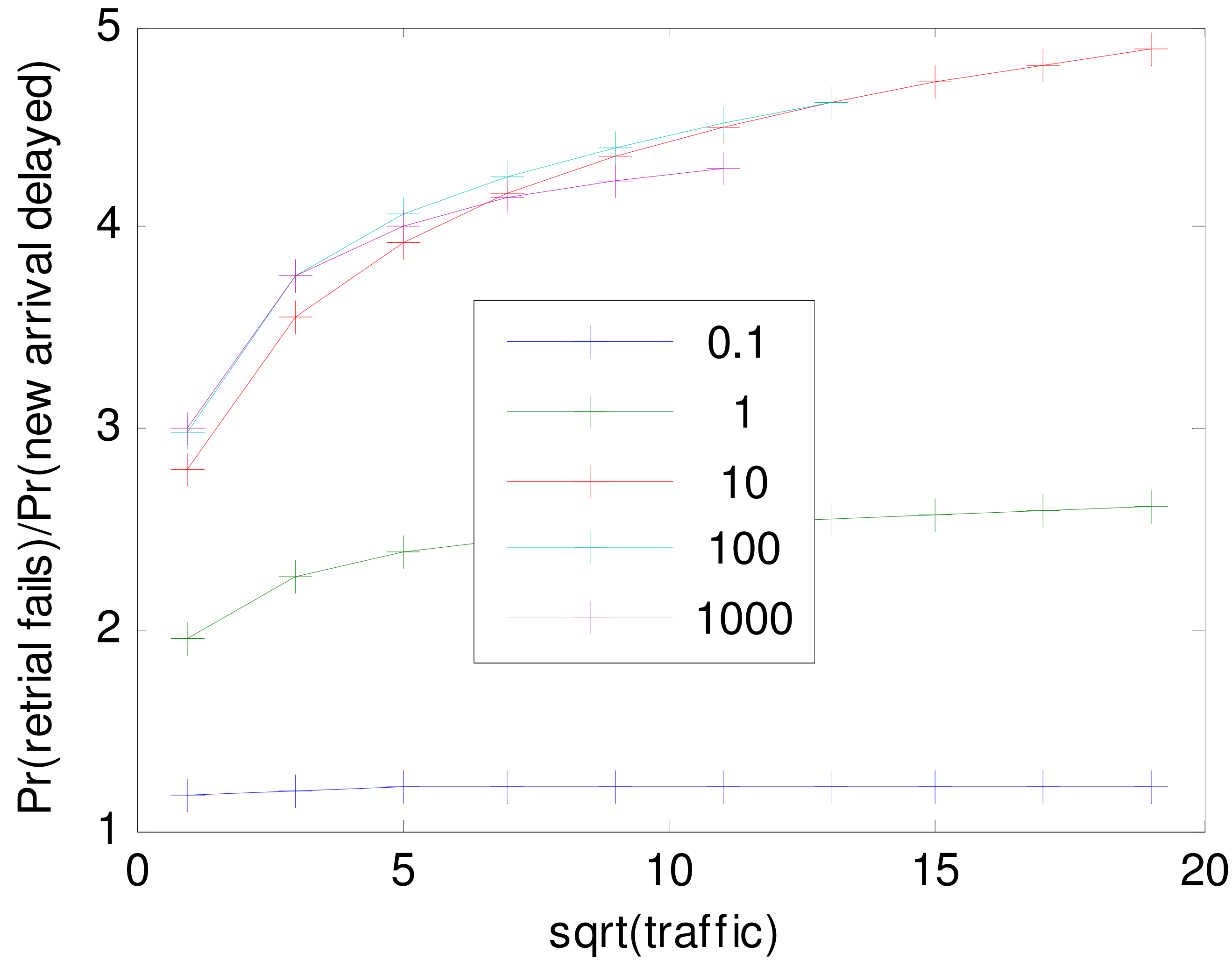
# General-Retrials literature

- Yang, Posner, Templeton, Li (1994): An approximation for M/G/1+G-retrials
- Many authors: only one person in orbit may retry (“constant retrial policy”)

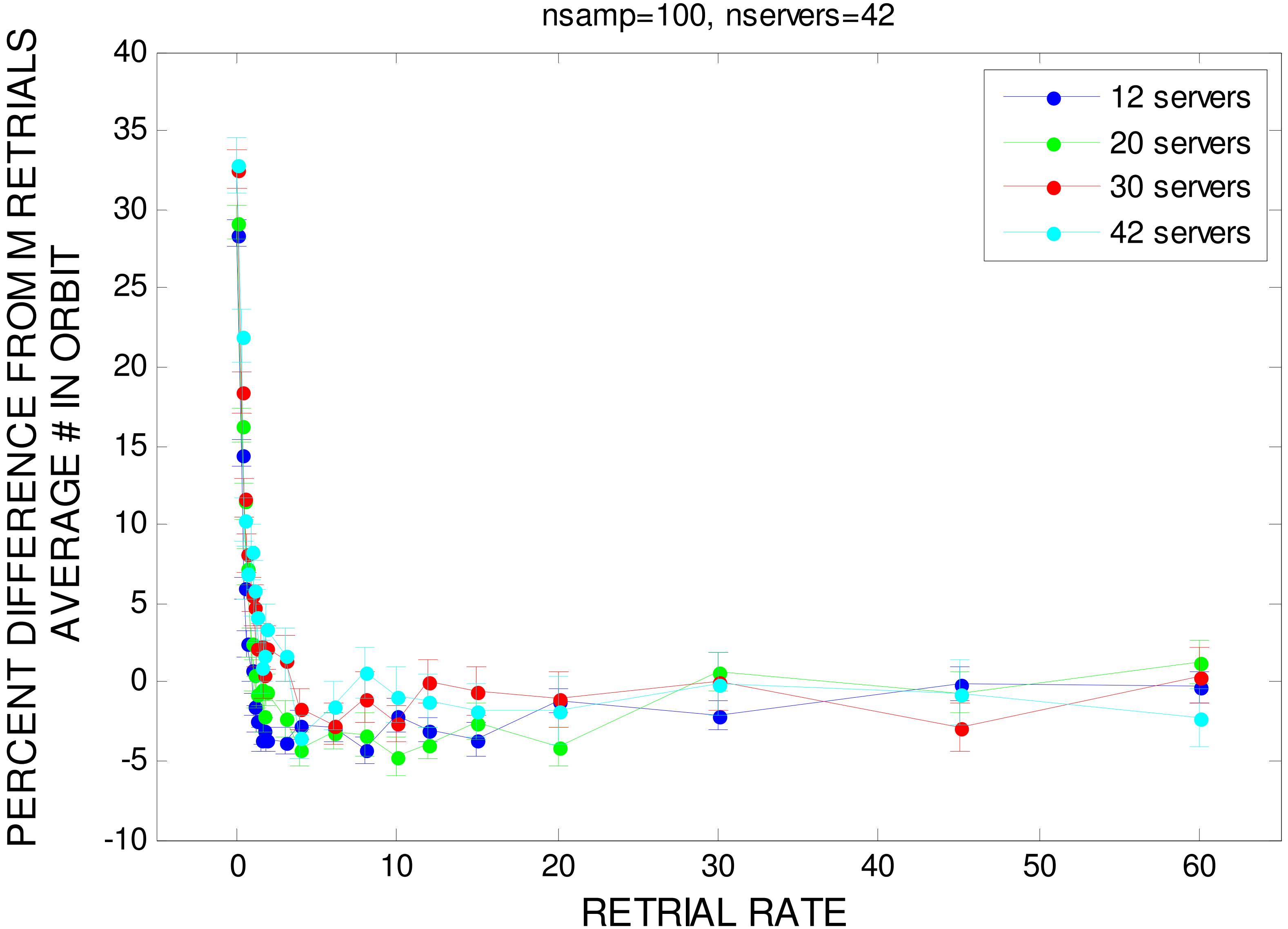
# Pr(retry fails) as system grows



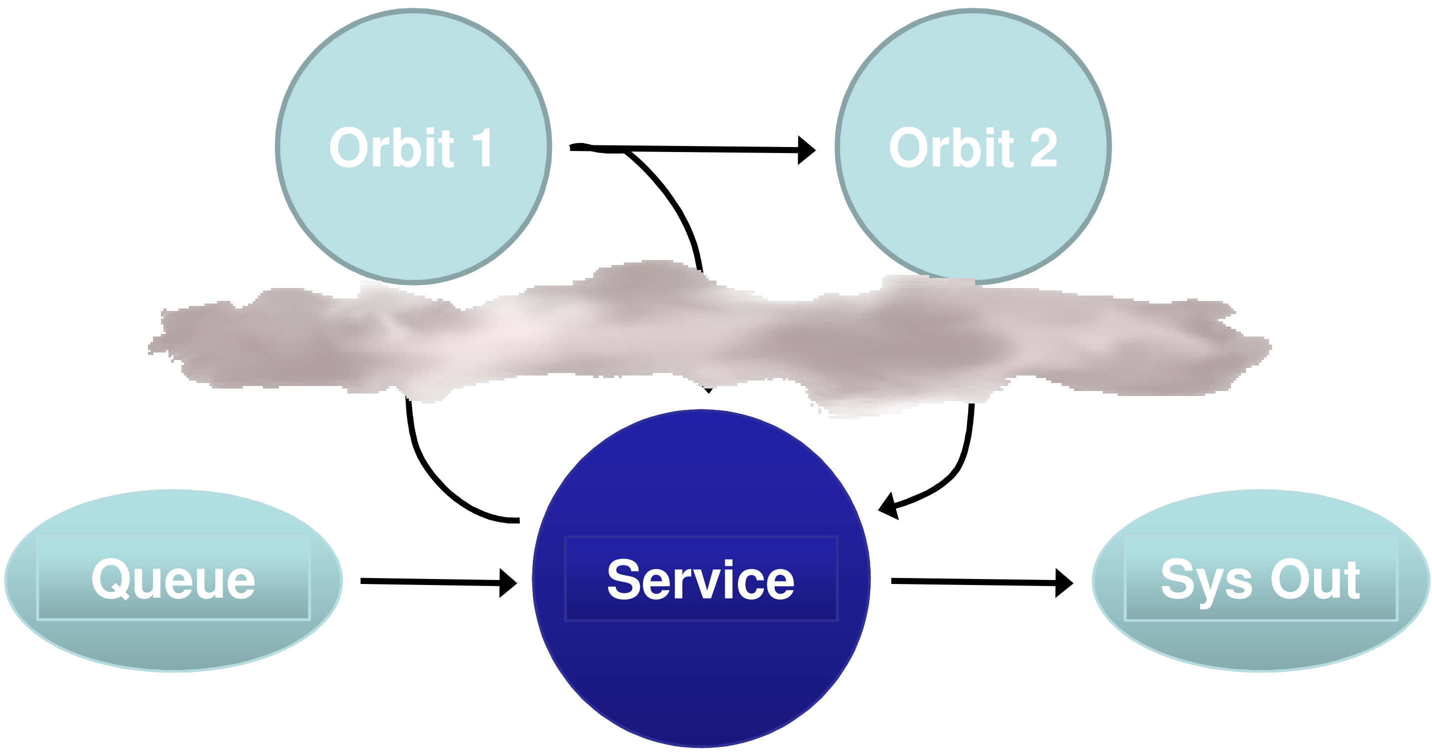
# P(retry fail)/P(new fail)



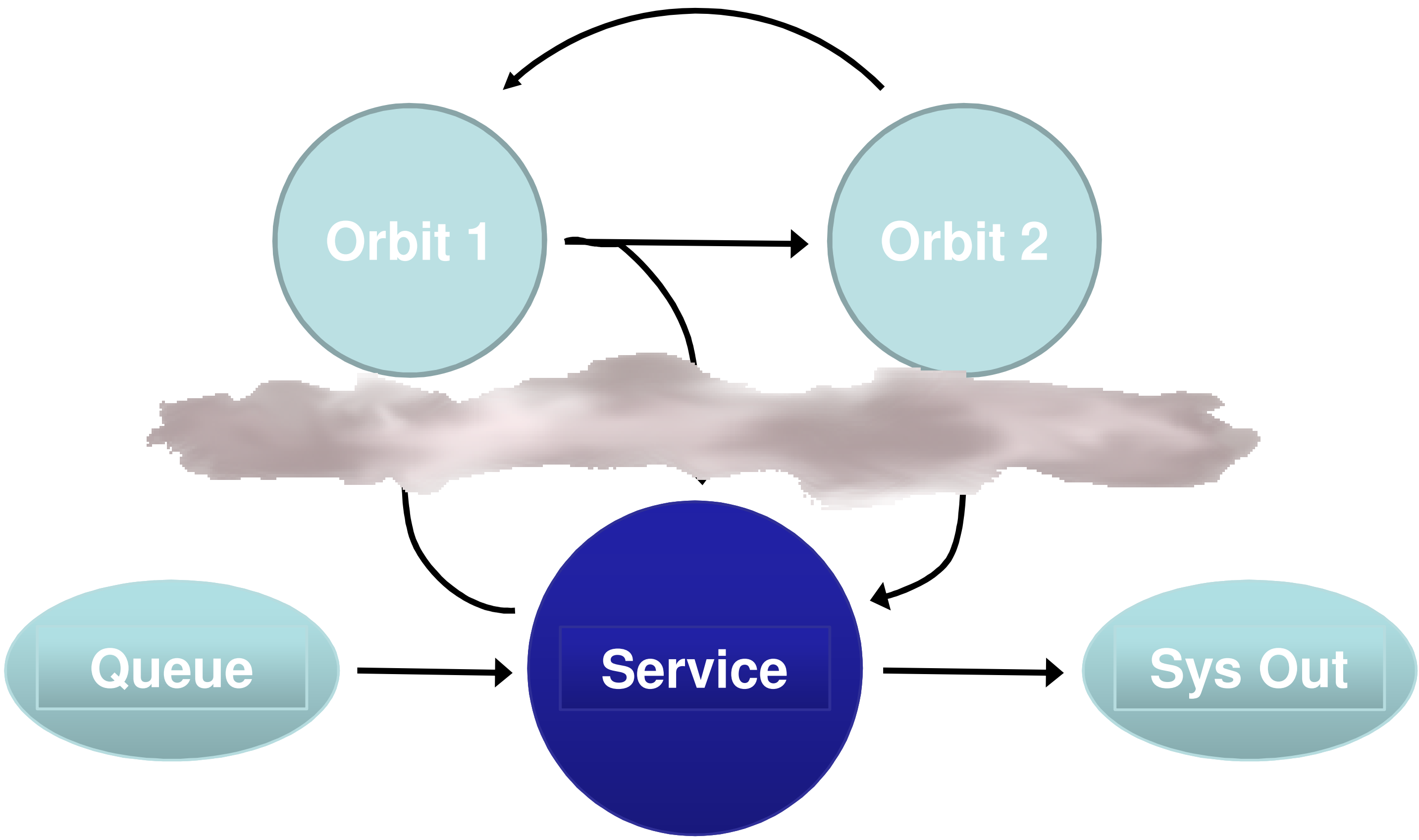
# % Diff: Average Number in Orbit



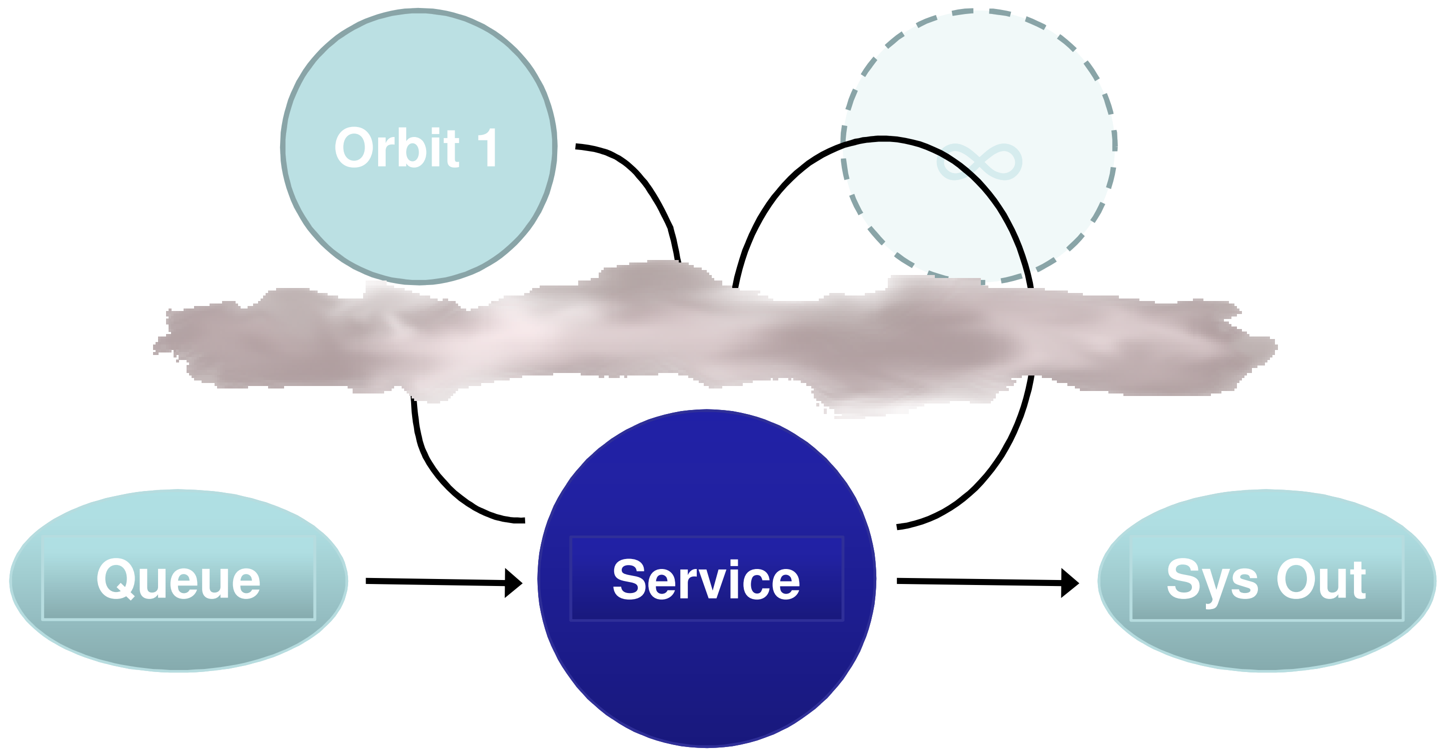
# Cox-Marie



# NP distribution



# $H_2$* distribution



# Very Low Retrial Rates, D-retrials

| RetryRate | Lo | StdErr | Expon. Lo | %diff from Exp | StdErr of % |
|---|---|---|---|---|---|
| 0.001 | 1931.987 | 3.332592 | 1414.9 | 36.54584 | 9.118938 |
| 0.01 | 194.7987 | 0.832853 | 142.39 | 36.80644 | 2.26279 |

| RetryRate | P(delay) | StdErr | Expon. Pd | %diff from Exp | StdErr of % |
|---|---|---|---|---|---|
| 0.001 | 0.149124 | 0.000172 | 0.13581 | 9.803055 | 0.001757 |
| 0.01 | 0.150076 | 0.000424 | 0.1362 | 10.1883 | 0.004161 |

# Very Low Retrial Rates, D-retrials

| RetryRate | Lo | StdErr | Expon. Lo | %diff from Exp | StdErr of % |
|---|---|---|---|---|---|
| 0.001 | 1932 | 3.3 | 1415 | 36.5 | 9.1 |
| 0.01 | 195 | 0.8 | 142 | 36.8 | 2.3 |

| RetryRate | P(delay) | StdErr | Expon. Pd | %diff from Exp | StdErr of % |
|---|---|---|---|---|---|
| 0.001 | 0.1491 | 0.0001 | 0.1358 | 9.8 | 0.002 |
| 0.01 | 0.1500 | 0.0004 | 0.1362 | 10.1 | 0.004 |