11294

ARPANET SATELLITE SYSTEM
ASS Note 12
August 22, 1972
See ASS Note 25 for corrections

ANALYTICAL RESULTS FOR THE ARPANET SATELLITE SYSTEM MODEL INCLUDING THE EFFECTS OF THE RETRANSMISSION DELAY DISTRIBUTION

L. Kleinrock and S. S. Lam

1. Introduction and Summary

The purpose of this note is to extend the analytic results on the ARPA Network Satellite System by L. Roberts in ASS Note 9. We shall define a more general model than those used in previous ASS notes [3,4,8,9] and arrive at less approximate solutions with the effects of the retransmission delay distribution included. These results when compared with previous (ASS notes) solutions indicate large disparities when the input traffic rate is low or when the random retransmission delay distribution has a small variance. Furthermore, the previous ASS notes results tend to be optimistic upper bounds of our more exact solutions.

The main results in this note are:

(1) The distinction between q, the probability of success for a new packet and q_t, the probability of success for a retransmitted packet. This builds into our model the effects of the distribution of retransmission delays upon the throughput rate and total packet delays. Expressions for channel efficiency and expected number of retransmissions are rederived.

(2) Assuming a uniform distribution (over k slots) for retransmission delays, an implicit equation (Eq. 5) for q_t is derived and solved numerically. A proof is given to show that in the limit as $k \to \infty$, q_t converges to $q = e^{-\gamma T}$. We also plot q_t, the channel throughput rate σ and expected packet delay as functions of k and the total traffic rate, γ.

In the limit as $k \to \infty$ the expected delay is found to increase linearly with k at a rate equal to $\frac{1 - q}{2q} T$ where T is the slot duration used; also as $k \to \infty$, the channel efficiency $\frac{\sigma}{\gamma}$ converges to $e^{-\gamma T}$ (where σ is the throughput rate). The optimal value of k for minimum packet delay plotted as a function of γ is found to be monotonically nonincreasing.

2. Preliminary Results

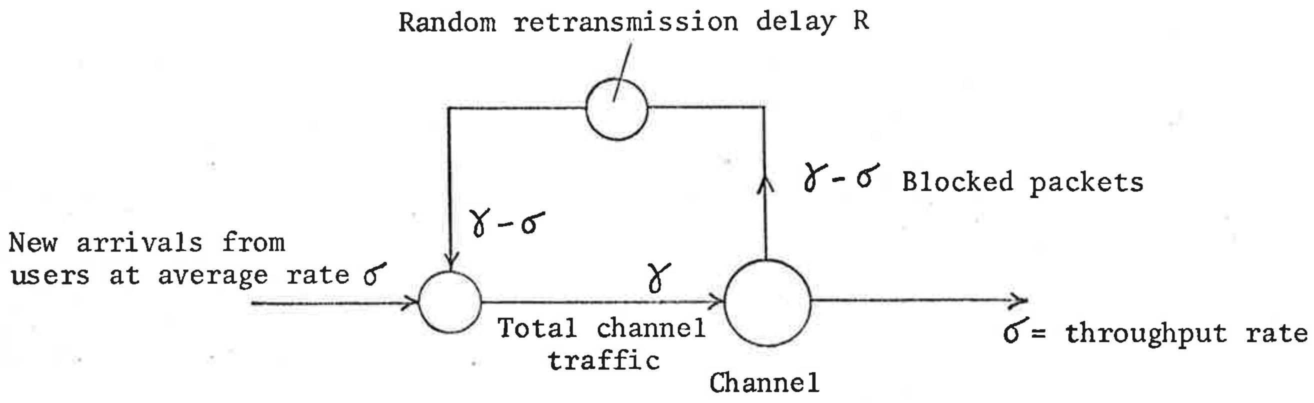


Fig. 1 The Model

Assumptions,

(A.1) satellite channel (non-capture system);

(A.2) synchronous transmission with slot size T;

(A.3) new arrivals form a Poisson process with parameter σ;

(A.4) total traffic arrivals form a Poisson process with parameter γ;

(A.5) throughput rate = rate of new arrivals under stationary conditions.

Define $p = 1 - q$

$= P$[a packet is blocked* in a slot/packet is new]

$p_t = 1 - q_t$

$= P$[a packet is blocked in a slot/previously blocked]

and we have immediately,

$$P[\text{no waiting}] = 1 - p = q$$

P[a packet has i retransmissions before success]

$$= pp_t^{i-1}q_t \tag{1}$$

E[total number of retransmissions for a packet]

$$= \sum_{i=1}^{\infty} i\, pp_t^{i-1}q_t$$

$$= \frac{p}{q_t} \tag{2}$$

The total channel traffic is therefore equal to the input rate σ times the total number of transmissions $(1 + \frac{p}{q_t})$:

$$\gamma = \sigma[1 + \frac{p}{q_t}] = \sigma \frac{q_t + p}{q_t}$$

or

$$\frac{\sigma}{\gamma} = \frac{q_t}{q_t + p} \tag{3}$$

---

*When a packet attempts a transmission which is unsuccessful due to other packet interference, we say it is "blocked".

which we shall call the efficiency of our satellite channel. The system throughput rate is

$$\sigma = \gamma \frac{q_t}{q_t + p} \tag{3a}$$

Define P = P[an arbitrary packet (regardless of whether it is a new packet or a retransmission) gets blocked in a slot]

$Q = 1 - P$

Hence P is the weighted average of p and p_t thusly,

$$P = p\frac{\sigma}{\gamma} + p_t \frac{\gamma - \sigma}{\gamma}$$

$$= p\frac{q_t}{q_t + p} + p_t \frac{p}{q_t + p}$$

$$= \frac{p}{q_t + p} \tag{4}$$

The above model is more general than models considered in previous ASS notes in that we distinguish between p and p_t.* This distinction is necessary to build into the model the effects of the distribution of retransmission delays upon total packet delay and channel throughput rate. L. Roberts recognized this point in ASS Note 9 when he defined q_t, but the distinction

---

*For the most general model we should let
$p_i = 1 - q_i$ = P[a packet gets blocked/packet has been blocked i times previously]
and then we have
P[a packet has i retransmissions before success] $= \prod_{\ell=0}^{i-1} p_\ell q_i$

(between p and p_t) was neglected as an approximation in his derivation of the expression for q_t. In earlier models, the effects of retransmission delays were usually ignored. No distinction between p and p_t was made and P has always been taken to be the same as $p = 1 - e^{-\gamma T}$ (assuming the total traffic arrivals to be a Poisson process with parameter γ). This approach simplifies the analysis and is not unjustified if the random retransmission delay distribution has a large variance. As will be shown in the next section, for a retransmission time with a uniform density (over k slots), p_t converges to p as $k \to \infty$.

3. An Implicit Equation for q_t and Some Limiting Results

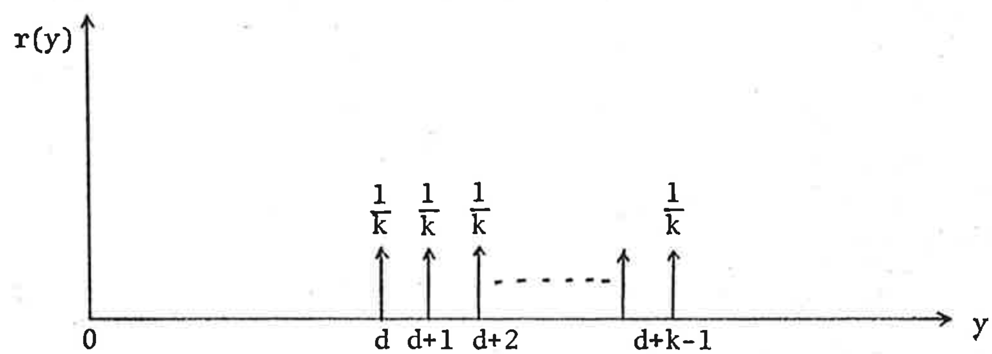


Fig. 2 Uniform Density Function of Retransmission Delay R (d is the "constant" propagation delay for a round-trip to satellite)

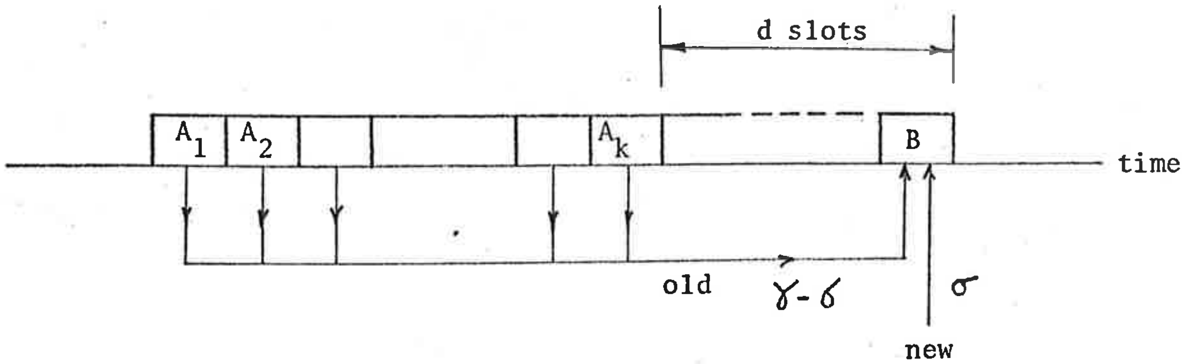


Fig. 3 Total Arrivals to a Specific Slot B

Assumptions,

(A.6) random retransmission delay R has a uniform density function as shown in Fig. 2;

(A.7) slot length $T = 1$ (at no loss of generality).

From our Poisson assumption for total traffic, we have, as shown earlier, that $p = 1 - e^{-\gamma}$. We shall find a more accurate expression for q_t than the one found in ASS Note 9. Let us condition on the event that a test packet was retransmitted from slot A_j to slot B as in Fig. 3.

Define q_i = P[no packets retransmitted from the i^{th} slot to slot B] $\quad 1 \leq i \leq k \quad i \neq j$

q_j = P[no packets (other than the test packet) is retransmitted from the j^{th} slot to slot B/blocking occurred on the last test packet transmission]

We have

q_t = P[successful transmission of test packet at slot B/ blocked previously]

$$= \prod_{\ell=1}^{k} q_\ell \; e^{-\sigma} = (q_i)^{k-1} q_j \; e^{-\sigma}$$

The calculation of q_i proceeds as follows:

$$q_i = \sum_{n=2}^{\infty}\left(\frac{k-1}{k}\right)^n \frac{\gamma^n}{n!}e^{-\gamma} + \sum_{n=0}^{1}\frac{\gamma^n}{n!}e^{-\gamma}$$

$$= e^{-\frac{\gamma}{k}} + \frac{\gamma}{k}e^{-\gamma} \qquad i = 1,\ldots,k \quad i \neq j$$

$$q_j = \frac{1}{P[\text{BLOCKED PREVIOUSLY}]}\sum_{n=1}^{\infty}\left(\frac{k-1}{k}\right)^n \frac{\gamma^n}{n!}e^{-\gamma}$$

$$= \frac{q_t + p}{p}\left[e^{-\frac{\gamma}{k}} - e^{-\gamma}\right]$$

since $P[\text{blocked previously}] = P = \dfrac{p}{q_t + p}$ from Eq. (4).

From Eq. (3) we have $\dfrac{\sigma}{\gamma} = \dfrac{q_t}{q_t + p}$ and so $e^{-\sigma} = e^{-\frac{q_t}{q_t+p}\gamma}$.

Finally then, we get

$$q_t = \frac{q_t + p}{p}\left[e^{-\frac{\gamma}{k}} + \frac{\gamma}{k}e^{-\gamma}\right]^{k-1}\left[e^{-\frac{\gamma}{k}} - e^{-\gamma}\right]e^{-\frac{q_t}{q_t+p}\gamma} \tag{5}$$

which is an implicit equation in q_t.

<u>Proposition 1</u> In the limit as $k \to \infty$, q_t converges to $q = e^{-\gamma}$.

<u>Proof</u>: From Eq. (5)

$$q_t = \frac{q_t + p}{p}\left[1 + \frac{\gamma}{k}e^{-\gamma+\frac{\gamma}{k}}\right]^{k-1}\left[e^{-\frac{\gamma}{k}}\right]^{k-1}\left[e^{-\frac{\gamma}{k}} - e^{-\gamma}\right]e^{-\frac{q_t}{q_t+p}\gamma}$$

Taking the limit as $k \to \infty$,

$$q_t = \frac{q_t + p}{p} e^{-\gamma(1-e^{-\gamma})} \left[1 - e^{-\gamma}\right] e^{-\frac{q_t}{q_t+p}\gamma}$$

We want to show that the equality holds for

$$q_t = q = e^{-\gamma} \quad \text{or} \quad p = 1 - q_t$$

$$\text{R.H.S.} = \frac{1}{1-q_t} \, q_t^{\,1-q_t} \, (1-q_t) \, q_t^{\,q_t}$$

$$= q_t = \text{L.H.S.} \qquad\qquad \text{Q.E.D.}$$

Proposition 2 In the limit as $k \to \infty$, the channel efficiency $\frac{\sigma}{\gamma} = \frac{q_t}{q_t + p}$ converges to $q = e^{-\gamma}$.

Proof: Obvious.

Expected Satellite Delay

Define D = E[total delay for a packet]

A = E[successful transmission delay]

R = E[retransmission delay]

From Eq. (2) we get,

$$D = A + \frac{p}{q_t} R$$

$A = C + T$ where C = propagation delay

T = slot length

From the density function $r(y)$ in Fig. 2 we have

$$R = d + \left(\frac{k - 1}{2}\right)T$$

$$= C + T + \frac{k - 1}{2}\,T$$

where d is assumed to be equal to A

$$= C + \frac{k + 1}{2}\,T$$

$$D = C + T + \frac{p}{q_t}\left[C + \frac{k + 1}{2}\,T\right] \tag{6}$$

<u>Proposition 3</u> In the limit as $k \to \infty$

$$\frac{\partial D}{\partial k} \text{ converges to } \frac{p}{2q}\,T = \frac{1 - e^{-\gamma T}}{2e^{-\gamma T}}\,T .$$

<u>Proof</u>: Obvious.

A program was written to solve for q_t in Eq. (5) iteratively and results for q_t, D, k_{opt} and the channel throughput rate σ are plotted in Figs. 4 to 10. In calculating D, we have used the constants $C = 0.27$ sec. and $T = 0.028$ sec. as in ASS Note 9; also T has been normalized to 1.

FIG. 4 PROB[SUCCESS/BLOCKED PREVIOUSLY] vs. k

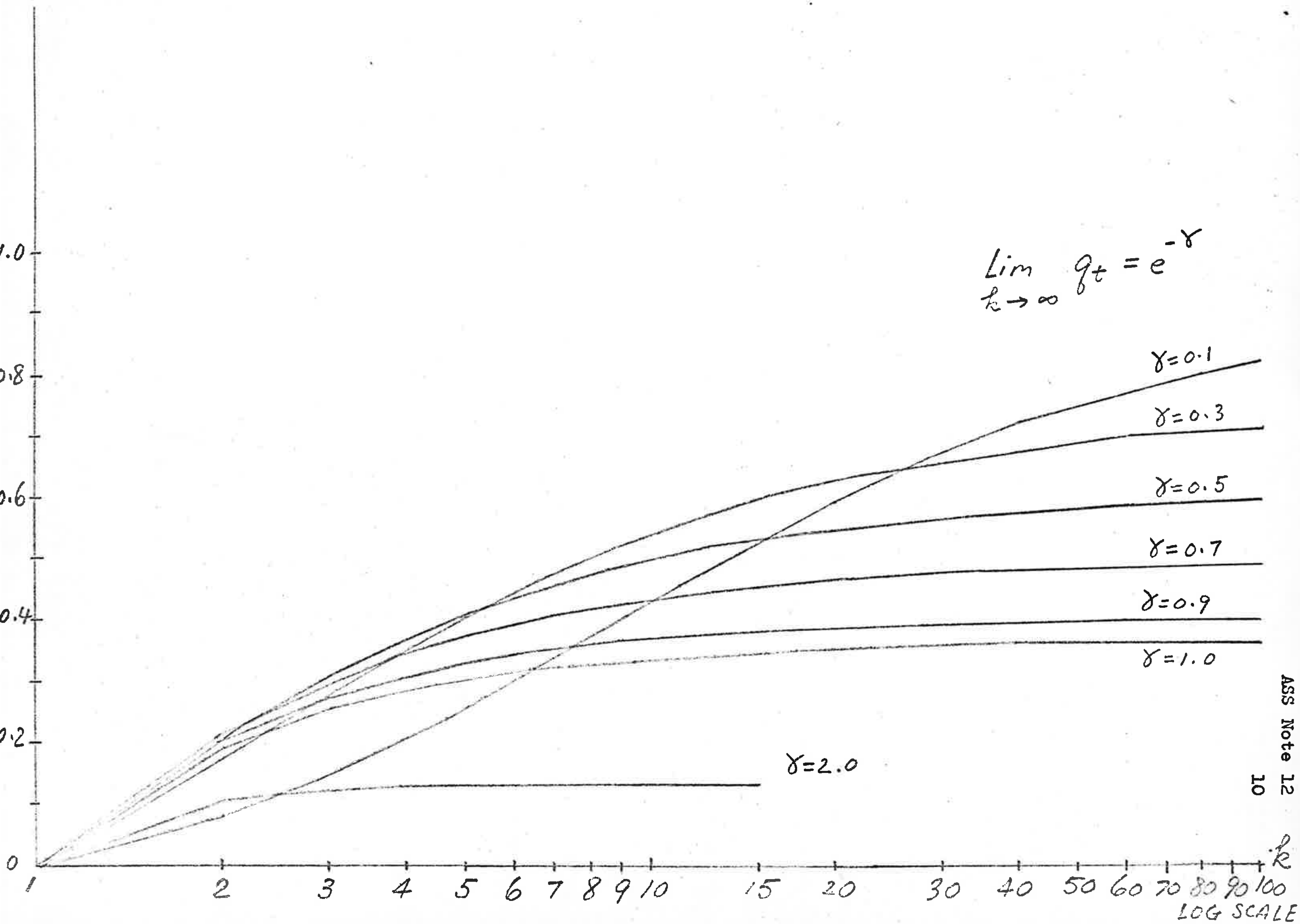


FIG. 5

Prob [SUCCESS | BLOCKED PREVIOUSLY] vs. k for $\gamma = 5.0$

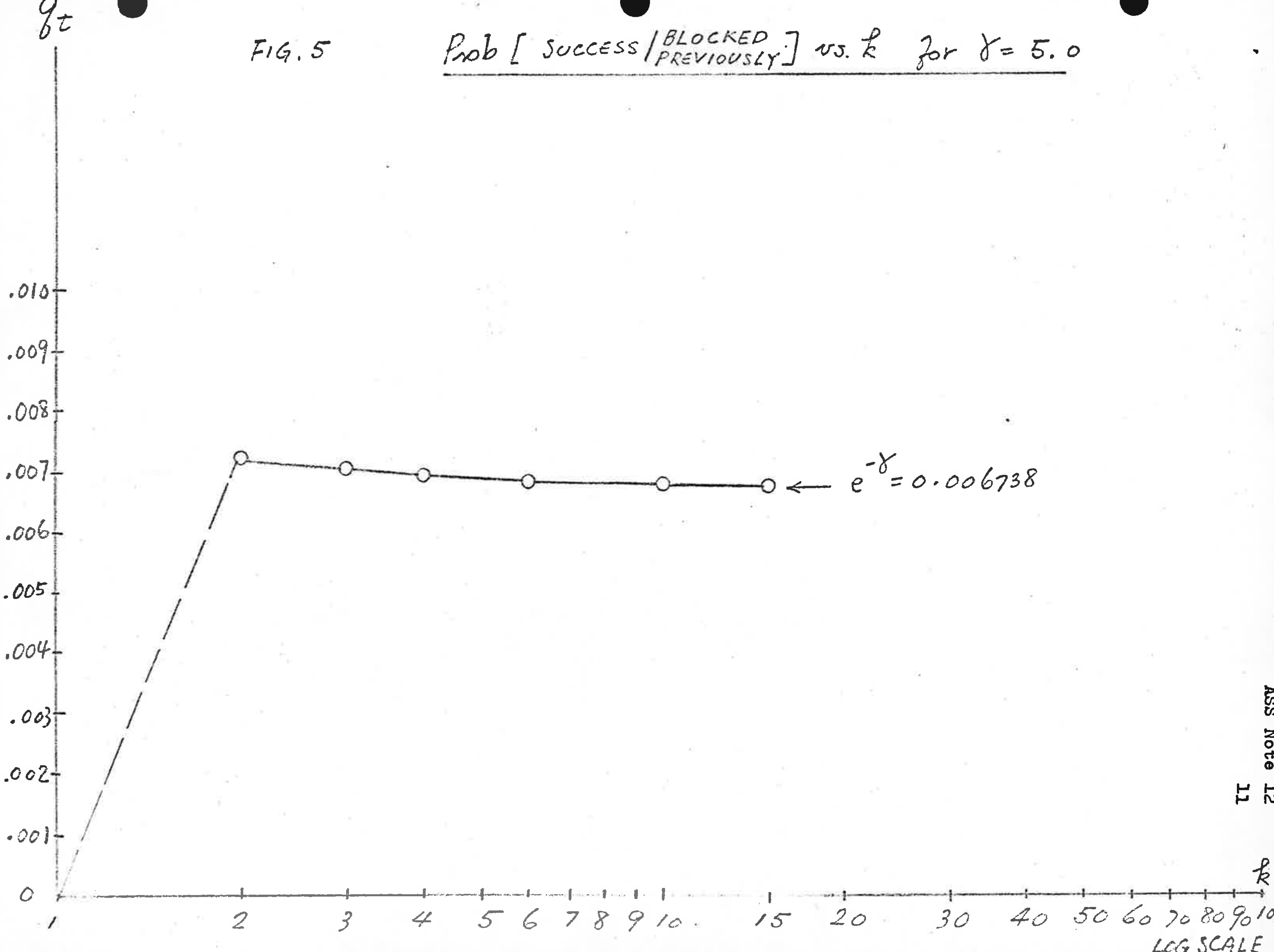


FIG. 6

PROB [SUCCESS / BLOCKED PREVIOUSLY] vs. k

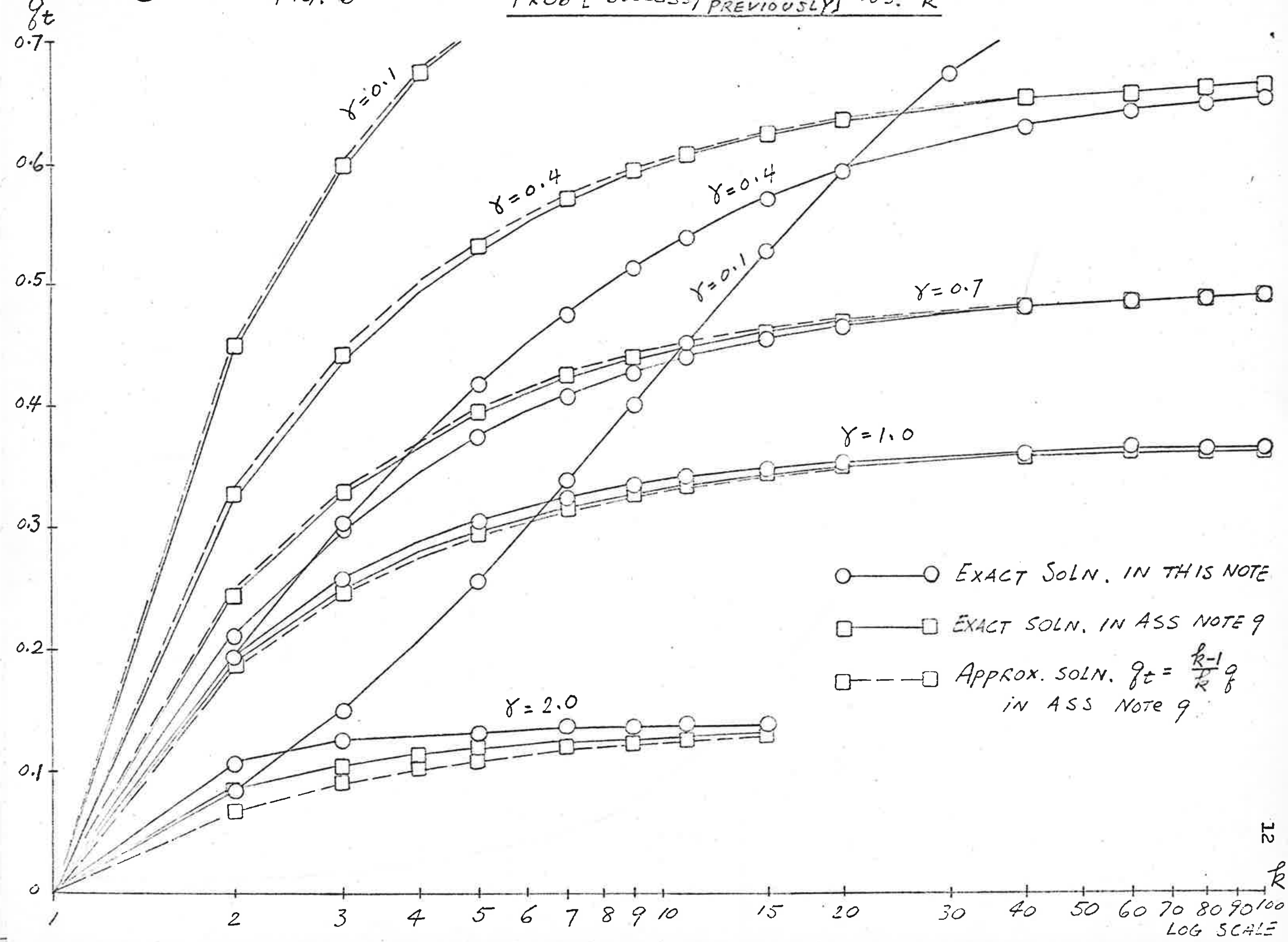


FIG. 7 THROUGHPUT RATE σ vs. γ

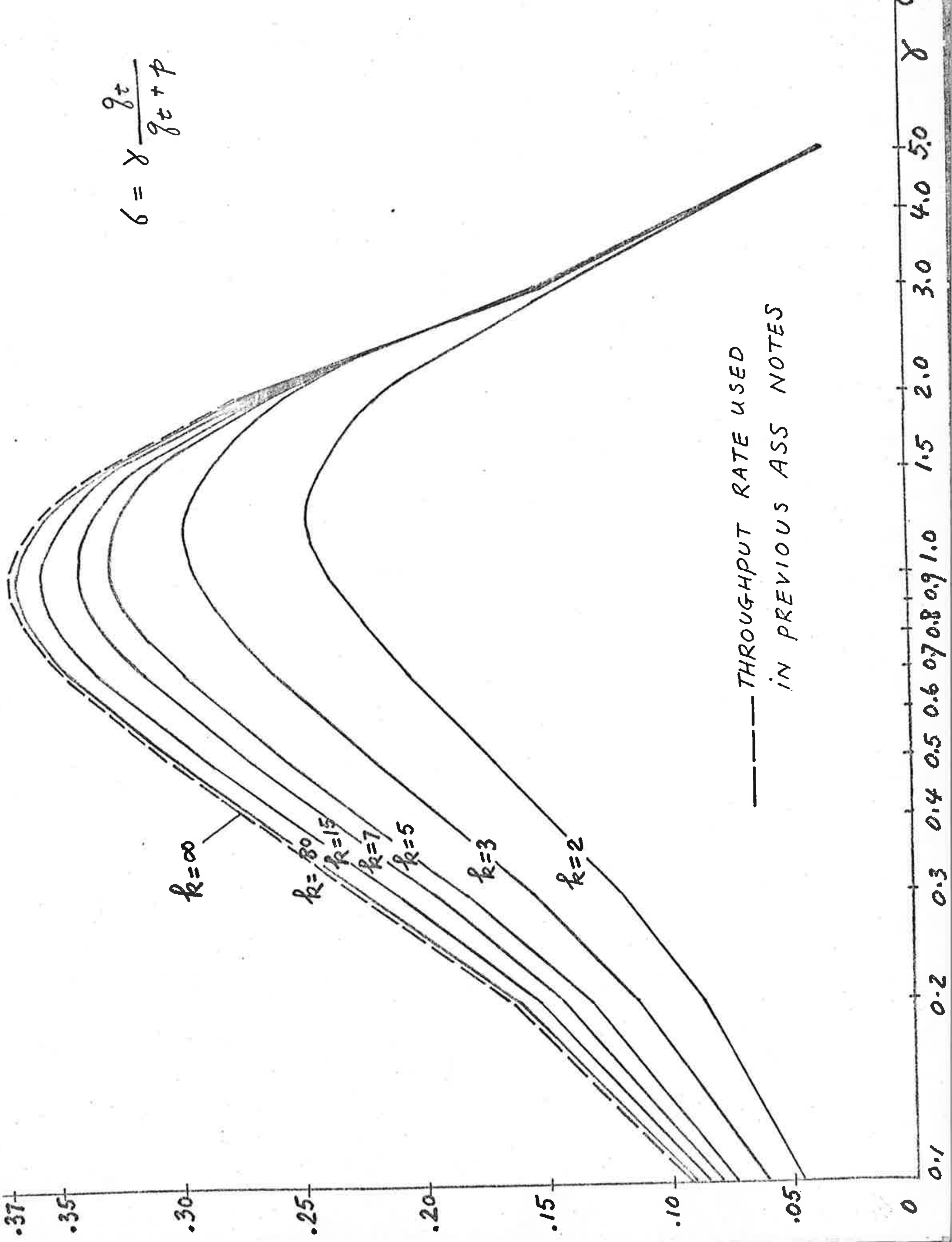


FIG. 8 THROUGHPUT RATE σ vs. k

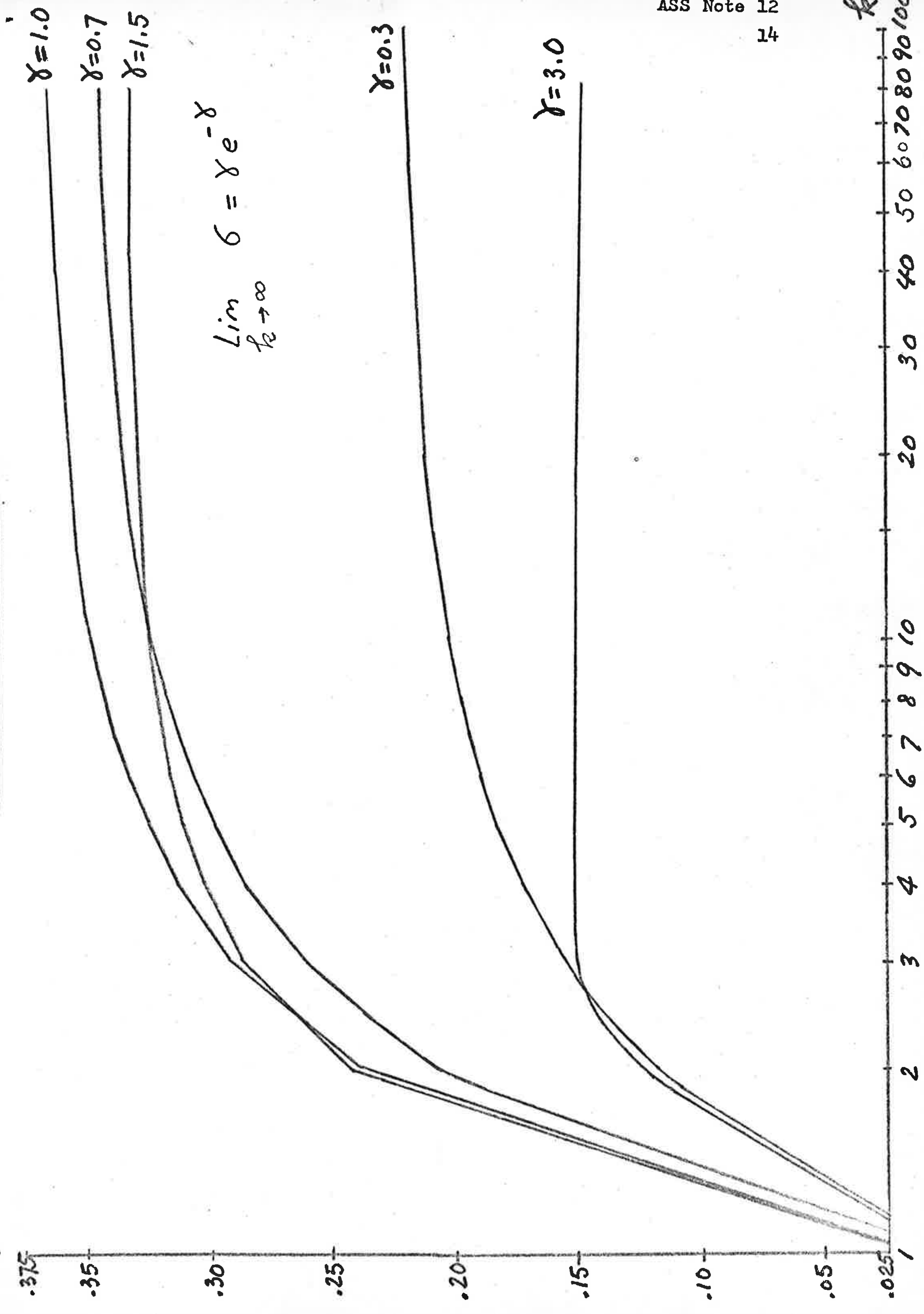


FIG. 9 EXPECTED PACKET DELAY vs k

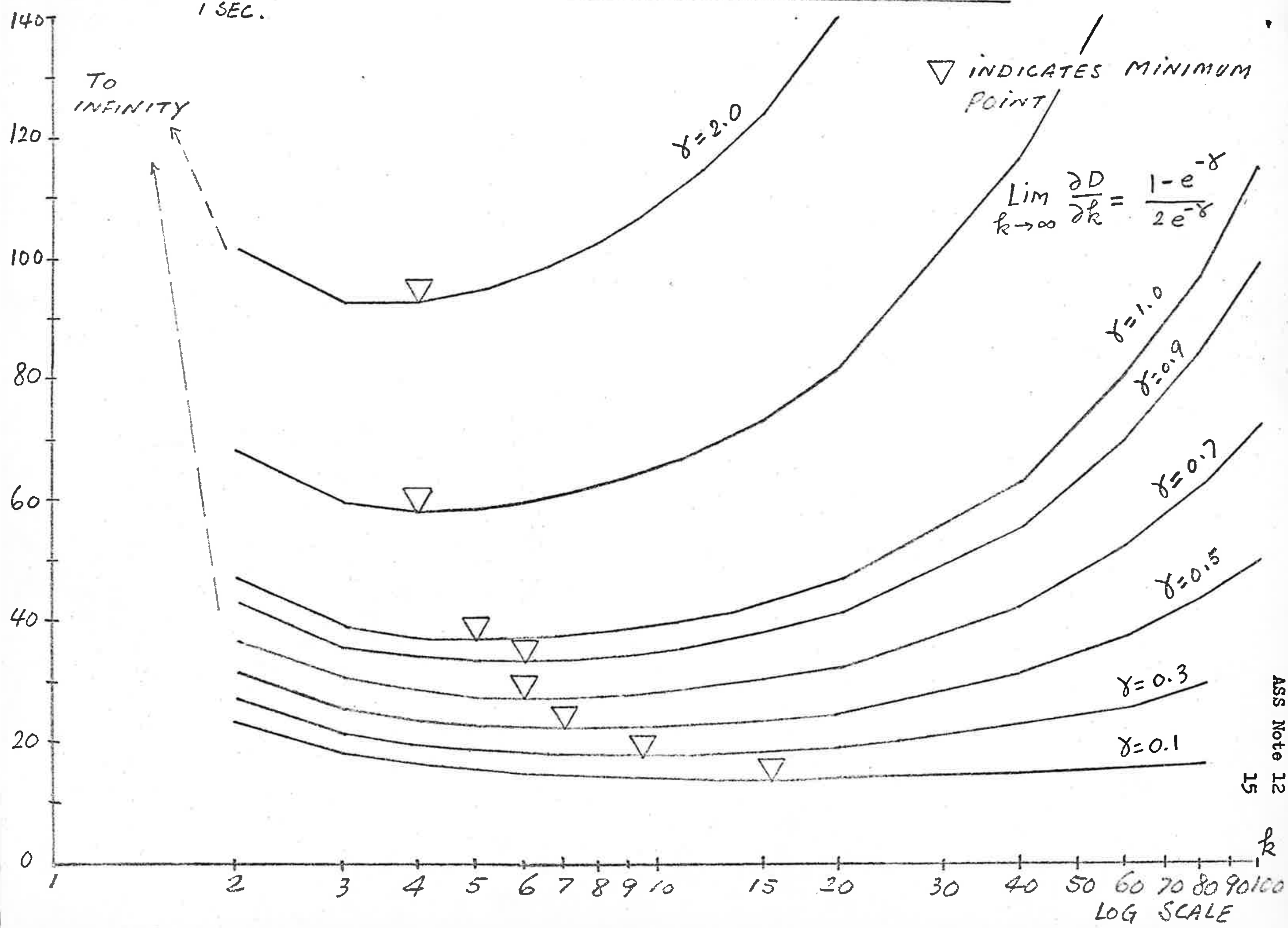
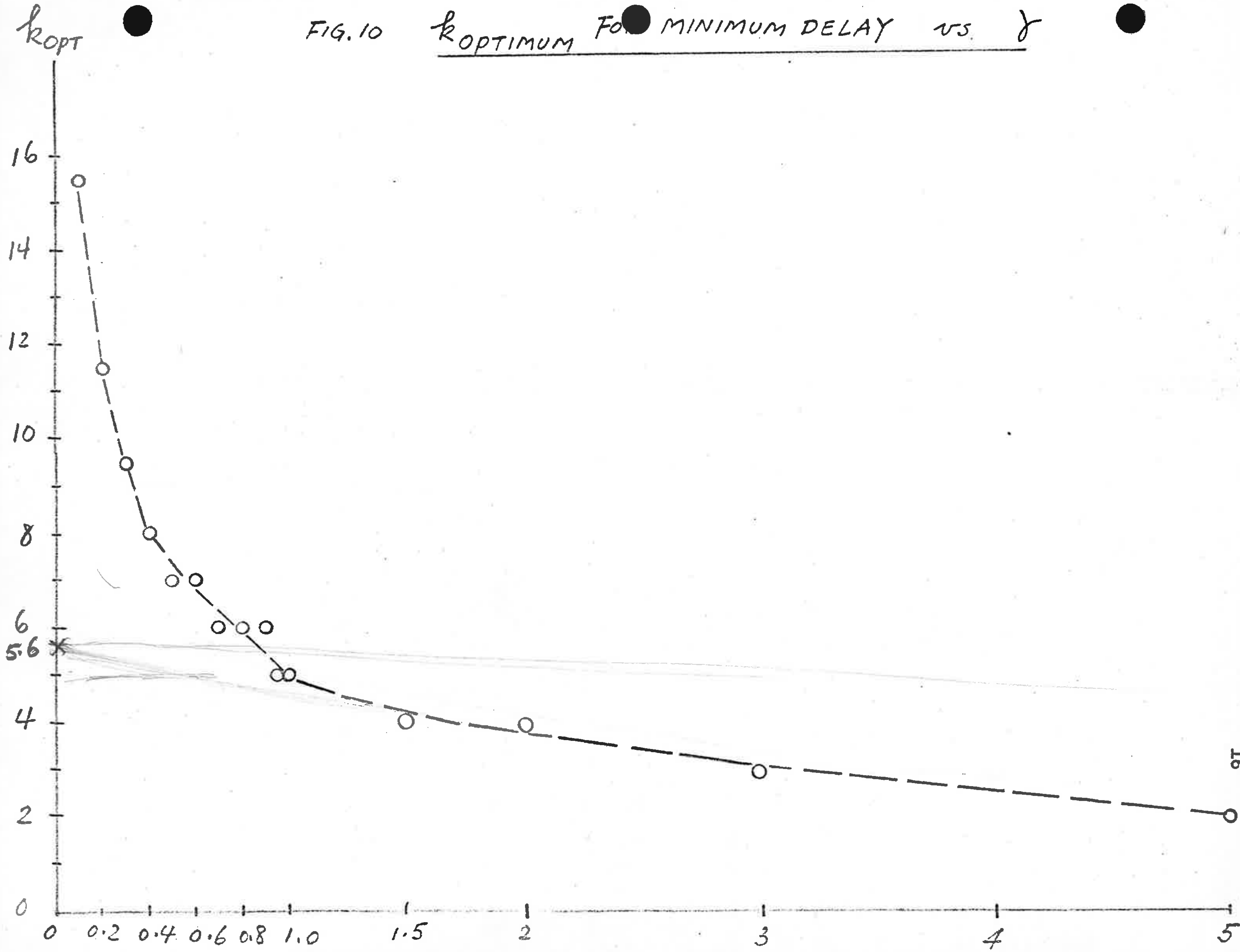


FIG. 10 $k_{OPTIMUM}$ FOR MINIMUM DELAY vs γ



4. Concluding Remarks

From Figs. 4, 5, 9 and 10 on P[success/previously blocked], expected packet delay and k_{opt}, we see that for large total traffic rates (γ), "small" retransmission delays (i.e. small k) are preferable. On the other hand, for small total traffic rates, "large" retransmission delays (i.e. large k) are optimal. This observation was not obvious before the analysis. The explanation is that under light traffic conditions, any blocking that occurs is usually "unlucky" and the congestion will most likely clear if the conflicted packets are retransmitted over a large number (k) of slots; whereas, under heavy traffic conditions, since congestion will not clear in a short time, retransmitting a packet after a long console wait will probably encounter more heavy traffic (in this case, it is better to retransmit as soon as possible to reduce the total packet delay due to console waits).

Our Eq. (5) is a more accurate expression for q_t than the one given in ASS Note 9. Comparing both solutions in Fig. 6, we see that large disparities exist for small values of γ (<0.6). Otherwise, the solutions are quite close, especially for large values of k.

From Fig. 7 we see that the throughput rate σ (as a function of γ) assumed in previous ASS notes is only an optimistic upper bound for the more exact values of σ as a function of γ and k. The reason for this is because in previous analyses Q = P[success for an arbitrary packet] has usually been taken to be the same as $q = e^{-\gamma T}$ = P[success/new packet]. From proposition 1 we see that q_t converges to q in the limit as $k \to \infty$, but is in general smaller than q. Thus, the previous analyses are only good in the limit as $k \to \infty$.

In conclusion, we feel that we have developed in this note a more accurate model for the ARPANET satellite system which takes into account the effects of retransmission delays with a uniform distribution. Similar analyses can be carried out using the same approach for other retransmission delay distributions deemed practical.