# BURSTY CELL STREAM CHARACTERIZATION

J. SOLÉ

J. DOMINGO

J. GARCÍA

RR 05/89

# BURSTY CELL STREAM CHARACTERIZATION

Josep Solé, Jordi Domingo, Jorge García

Universitat Politècnica de Catalunya
Departament d'Arquitectura de Computadors
Fax number: + 43 3 4017040

Facultat d'Informàtica de Barcelona
Pau Gargallo, 5 08028 BARCELONA (Spain)
Phone number: + 43 3 401 69 82

## Abstract

Traffic cell stream generated by a single source (or a single connection traffic cell stream) may be characterized by its mean traffic intensity (or normalized mean cell rate)$A_i$, but also by its burstiness.

"Burstiness" is an important aspect of IBCN's traffic to be considered, because it is a characteristic of most services in an ATM (Asynchronous Transfer Mode) environment [CHE88], [GEL89], [RID89]. Usually ATM cell streams alternate high activity periods (burst) and low or null activity periods (silences).

This paper is focused in characterizing the traffic cell streams from the burstiness point of view. In this sense, first we discus the conceptual aspect of burstiness and then we propose a mechanism to quantify the traffic cell stream burstiness. To validate this mechanism some examples are presented, and traffic cell streams of different burstiness are selected and its bursty characteristics commented. Finally, as an application, the behavior of a bus matrix switch element structure with each one of the two kinds of traffic selected is evaluated by simulation and the results are discused.

**Abstracte**

El flux de cel.les (paquets ATM, "Asynchronous Transfer Mode") d'una font qualsevol, o flux de cel.les d'una connexió, es pot caracteritzar per la seva intensitat de tràfic mitjana, $A_i$, (velocitat de transmissió de cel.les mitjana normalitzada); però també pel grau de fructuació, o grau de rafegueig, d'aquest flux.

El grau de rafegueig ("Burstiness") és un aspecte del tràfic de les xarxes IBCN ("Integrated Broadband Communications Network") que cal considerar, perquè és una característica de la majoria dels serveis en un entorn ATM; usualment, el flux de cel.les en ATM, alterna períodes d'alta activitat (ràfegues) amb períodes de baixa, o nul.la, activitat (silencis).

En aquest report s'hi tracta en la carectiratzació d'un determinat flux de cel.les des del punt de vista del grau de rafegueig, així, primer s'entra en discussió sobre el mateix concepte de grau de rafegueig i es proposa un mecanisme per quantificar-lo, acte seguit es presenten un parell d'exemples que il.lustren la utilització d'aquest mecanísme i, finalment, com a aplicació de l'estudi realitzat, se seleccionen dues seqüències de tràfic de grau de rafegueig força diferenciades i s'avalua, per simulació, el comportament d'un commutador "crossbar" amb cadescun d'aquests tràfics i se'n comenten els resultats.

## 1. Burstiness notion

The idea of burstiness of a traffic cell stream we are interested in is the one that is close related with the ATM concept. Later we will focus on the evaluation of its influence on the behavior of ATD (Asynchronous Time Division) switching elements.

First we will pay attention to the burstiness notion with no other considerations like the presence of the switch. There are several definitions for burstiness. One of them is the ratio between the peak cell rate ($A_p$ if normalizated) and the mean cell rate of a source ($A_i$) [KUL84]. Other definitions used are the ratio between the standard deviation of the cell rate (S) and the mean cell rate or the square of this ratio [COS88]. Another definition states the burstiness (B) as the mean density of bursts in time, that is, the percentage, over the total time, the burst is active [FIL89].

All of them try to give an idea of the same concept: the degree of regularity in the cell generation rhythm, irregularity implies burstiness. If we attempt to quantify the burstiness by the $A_p/A_i$ ratio we will have no information about how often this peaks of intensity have occurred. The standard deviation and mean traffic intensity ratios ($S/A_i$ or $S^2/A_i^2$) tell us the relative traffic intensity dispersion, but do not say anything about high density traffic concentration timing. In contrast, B gives this information in percentage of time, the fact is that all this are not enough to compare two traffic cell streams from the burstiness point of view –if two traffic cell streams with the same mean cell rate have $B1 < B2$, which one of them is the more burstiness?– at least the knowletge of the mean burst traffic intensities and the mean burst sizes seems to be necessary, and still in some particular cases they will not be enough.

Actually, the basic parameters that allow us to characterize a bursty traffic cell stream are the mean duration of the burst (L), the mean interburst periods (I), the mean traffic intensity for the burst period ($A_a$), the mean traffic intensity for the interburst period ($A_s$) and the mean percentage of bursts over the total time ($B = E[\mathbf{L}/\mathbf{L} + \mathbf{I}]$) –from now on we will refer a random variable with a bold face simbol, for instance **B**, and its mean value, E[**B**], with B–. Assuming that burst and low traffic intensity are independent of their lengths, certain combinations of these parameters, as for example: $R = A_aB + A_s(1\text{-}B)$, the distance between $A_a$ and $A_s$ or between $d_a = A_a\ L$ and $d_s = A_s\ I$, may quantify burstiness better than the ratios aforementioned. An other interesting proposal are the probability density function of the traffic intensity (per interval) weighted by the number of cells in the interval, that we will namely **D**, and the probability density function of **R**, (**R** = $\mathbf{A_a B} + \mathbf{A_s}(1-\mathbf{B})$) each one of them contain information of burst consistency and burst size together.

## 2. Burstiness measurement

Given an ATM cell stream we may know, without difficulty, something about its traffic intensity: the mean value, the maximum (or peak) value and, perhaps, speculate over the probability density function of the cell interarrival time. With this information some of the burstiness characterization mentioned may be used. If a more strict knowledge of burstiness is wanted (for instance the value R) we need to analyse accurately the complete cell stream and measure its basic parameters.

The essential requirement to determine the burstiness is to identify the high and low activity periods, as shown in figure 1. By analysing during a long enough time interval successive burst-silence periods several statistical measurements (i. e. probability density functions and mean and standard deviation values) can be obtained, namely, duration and traffic intensity of the burst, duration and traffic intensity of the silences, density of high activity periods over the total time, **B**, the traffic intensity of burts-silence couples, **R**, etc.

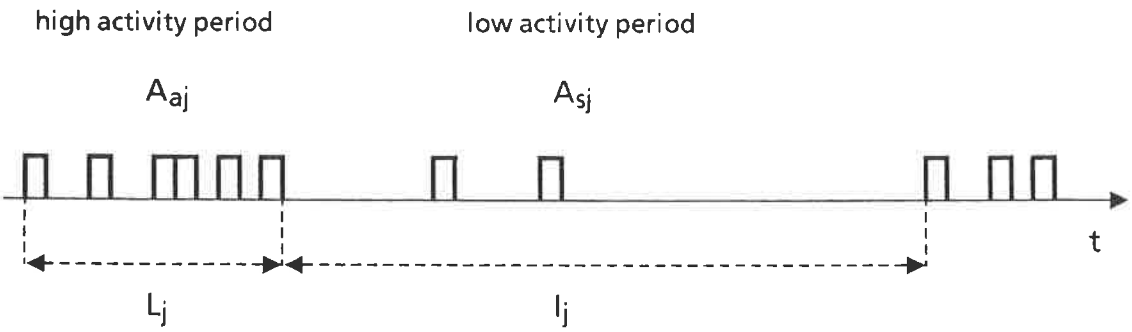


Figure 1. Bursty cell stream characterized by alternating high activity and low activity periods

To put the burstiness measurement in practice the partitioning criterion that we propose to adopt is based in that used in [JAI 86] to model the "train arrival process". A parameter MAIG (Maximum Allowed Intercar Gap) fixes the maximum intercell distance allowed to continue in a high activity period. So MAIG is used to decide when to close a high activity period (burst) and open a low activity one, and for the inverse case, to decide when to close a low activity period and to open a high activity one. In the later situation an additional condition needs to be applied to guarantee the burst entity: a high activity period must exceed a minimum length to be taken as a burst; otherwise, for instance, two consecutive cells in the middle of a silence would be considered as a burst and would distort the results.

**3. Examples**

In our implementation of the burstiness measurement algorithm (see the appendix for details) we have normalized all the parameters to the cell duration to make them idependent of the cell structure and the channel bit rate. We have set MAIG to the mean intercell distance value ($MAIG = \lceil 1/A_i \rceil$), this option seems to be coherent in the sense that it has a logical support: we are considering "silence" each interval of time in which the traffic intensity is lower than $A_i$ and "burst" each interval of time in which the traffic intensity is greater or equal than $A_i$.

In the other hand, is not so easy to find a justification for the burst entity insurance criterion. We have taken into acount two possibilities: fixing the minimum number of slots for the burst or fixing the minimun number of cells in a burst. In the examples we now present we have choosen the first option, because is the less severe one, setting the minimum burst lenght to MAIG slots: a high activity period (the mean traffic intensity in this period exceeds the mean traffic cell stream intensity) is only considered a burst if it is longer than MAIG.

Naturally other considerations about the MAIG value and the minimum burst length criterion, and value, may be defined. We have left the discution of other choices for further studies.

We have programmed the algorithm to obtain the probability density functions and, the mean and standard deviation values of the most significant bursty parameters of some traffic cell stream input trace.

*EXAMPLE 1.*

The objective of the first example was to verify the measurement program goodness. We have analyzed a completely deterministic periodic sequence of burst–silence intervals, set to the following values:

$$L = 2000, I = 4000, A_a = 0.125, A_s = 0.0125, \text{so } A_i = 0.05$$

The table 1 gives the mean values and the standard deviation values obtained after an execution over a sequence of 10E6 slots long. This table shows that the values of the basic parameters (L, I, $A_a$ and $A_s$) are correctely counted, the other result in this table, B, also fits well with the fact that, in this case, $B = E[L/L + I] = L/L + I = 1/3$, that implies $R = A_i = 0.05$. The value of NCF corresponds to the number of couples burst-silence found, and is close enough to the expected one, that is the total sequence lenght divided by the sequence period lenght L + I (10E6/6000 = 1666).

| MAIG = 20 | Periodic sequence |
|---|---|
| $A_i$ | 0.049 |
| $A_a$ | 0.1249 |
| $A_s$ | 0.0125 |
| B | 0.3333 |
| L | 2000 |
| I | 4000 |
| NCF | 1663 |

Table 1.

The figures 2 shows the probability density function of **B**, **R** and **D** obtained in this execution, and also corresponds to the expected ones.

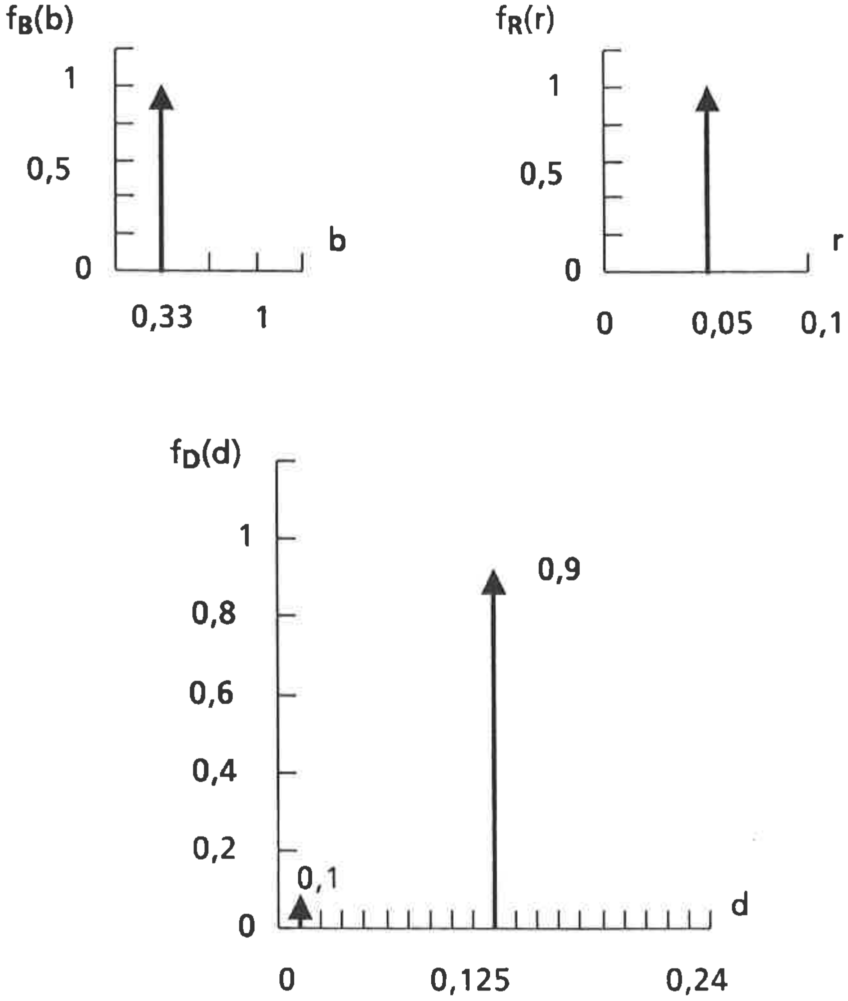


Figura 2.

*EXAMPLE 2.*

Real traffic traces would have been the most interesting to be tested, unfortunately we do not have them.

For the second example we have taken an analytic model to generate pseudoramdom sequences to treat them as real traffic traces. The model used was a three state Markov chain characterized by a probability matrix and, an average output traffic and a state mean sujourn time per state. The particular values we have set for these parameters are not relevant for our study, because the only thing we were interested in was to obtain a set of very different traffic cell stream with the same mean traffic intensity (although, the knowledge of the model allowed us to do a prediction, and the goodness of the results has been verified, again, in each particular case). So we have measured sequences generated by three different model configurations (α, β and γ) and two different mean traffic intensities $A_i = 4.8E\text{-}3$ and $A_i = 1.5E\text{-}2$.

The obtained mean values on measuring a 10E6 slots lenght sample of each one of the three kind of sequences, are collected in tables 2 and 3.

| MAIG = 209 | α | β | γ |
|---|---|---|---|
| $A_i$ | 4.87E-3 | 4.91E-3 | 4.9E-3 |
| $A_a$ | 1.69E-2 | 1.33E-2 | 1.63E-2 |
| $A_s$ | 7.94E-4 | 1.62E-3 | 1.03E-3 |
| B | 0.52 | 0.32 | 0.5 |
| L | 1005 | 438 | 795 |
| I | 3300 | 1547 | 2500 |
| NCF | 2274 | 4854 | 2966 |

Table 2. Results of the $A_i = 4.8E\text{-}3$ sequences measurements

| MAIG = 67 | α | β | γ |
|---|---|---|---|
| $A_i$ | 1.47E-2 | 1.49E-2 | 1.5E-2 |
| $A_a$ | 5.39E-2 | 4.3E-2 | 5.13E-2 |
| $A_s$ | 3.32E-3 | 5.95E-3 | 4.05E-3 |
| B | 0.54 | 0.33 | 0.53 |
| L | 320 | 122 | 240 |
| I | 1100 | 475 | 780 |
| NCF | 6474 | 14829 | 9056 |

Table 3. Results of the $A_i = 1.2E\text{-}2$ sequences measurements

As we have said in section 1 (burstiness notion) it is interesting to pay attention not only to the value of B but also to the average value of the gaps between $\mathbf{A_a}$ and $\mathbf{A_s}$ ($A_a$-$A_s$), between the burst cells size $\mathbf{d_a}$ and the silence cells size $d_s$ ($LA_a$-$lA_s$), and the value of R. In this example we can observe that the measurements corresponding with the sequence α (for the two cases $A_i$ = 4.8E-3 and $A_i$ = 1.2E-2) give the highest values obtained (tables 4 and 5). That leads us to say that α and β are the most and the less bursty cell stream traces respectively.

| $A_i$ = 4.8E-3 | R | $A_a$-$A_s$ | $d_a$-$d_s$ |
|---|---|---|---|
| α | 0.0092 | 0.0162 | 13.8 |
| β | 0.0053 | 0.0117 | 3.3 |
| γ | 0.0086 | 0.0152 | 10.4 |

Table 4.

| $A_i$ = 1.5E-2 | R | $A_a$-$A_s$ | $d_a$-$d_s$ |
|---|---|---|---|
| α | 0.032 | 0.0506 | 13.6 |
| β | 0.018 | 0.0370 | 2.4 |
| γ | 0.029 | 0.0472 | 9.2 |

Table 5.

Figures 3, 4 and 5, show the density probability functions of **B**, **D** and **R** respectively, all of these figures present the different functions obtained for the three configuration models overlapped. Only the ones for the case $A_i$ = 1.5E-2 have been included. Observing these figures several details can be appreciated:

1. The **B** density probability functions shows that β is the sequence in which the percentage of time of burst over the total burst-silence duration is under 0.5 with a high probability, in contrast, α and γ have values over 0.5 with high probability. This consideration alone does not allow to state a significative conclusion because values of **B** may be associated either with hard or soft traffic intensity gaps between burst and silences, this is the factor that will complete the bursty characterization.

2. The **D** and **R** density probability functions give us information about the traffic intensity gap between bursts and silences; case of peaks of these functions are near of the mean traffic intensity value ($A_i$) means high probability soft gaps and vice versa.

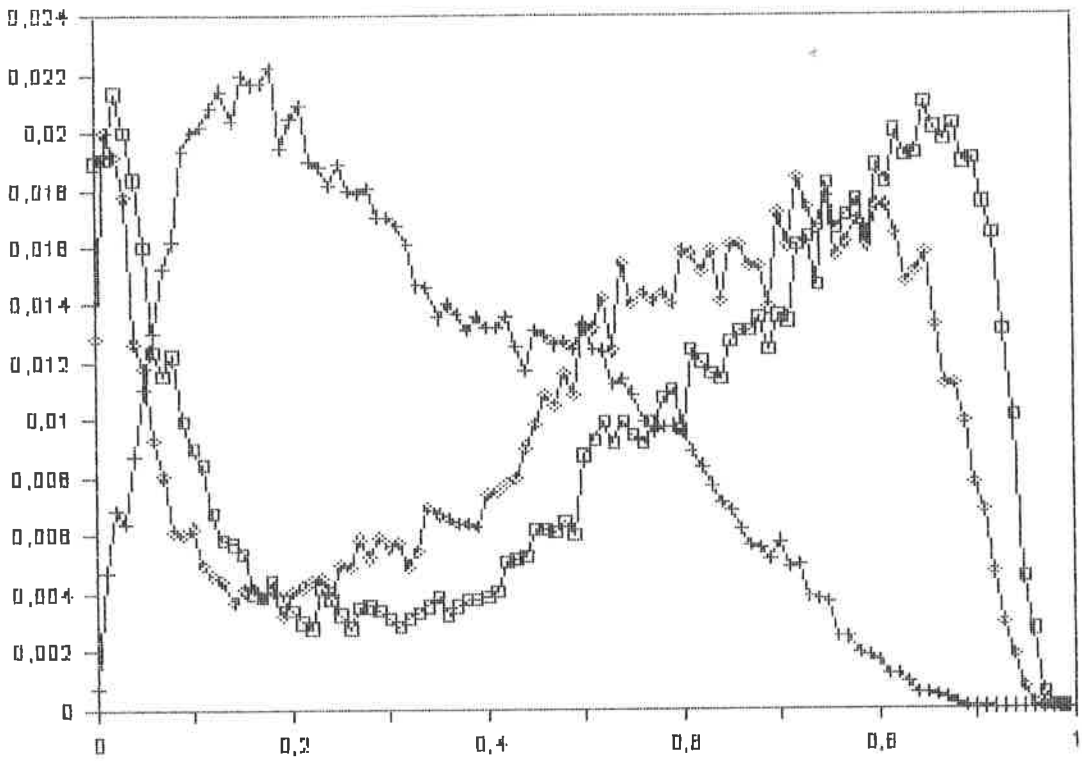


Figure 3. □ $f_B(b)$ of α; + $f_B(b)$ of β; ◊ $f_B(b)$ of γ

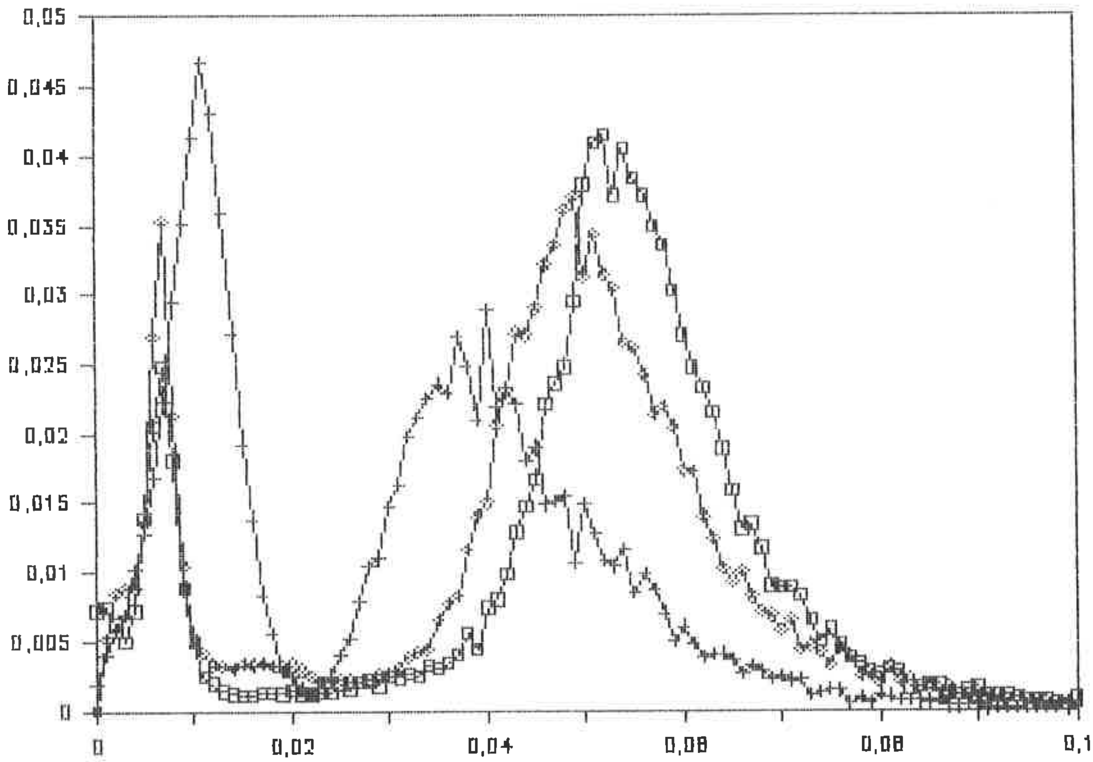


Figure 4. □ $f_D(d)$ of α; + $f_D(d)$ of β; ◊ $f_D(d)$ of γ

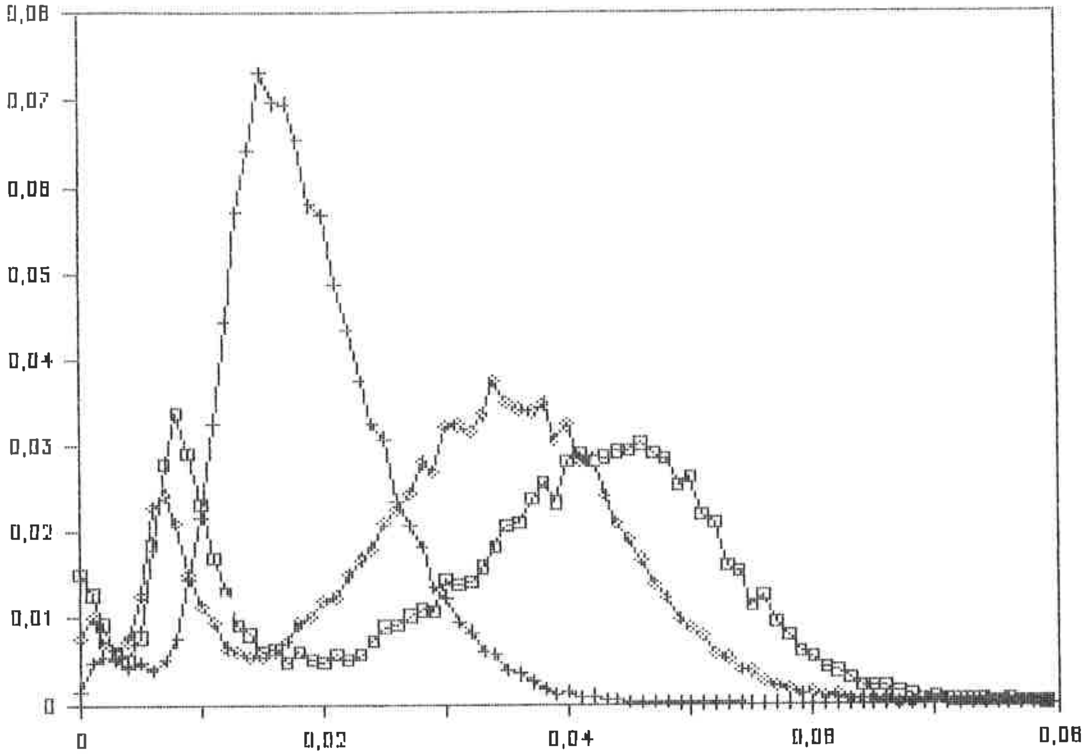


Figure 5. □ $f_R(r)$ of α; + $f_R(r)$ of β; ◊ $f_R(r)$ of γ

3. In **R** density probability functions may be see that, from burst-silence couples traffic intensity point of view, α and γ sequences have a great dispersion, while sequence β seems to be a chain of very similar burst-silence couples. In α and γ sequences very different traffic intensity burst-silence couples may be found, this is also a high burstiness indicator.

As we have mentioned above, after observing figures 3, 4 and 5,we can still mantein that α is the trace of the most bursty traffic cell stream.

Conflictive situations may arise when all these parameters or functions do not indicate the same direction, this is because is not easy to precisse estrictely the cell generation rhythm irregularity concept. At this point is when the external condition is needed; for instance one may consider that a negative influence over the switch behavior is also an indicator of burstiness.

## 4. Applications

The main objetive of our study was to prove that the burstiness plays an important role in the switch behavior. Oviously, mean input ports traffic intensity affects the switch behavior, but intuitively seems clear that burstiness of single connections (not of multiplexed connections) also will influence into the delay and cell lost values.

We have evaluated by simulatin a bus matrix switch structure behavior over two kinds of traffic cell streams: traffic cell streams of class $\alpha$ and traffic cell streams of class $\beta$. Recall that these twc sequences were the most burstiness distanciated in the example 2.

*SIMULATION ENVIRONMENT*

For this study we take a switching bloc of size 16 x 16, which is a crossbar with queues at each crosspoint (matrix of slotted buses, figure 6)

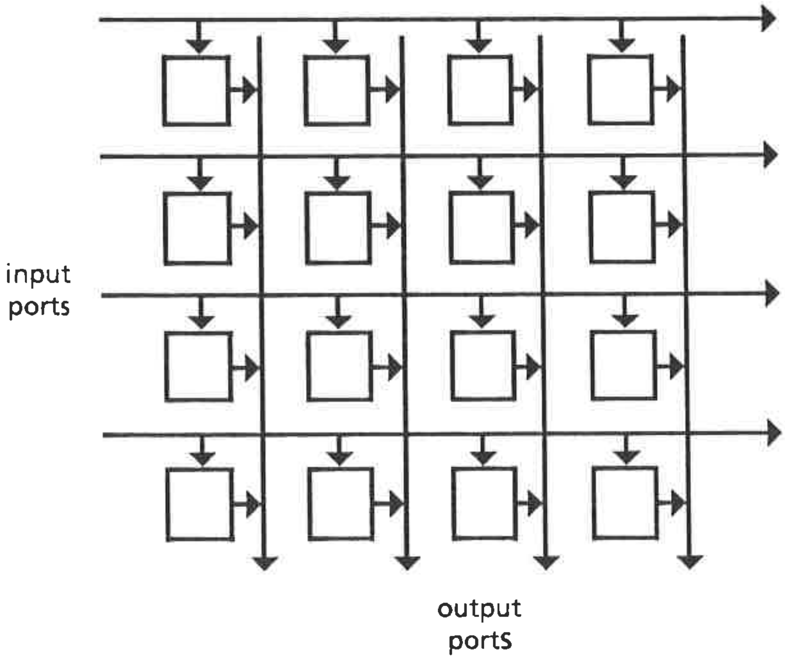


Figure 6. Crossbar switching element with queues at the crosspoints (Size 4 x 4)

[COS88]. The management of the queues corresponding to one output port (queue of the same column) is done as a global FIFO policy. That means that in case of more than one queue having a cell to output the oldest one is taken; in

Looking at the table 7 only, is interesting to observe that actually burstiness have an important influence in switch behavior, but if we compare the two tables we can see that this influence depends of the mean value traffic intensity.

Other applications of this measurement technic in study are:

1. The reproduction of some ATM traffic cell stream burstiness by a very simple source model (a two states one if it is possible).

2. The identification of the most significative traffic cell streams in an ATM environment to develop an ATM switcing architectures test–bench.

## References

[KUL84] J. J. Kultzer, W. A. Montgomery, "STATISTICAL SWITCHING ARCHITECTURES FOR FUTURE SERVICES". ISS 1984, 43.A1.

[JAI 86] R. Jain, S. A. Routhier, "PACKET TRAINS -MEASUREMENTS AND A NEW MODEL FOR COMPUTER NETWORK TRAFFIC". IEEE Journal on Selected Areas on Communications, Vol. SAC-4, no.6 (September, 1986) pp. 986-995.

[COS 88] Expert Group on ATM Switch Structures. "FINAL REPORT OF COST 202bis". December, 1988.

[CHE 88] T. M. Chen, D. G. Messerschmitt, "INTEGRATED VOICE/DATA SWITCHING". IEEE Communication Magazine, vol. 26, no. 6 (June, 1988), pp. 16-26.

[FIL 89] J. Filipiak, "M-ARCHITECTURE: A STRUCTURAL MODEL OF TRAFFIC MANAGEMENT AND CONTROL IN BROADBAND ISDN". IEEE Communication Magazine, vol. 27, no. 5 (May, 1989) pp. 25-31.

[GEC 89] J.Gechter, P. O'Reilly, "CONCEPTUAL ISSUES FOR ATM". IEEE Network (January, 1989) pp. 14-16.

[RID 89] M. J. Rider, "PROTOCOLS FOR ATM ACCESS NETWORKS". IEEE Network (January, 1989) pp. 17-22.

case of conflict, random selection is done. In this way the behavior of one of the queues may be generalized to the other ones.

Other parameters of the system are: the bitrate of the channel is 600 Mbps; the cell length is 48 information bytes plus 5 bytes for the header, that is 0.7 μsec per cell, as proposed recently in the last CCITT meeting; the mean traffic intensity per input channel is fixed to be 0.72 (channel load), this implies 48 multiplexed sources per input channel in case of $A_i = 1.5E\text{-}2$, and 150 in case of $A_i = 4.8E\text{-}3$ ($A_i$ still indicates the mean traffic intensity of a single source), for each one of the proves done (over α sequences and over β sequences). The unit of time is the cycle taken as the cell transmission time.

## *RESULTS*

The results obtained in this simulation are collected in tables 6 and 7. Table 6 contens the results for the case of lower traffic intensity sources ($A_i = 4.8E\text{-}3$) and in table 7 are included the results for the other case ($A_i = 1.5E\text{-}2$). All the values have been normalized to the cell transmission time..

| $A_i = 4.8E\text{-}3$ | QUEUE LENGHT mean and maximum values | QUEUE LENGHT standard deviation value | DELAY, mean and maximum values | DELAY, standard deviation value |
|---|---|---|---|---|
| α sources | 0.131 / 14 | 0.40 | 2.69 / 113 | 3.31 |
| β sources | 0.102 / 4 | 0.32 | 2.23 ./ 19 | 1.53 |

Table 6.

| $A_i = 1.5E\text{-}2$ | QUEUE LENGHT mean and maximum values | QUEUE LENGHT standard deviation value | DELAY, mean and maximum values | DELAY, standard deviation value |
|---|---|---|---|---|
| α sources | 0.303 / 120 | 2.87 | 6.65 / 1014 | 47.8 |
| β sources | 0.079 / 3 | 0.27 | 1.88. / 10 | 1.13 |

Table 7.

## Appendix: *ALGORITHM DESCRIPTION*

The basic idea of the algorithm in which the measurement mechanism is based is to count the interval lengths between two cells. The total number of slots –number of empty cells plus the last of the two full cells– of each of these intervals is assigned to a variable *Il* (Interval length), see figure A.1.

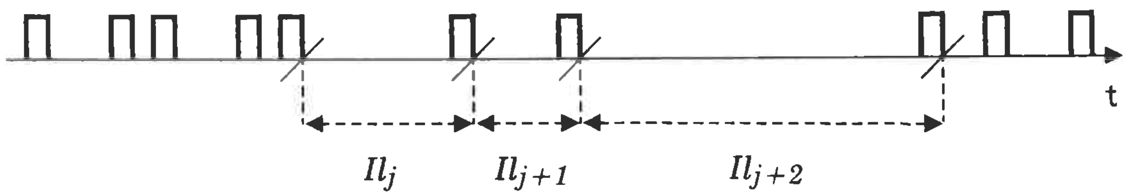


Figure A.1. *Il* values for the succesive inervals j, j + 1, j + 2 and so on

To mark the kind of interval we are studying (burst or silence) a boolean variable *b* (burst) is used. The variable *b* is initializated to a random value and after a transient running time *b* matches the reality. The PROCEDURE measure calls again and again the traffic source since a significative interval *Il* is identified, then the following conditions are tested:

```
If ((Il > = MAIG) and (b = TRUE)) then   (*We were in a burst of L slots long and d
                                           cells, and we have just entered in a silence
                                           period*)

begin
   do_sta(b,L,d);                        (*The statistics of the closed burst are
                                           done*)
   L: = 0;
   d: = 0;
   b: = NOT(b);
end;

If ( (Il < MAIG) and (b = FALSE)) then   (*One of the conditions to close a silence
                                           period and open a burst is verified*)

begin
   dp: = dp + 1;
   Lp: = Lp + Il;
```

```
    if (Lp > MAIG) then                 (*The initiated high density period is
                                        long enough to be a burst, the second
                                        condition is verified. This condition
                                        may be applied over dp and also
                                        another bound, different of MAIG, can
                                        be used*)

    begin
        d: = d-dp;
        Lp: = Lp-Il;
        L: = L-(Lp + 1);
        do_sta(b,L,d);                  (*The statistics of the precedent
                                        silence of length L and d cells are
                                        done*)

        L: = (Lp + 1);
        d: = dp;
        b: = NOT(b);
        Lp: = 0;
        dp: = 0;
    end;
end;

If ( (Il > = MAIG) and (b = FALSE)) then    (*A high activite period have not
                                            arrived to be a burst*)

begin
    Lp: = 0;
    dp: = 0;
end;

L: = L + Il;                            (*L acumulates the present interval
                                        lenght Il to make the posteior
                                        statistics*)
```

This program sees the traffic cell stream as an alternate sequence of high and low activity periods, where the end of one of them is defined by the beginning of the other (i.e. the end of a burst is marked by the beginning of the adjoining silence, when an interval $Il$>MAIG is found, and vice versa).

We do two kinds of analysis, one over each class of the periods (bursts or silences) and, another, over the burst-silence couples. With the first one we compose the statistics (probability density function and mean and standard deviation values) of the duration, the traffic intensity and the number of cells, in the burst and in the silences. With the second one we do the statistics of the percentage of burst (or silence, there will be complementary magnitudes) over the total burst-silence duration ,**B** (B = E[**B**]), and the statistics of the traffic intensity in the burst-silence couples, **R**, which mean we have namely R (R = E[**R**]).

In order to do that at each interval the variables L and d are updated to the values of total of slots and total of cells, figure A.2 shows how this is done. Lb and db stores the values of the prior burst while the present silence is catched to allow the burst-silence couple analysis.

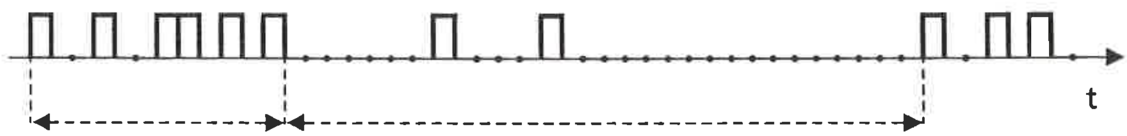


Figure A.2. Bursts and silences measurement

The PROCEDURE do_sta does the statistics for each kind of period and the statistics for the burst-silence couples: density probability function, mean value and standard deviation of the traffic intensity (**$A_a$** and **$A_s$**), of the interval duration (**L** and **I**) and of the number of cells in the interval (**$d_a$** and **$d_s$**), for bursts and silences, and, density probability function, mean value and standard deviation of the percentage of burst (**B**) duration and of traffic intensity for the couples (**R**).

The PROCEDURE do_sta also does the statistics of the traffic intensity counted not cell by cell but period by period, that we have namely **D**. This function is obtained weighted by the number of cells in the period (i.e. f**D**[dd]: = f**D**[dd] + d).

The transcription of the procedures where the density probability functions and, the mean and standard deviation values are calculated is not outstanding for the algorithm description.

It is necessary to be noted that the use of this algorithm implies to loose the indepencence between the burst and silences traffic intensity and their lenghts –the interval borders are identified by the value variations of the traffic intensity–.