(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 December 2001 (27.12.2001)

PCT

(10) International Publication Number
WO 01/99342 A2

(51) International Patent Classification⁷: H04L 12/00

(21) International Application Number: PCT/US01/19006

(22) International Filing Date: 13 June 2001 (13.06.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/596,609 19 June 2000 (19.06.2000) US

(71) Applicant: VITESSE SEMICONDUCTOR CORPORATION [US/US]; 741 Calle Plano, Camarillo, CA 93012 (US).

(72) Inventors: BARTON, David, F.; 715 Rocky Mountain Place, Longmont, CO 80501 (US). WORRELL, Kurt, J.; 5200 Glen Drive, Berthoud, CO 80513 (US).

(74) Agents: GALLENSON, Mavis, S. et al.; Ladas & Parry, 5670 Wilshire Boulevard, Suite 2100, Los Angeles, CA 90036-5679 (US).

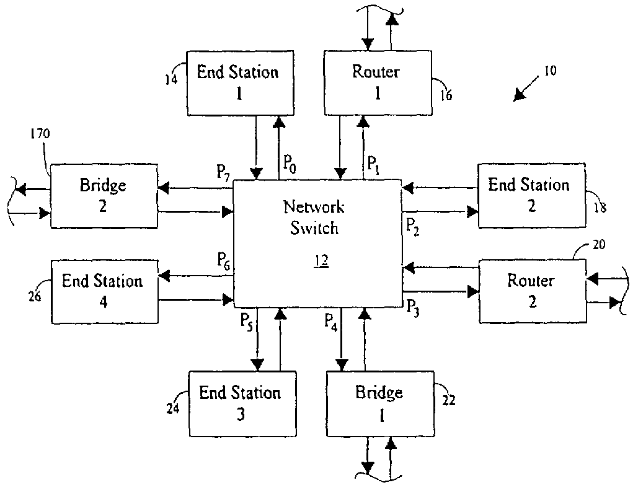
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

[Continued on next page]

(54) Title: DIGITAL NETWORK BANDWIDTH ALLOCATION METHOD AND APPARATUS



(57) Abstract: Advanced digital network bandwidth allocation is described. In one feature, flows require a minimum bandwidth allocation on the network link and are transferred using variable length cells multiplexed from different flows. In another feature, packets originating from data flows are divided into a series of cells such that each of the cells is identifiable as forming a part of specific packets. At least the specific packets are transferred on the network link incrementally by alternating each series of cells for the respective packets. At a second end of the network link, the identification information is used to reconstruct the series of cells for each specific packet to produce the specific packets at the second end of the network link. In still another feature, cells are formed using the data flows such that each data flow is made up of a series of one or more cells. At a time when the use of the network link is not allocatable to any of the data flows consistent with established priority allocations assigned to the data flows, at least one additional cell forming part of one of the flows is transferred, irrespective of the priority allocation assigned to that flow. Other features include response to variable network link bandwidth and burst priority re-assignment.

DIGITAL NETWORK BANDWIDTH ALLOCATION METHOD AND APPARATUS

BACKGROUND

The present disclosure relates generally to digital networks and, more particularly, to a method and system for providing advanced levels of service in a digital network while optmizing bandwidth utilization.

A number of approaches have been used in the prior art to cope with traffic on digital networks. Early digital network technologies did not suffer from their lack of sophistication in terms of bandwidth allocation since relatively little data required movement between interconnected components. For example, a network switch in such an early system might have employed a round robin allocation in which incoming data flows, requesting outgoing transfer via a particular network link, alternately received grants to use the particular network link. Even though such a simple allocation approach necessarily wastes a portion of the available network bandwidth, the approach was simple to implement while the waste was of little or no consequence in view of the surplus of available bandwidth.

As components interconnected by digital systems have greatly advanced in sophistication, there has been a corresponding remarkable increase not only in the amount of digital data requiring transfer between these network interconnected components, but in the nature of the way data is presented to the network. That is, for instance, some network data sources are characterized by bursting in which the sources require almost immediate transfer of a relatively large amount of data, followed by a quiescent period during which no data is presented to the network by that source.

Existing network technologies have necessarily undergone a continuous evolution to cope with such expanding sophistication. One of the more popular networking techniques, at the time of this writing, is known as Asynchronous Transfer Mode (hereinafter ATM). One example of improved ATM technology is described in U.S. patent number 5,515,363 issued to Ben-Nun, et al (hereinafter the '363 patent). While the latter discloses improvements which address traffic shaping to improve quality of service for virtual circuits, Applicants consider that the '363 patent, fails to address certain weaknesses in the overall ATM technology scheme, as will be described immediately hereinafter.

As in other state-of-the-art network technologies, ATM transfers packets containing data across the network incrementally in the form of cells. Packets vary in length depending upon the amount of data carried. For transfer, each packet is broken up into a series of cells. One weakness in ATM network technology resides in a reliance upon fixed cell length. Because there is no relationship between packet length and cell length, a particular packet will rarely divide in a way which completely fills the last cell of the plurality of cells. For this reason, some space in the last cell of the plurality is often wasted, since even the last cell must be of the required fixed length. It is recognized herein that reliance on fixed cell length results in wasted bandwidth which is now valuable.

Another weakness in ATM technology is submitted to relate to data flow bandwidth allocation. In ATM, flows are segregated into two general categories. The first category of flows is guaranteed bandwidth on the network, while the second category of flows is relegated to any remaining bandwidth, which is termed as "leftover" bandwidth. This scheme is problematic since flexibility in dealing with a flow, after assignment to one of the categories, may be limited. For example, it is difficult to allocate at least some of the leftover bandwidth to a flow in the guaranteed bandwidth category. As another example, flows being transferred in a particular system may simply be a mismatch as to these two narrowly defined categories. Moreover, it is submitted that ATM response to a flow which is guaranteed some predetermined bandwidth and which fails to use its guaranteed bandwidth wastes considerable bandwidth. Generally, if an ATM system responds at all to such a situation, the response is through the mechanism of recognizing over time that the flow is not using its allocated bandwidth and, thereafter, modifying the allocation given to that flow to bring the allocation more into line with the actual bandwidth used by the flow. Unfortunately, the bandwidth is wasted up until the response takes place.

Still another weakness in ATM technology is submitted to concern packet latency. That is, the cells which make up a particular packet are generally transferred in an ATM system ignoring the relationship of the cells to the packet from which they were formed. As will be described in detail at an appropriate point below, the present invention contemplates the use of cell position identification within a packet to remarkably improve packet latency.

Yet another weakness in ATM concerns a general provision that connected devices are given a physical transfer rate, equivalent to a fixed bandwidth allocation on the network, which is fixed at the time the devices are configured (start-up time). Following this initial configuration, changing the transfer rate on a network link requires ceasing all transfers over the network link, even those not involving a particular device of concern; reconfiguring the particular device with a different physical transfer rate; and then restarting transfers on the network link.

The present invention is submitted to address the foregoing in a highly advantageous and heretofore unseen way, as well as providing still further advantages to be described below.

SUMMARY OF THE INVENTION

As will be described in more detail hereinafter, there is disclosed herein an advanced digital network bandwidth allocation method and associated apparatus.

In accordance with one aspect of the present invention, a network arrangement is configured for transferring a plurality of data flows using a network link, each of which flows requires a minimum bandwidth allocation on the network link. The data flows are transferred by initially receiving the data flows at a first end of the network link. The data flows are then divided into cells such that at least some of the cells vary in length with respect to one another. Transfer of the cells on the network link is accomplished in a predetermined way such that cells of different data flows are multiplexed and the minimum bandwidth allocation of each data flow is maintained on the network link. At a second end of the network link, the cells are received, including the cells which vary in length, and combined so as to reconstruct the plurality of data flows at the second end of the link. In one feature, the data flows are divided into cells such that the length of each cell does not exceed a maximum cell length and at least a portion of the cells include a length that is less than the maximum cell length. In another feature, each data flow accumulates priority credits and is debited an amount of credit corresponding only to the length of a particular cell that is transferred. In still another feature, a plurality of different types of priority credits are simultaneously accumulated by at least one of the data flows such that priority is obtained based on having a predetermined combination of the plurality of different types of credits as to type of credits and/or amount of credits.

In another aspect of the present invention, a network arrangement is configured for transferring packets originating from a plurality of data flows using a network link. The packets are transferred by initially receiving the packets at a first end of the network link. For at least specific ones of the packets, each of which includes a length that is greater than a maximum cell length, the specific packets are divided into a series of cells such that each of the cells is identifiable as forming a part of that specific packet. At least the specific ones of the packets are transferred on the network link incrementally by alternating each series of cells for the respective packets. At a second end of the network link, the cells, as transferred from different packets, are received and the specific packets from which each cell originated are identified. Thereafter, that identification information is used to reconstruct the series of cells for each specific packet to produce the specific packets at the second end of the network link.

In still another feature of the present invention, a network arrangement is configured for transferring a plurality of data flows using a network link, each data flow requires a priority allocation of at least a minimum bandwidth on the network link. The data flows are transferred using the network link by initially receiving the data flows at a first end of the network link. Cells are formed using the data flows such that each data flow is made up of a series of one or more cells. Transfer of the data flows is initiated on the network link by transferring one or more cells on the network link which form portions of the flows consistent with the established priority allocation. At a time when the use of the network link is not allocatable to any of the data flows consistent with the established priority allocations of the data flows, at least one additional cell forming part of one of the flows is transferred irrespective of the priority allocation assigned to that flow. In one feature, certain ones of the cells are transmitted consistent with the established priority allocation for a particular flow along with at least one cell transmitted irrespective of the established priority allocation for the particular flow such that the certain cells include a cell configuration and the one cell transmitted irrespective of the established priority allocation for the particular flow includes a configuration that is identical to the cell configuration of the certain cells.

In yet another aspect of the present invention, a network arrangement is configured for using a network link having an available network link bandwidth which potentially varies over time and which is configured for incrementally transferring a plurality of data flows, for at least specific ones of which preferred minimum bandwidth allocations are specified. The data flows are transferred by receiving the data flows at a first end of the network link for transfer thereon to a second end of the network link. Based on the preferred minimum bandwidth allocations, sustained percentages of the available network link bandwidth are established to be assigned, respectively, to each of the specific data flows for use during incremental transfer of the specific data flows on the network link. During transfer and responsive to variation, over time, of the available network link bandwidth, actual use of the available network bandwidth by the specific data flows is controlled during the incremental transfer of the specific data flows across the network link such that each of the specific data flows is allocated approximately its respective sustained percentage of the available network bandwidth over time.

In still another aspect of the present invention, using a network link configured for transferring a plurality of data flows, each of which requires a minimum bandwidth allocation on the network link and which flows include bursts having assigned priorities, an arrangement and method are provided for transferring the data flows on the network link. Accordingly, a first configuration receives at least one burst as part of one of the flows at a first end of the network link with a particular assigned priority. A second configuration then transfers the burst to the second end of the network link with a new priority that is higher than the particular assigned priority. In one feature, the new priority is based on the use of at least two different types of priority allocation credits

BRIEF DESCRIPTION OF THE DRAWINGS

The present invention may be understood by reference to the following detailed description taken in conjunction with the drawings briefly described below.

FIGURE 1 is a diagrammatic representation of a network switch manufactured in accordance with the present invention.

FIGURE 2 is a more detailed illustration of the network switch of Figure 1 showing further details of its structure.

FIGURE 3 is a block diagram showing details of a single output port of the network switch of Figures 1 and 2 illustrating details of the structure of the output port.

FIGURE 3a is a diagrammatic representation of a data cell formed in accordance in the present invention and configured for transfer in accordance with the present invention.

FIGURE 4 is a table illustrating the flow configuration parameter implementation of the present invention which is used to fix flow parameter values in accordance with the sustained/burst priority concepts of the present invention.

FIGURE 5 is a diagrammatic representation, in block diagram form, showing one possible implementation of a bandwidth allocation state machine in accordance with the present invention for one of the flow paths forming a portion of the output port shown in Figure 3.

FIGURE 6 is a flow diagram illustrating the process carried forth in the bandwidth allocation state machine of Figure 5, illustrating one implementation of the sustained/burst priority concepts of the present invention.

FIGURE 7 is a block diagram illustrating an active packet priority (APP) structure for one of the flows of Figure 3 and the manner in which the structure produces APP information.

FIGURE 8 is a flow diagram illustrating one possible method by which the APP state determined by the APP structure of FIGURE 7 is calculated in accordance with the present invention.

FIGURE 9 is a block diagram illustrating one implementation of a bandwidth allocation arbiter produced in accordance with the present invention for use in the overall output port of Figure 3.

FIGURE 10 is a flow diagram illustrating one technique for updating the current selected cell register forming part of Figure 9 in accordance with the priority allocation concepts of the present invention.

FIGURE 11 is a flow diagram illustrating one process for updating the round robin arrangement show in Figure 9 in accordance with the present invention.

FIGURE 12 is a block diagram illustrating a priority encode comparator suitable for use in the bandwidth allocation arbiter of Figure 9 in accordance with the present invention.

FIGURE 13 is a diagrammatic representation of a first alternative encoding for the candidate and current cell registers of Figure 12 which promotes active packet priority above other priority types.

FIGURE 14 is a diagrammatic representation of a second alternative encoding for the candidate and current cell registers of Figure 12 which is implemented on a flow basis.

FIGURE 15 is a data flow diagram showing ingress data, queued data and egress data for three flows transmitted in accordance with the sustained priority concept of the present invention.

FIGURE 16 is a table illustrating the development of the data flow show in Figure 15, further including credit counts and pointer indications for the flows.

FIGURE 17 is another data flow diagram showing ingress data queued data and egress data for three flows and further illustrating the manner in which available unused bandwidth is utilized in accordance with the present invention.

FIGURE 18 is a table illustrating the development of the data flow shown in Figure 17 including credit counts and pointer indications for the flows.

FIGURE 19 is a data flow diagram shown here to illustrate the response of the present invention in a reduced network bandwidth situation.

FIGURE 20 is a data flow diagram shown here as part of a sustained credit counting example to illustrate the operation of the present invention in the instance of transmission of variable length multi-word cells.

FIGURE 21 is a table illustrating the development of the data flow shown in Figure 20 including credit counts and pointer indications for the flows.

FIGURE 22 is a plot of per flow sustained credit over time for each of the flows transferred in Figures 20 and 21, shown here to illustrated in detail changes in the credit counts for each flow in accordance with the present invention.

FIGURE 23 is a data flow diagram shown here as part of a burst fraction example to illustrate the operation of the present invention in the instance of transmission of data with a bursty flow and using multiple types of priority credits.

FIGURE 24 is a table illustrating the development of the data flow shown in Figure 23 including credit counts and pointer indications for the flows.

FIGURE 25 is a data flow diagram illustrative of a case which does not utilize the burst fraction, shown here for purposes of comparison of this result with the result of Figure 24 when one of the flows exhibits bursty ingress data.

FIGURE 26 is a table illustrating the development of the data flow shown in Figure 25 including credit counts and pointer indications for the flows.

FIGURE 27 is a data flow diagram shown here as part of an active packet priority (APP) example to illustrate the operation of the present invention in the instance of transmission of data using the APP concept of the present invention.

FIGURE 28 is a data flow diagram illustrative of a case which does not utilize active packet priority, shown here for purposes of comparison of this result with the result of Figure 24 to illustrate the advantages of APP.

FIGURE 29 is a table illustrating the development of the data flow shown in Figure 28, shown here to further illustrate priority allocation using the credit priority concept of the present invention.

DETAILED DESCRIPTION OF THE INVENTION

Turning to the drawings, wherein like components are designated by like reference numerals throughout the various figures, attention is initially directed to Figure 1 which illustrates a portion of a computer network manufactured in accordance with the present invention and generally indicated by the reference numeral 10. Network 10 includes an eight port full duplex switch 12. The latter is bi-directional and is interconnected with a number of other network devices, as will be described immediately hereinafter.

Still referring to Figure 1, the eight ports of network switch 12 are indicated as P_0-P_7. Network devices connected to the ports include an a first end station 14 on P_0, a first router 16 on P_1, a second end station 18 on P_2, a second router 20 on P_3, a first bridge 22 on P_4, a third end station 24 on P_5, a fourth end station 26 on P_6 and a second bridge 28 on P_7. The first through fourth end stations may comprise, for example, workstations or personal computers. Typically, the end stations are leaf nodes of the network while the routers or bridges are internal nodes of a graph and are connected to other network devices using other links that are only partially illustrated. It should be appreciated that an 8x8 network switch is used in the present example for illustrative purposes only and that the teachings herein are readily applied in the instance of network switches having virtually any port count. Network switch 12 interconnects each of the eight attached devices so that any one device may communicate through the switch to any of the other devices, including a device communicating with itself through the switch.

Referring to Figure 2, the internal structure of network switch 12 is illustrated in more detail, showing that each port includes an input port, PI_0-PI_7 and an output port PO_0-PO_7. The connections between the ports of the network switch and the connected devices may be referred to as network links or, more simply, links. For example, the link between port 4 and first bridge 22 may be referred to as link 4. Each input port is connectable to all eight output ports via an Input-to-Output port connection network 30. It should be appreciated that there are a variety of methods known to those of ordinary skill in the art for selectively connecting the input and output ports. All known techniques as well as any appropriate techniques yet to be developed are considered as being useful herein so long as each input port is selectively connectable to each output port.

Referring to Figure 3, details of one output port of network switch 12 are shown. In the present example, PO_7 is generally indicated and is shown diagrammatically. Eight port inputs P_0-P_7 are received from port connection network 30 (Figure 2) and are separated into any flows, F_1 to F_N that are present. Within network switch 12, flows are transferred in the form of packets each of which is made up of a series of cells. As will be described in more detail below, the present invention contemplates the use of variable length cells in contrast to prior art network technologies such as, for example, ATM which relies on fixed length cells.

Referring to Figure 3a, a cell, formed according to the present invention, is diagrammatically illustrated and generally indicated by the reference number 31. Cell 31 includes a header 32 and a data payload 33. Header 32 is made up of packet information 34, routing data 35, cell length data 36, a flow number 37, output link property data 38 and a channel ID 39. The latter is typically used to identify cells originating from a common packet in the reconstruction of the packets. Of course, data 33 is the object of transfer on the network. The specific amount of data contained in data payload 33 is specified by cell data length 33, in terms of octets. In contrast, ATM does not include cell size in its cell headers. Such information is not relevant in ATM since the cell size is fixed. Moreover, the ATM cell size is known, due to its fixed length, throughout an ATM network including at all input and output ports. The present invention contemplates that adaptability of existing ATM networks, in order to utilize the advantages disclosed herein associated with variable length cells, may be limited since reliance on fixed cell size is seen in at least some ATM hardware designs. That is, the fixed ATM cell size may simplify certain aspects of hardware. Other aspects of cell 31 of the present invention will be described at appropriate points below.

Referring again to Figure 3, flows F_0-F_7 are received by output port PO_7 following pre-processing performed, for example, by each flow's originating input port. In this pre-processing, the respective input port separates the header of each cell, containing control and routing information, from the data of the cell which need not be processed, but which data must be forwarded through the network switch. A cell header reception circuit 40 receives the cell headers for each flow while a cell data storage circuit 42 receives the cell data from each of the flows. It is noted that the cell size, forming part of the header information, is used by cell header reception circuit 40. Cell data storage circuit 42 may comprise, for example, commodity dynamic rams or any suitable storage configuration. As described above, the switching configuration for physically routing the cells between ports and for separating cell header information from cell data may take on a variety of configurations. Incoming data is stored by cell data storage section 42 for selective retrieval at appropriate times. Techniques for identifying the correct cell data, based on header information, are known in the art and readily adaptable to the present application by one having ordinary skill in the art. The cell header data arriving on one of the flow inputs to cell header reception circuit 40 is analyzed by the latter in a way which identifies the appropriate flow for each cell.

Each output port, including exemplary port PO_7, includes a flow cell header queue arrangement 50 (indicated by a dashed line) having a flow cell header queue corresponding to each flow. In the present example, only three cell header queues are shown indicated as flow 0 cell header queue 60, flow 1 cell header queue 61 and flow N-1 cell header queue 67. While the latter is specified as being associated with flow N-1, it should be appreciated that the highest order designated flow in the present example is flow 7. The use of flow N-1 emphasizes that the present invention is not restricted to any specified number of flows. After identification of the flow associated with cell headers, the cell headers are placed into the appropriate flow queue to await priority determinations requisite to transfer, as will be described at an appropriate point hereinafter.

Referring to Figure 4 in conjunction with Figure 3, a flow configuration parameter section, generally indicated by the reference number 70, stores priority allocation information associated with the various flows. Information provided initially by an external agent (not shown) is used in determining specific values for the parameters to be described. The specific parameters associated with each flow are illustrated. For purposes of brevity, only those parameters associated with flow 0 will be described, since all other flows include parameters having the same designations. Accordingly, F_0 has an F_0 bandwidth allocation (hereinafter BWA) sustained fraction 72, an F_0 BWA sustained limit 74, an F_0 BWA burst fraction 76 and an F_0 BWA burst limit 78. The BWA parameters for each flow are used by associated BWA state machines that are indicated by the reference numbers 80, 82 and 86 for flows F_0, F_1 and F_{N-1}, respectively. Each of the BWA state machines stores a BWA state value. Further associated with each flow are Active Packet Priority (hereinafter APP) computation elements. Specifically, each flow has an associated APP section, indicated as APP_0, APP_1 and APP_{N-1}, respectively.

Still referring to Figures 3 and 4, the BWA parameters in flow configuration parameter section 70 may be determined with the objective of maintaining flow control, for instance, based on any number of considerations including, but not limited to: 1. Ingress data rates for traffic destined to be transmitted on a particular output port/link (based upon the setting of sustained fraction 72). 2. The burst characteristics of a flow; i.e., is the flow a steady flow of N megabytes per second, uniformly distributed over time or is the flow bursty, needing low bandwidth over certain intervals and needing significant bandwidth over other intervals (based on the setting of burst fraction 76). 3. The amount of credit that can be accumulated by a flow without adversely affecting the other flows present on the same output link. That is, allowing a flow to accumulate credit to such an extent that one or more other flows are unable to attain their allocated bandwidths due to the disparity in priority (based on a combination of sustained and burst fraction settings). 4. Burst characteristics of upstream and/or downstream devices (based on the burst fraction). 5. Economic factors such as, for example, when an entity is willing to pay a surcharge to obtain a relatively larger guaranteed allocation on a network link.

Continuing with a description of the elements shown in Figure 3, the APP sections are interposed between their particular flow queues and a bandwidth allocation arbiter 90. These elements cooperate in a highly advantageous way for allocating network link bandwidth to the various flows, as will be described. For the moment, it is sufficient to note that lines 92, 94 and 98 extend from the flow queue headers through the APP components of each flow to the bandwidth allocation arbiter. These lines are configured for transferring priority control information as well as the actual cell headers, once selected for transmission. Using this configuration, data passed from the flow cell header queues is augmented with APP information in route to arbiter 90. The remaining structure in the figure includes a transmitter 100 interconnected to the bandwidth allocation arbiter via a transmit header queue 110. Once a cell is selected for transmission by bandwidth allocation arbiter 90, its corresponding cell header is placed or "released" into transmit header queue 110. Cell headers line up in the transmit header queue awaiting transfer by transmitter 100 on a physical network link 120. It is noted that the latter may be connected at an opposing end (not shown) to another network switch or, for example, to another network switch or, for instance, to an station or bridge (see Figure 1). If network link 120 is connected at its opposing end to another network switch, the network link would simply be connected to an input port, for example, such as one of the input ports shown in Figure 2. In this regard, it should be appreciated that one network link may carry a plurality of flows, each of which may be directed to a different destination. That is, each flow in Figure 3 does not represent a network link. Rather, the flows may be derived from any of the input ports shown in Figures 2 and 3, but all are directed to link 120, in this example, and thus PO_7, as represented in Figure 3.

Transmitter 100 is configured for sequentially transmitting cells when their associated cell headers are received, in turn, from the transmit header queue since the priority allocation method of the present invention has, at this point, been accomplished by preceding elements. Transmitter 100 includes a cell data fetcher section 122 which receives cell headers and then retrieves associated data from cell data storage arrangement 42 based on a known relationship between each cell header and the corresponding stored data. When a cell is chosen from a particular flow (e.g., flow 1) the cell header is dequeued from the flow queue (e.g., Flow 1 cell header queue 61) and is placed on the transmit header queue 110 for subsequent data retrieval and transmission. Select line 124 serves two purposes. First, providing an indication to the bandwidth allocation arbiter from the transmitter whenever a cell is transmitted. Second, in every instance that a word is transmitted, providing an indication to the BWA block of each flow such that the state of the flow associated with the transmitted word is properly updated. The select line is comprised of a combination of signals further including the flow ID of a transmitted cell and the priority with which a cell was transmitted. Accordingly, bandwidth allocation arbiter 90 then updates the BWA State for each flow. At times when output network physical link 120 is available, cells are sequentially transmitted as per the physical link protocol in first-in-first-out order from transmit header queue 110. As described above, the cell header is retrieved from the transmit header queue 110 by cell header fetcher 122 and the fetcher locates the data associated with the cell, as well as any output physical link parameters stored with the data. Examples of such output physical link parameters which may be stored with the data include but are not limited to physical link header information, physical link error detection/correction information or physical link timing information. If the input physical link and output physical link are identical physically, the input link physical data may be applied without change to the output physical link. Alternatively, physical network information may be modified or created by the network switch itself. Irrespective of the specific nature of data such as physical network information, the objective is to ensure that all information necessary to accomplish transmission of a particular cell is to hand using the combination of a cell header queue with referring to cell data storage section 42. The recombined cell is then transferred on network link 120.

Referring now to Figures 3 and 4, attention is now directed to the manner in which the present invention allocates outgoing network link 120 bandwidth to incoming flows F_0 to F_{N-1}. Each flow cell header queue promotes a candidate cell to bandwidth allocation arbiter 90. The candidate cell (not shown), represented by its corresponding cell header within each cell header queue, is first in line in the flow cell header queue. The values present in flow configuration parameter section 70 are used by the BWA State components in determining the transmission order of outgoing cells in attempting to meet desired outgoing bandwidth guarantees. As will be described below, the BWA sustained and burst bandwidth fractions do not actually specify a guaranteed bandwidth on the outgoing network link for certain reasons. Rather, the sustained and burst fractions represent a portion or percentage of outgoing available bandwidth present on the outgoing network link. In and by itself, the use of such parameters as representing a percentage of available bandwidth is considered as being highly advantageous and has not been seen heretofore by Applicants.

At least one reason this percentage allocation approach is so advantageous resides in the fact that some systems allow devices connected to a network link to control the clock rate used on the link. In this manner, a downstream device is able to conform the network link to its own capabilities. For example, the downstream device can instantaneously and selectively change the clock rate either up or down depending upon its own bandwidth requirements. The down stream device may even set the clock rate to zero which halts the link and then restart the link at will. As will be further described, during such a halt, the present invention preserves the priority state of each flow since priority credits are neither issued nor debited in the absence of link transmissions. At the same time, however, changing the clock rate in this way affects not only the flow being received by the controlling downstream device, but all of the flows transmitted while the downstream device is controlling this aspect of link operation. In effect, if the downstream device slows the clock rate, the available network link bandwidth can be "constricted" and can be quite limited while the rate is imposed. The percentage bandwidth allocation concept of the present invention responds in a highly advantageous way to this situation, as will be pointed out in further discussions.

With continuing reference to Figures 3 and 4, bandwidth allocation arbiter 90 utilizes flow configuration arrangement 70, the status of output link 120, the BWA State value and APP information for each flow in allocating priority for the use of the outgoing network link bandwidth. Further details regarding use of the APP information will be provided at an appropriate point below. Using F_0 as an example, the values relevant to F_0 are BWA F_0 sustained fraction 72, BWA F_0 sustained limit 74, BWA F_0 burst fraction 76 and BWA burst limit 78. While the manner in which these values are used in the overall method of the present invention will be described in detail hereinafter, it is noted for the moment that BWA F_0 sustained fraction 72 limits allocation of available network link bandwidth in an average sense over time. The BWA F_0 burst fraction is used in the instance of flows exhibiting "bursty" behavior to limit the percentage of available network bandwidth usurped by a flow during a particular burst.

Referring solely to Figure 4, flow configuration 70 is generally semi-static as supplied by the external agent, however, the flow parameters can be changed by the external agent at any time to provide new desired minimum bandwidth guarantees once the configuration change is complete. It is noted that the present invention implements its priority allocation scheme using credits accumulated by the flows and, thereafter, decremented in a highly advantageous way. In this regard, two different types of credits are allocated to each flow: sustained credits are allocated relative to the BWA sustained fraction and sustained limit parameters while burst credits are allocated relative to the BWA burst fraction and burst fraction parameters. Still considering the example of F_0 and again noting that each flow shares functionally similar entries, BWA F_0 sustained limit 74 limits the number of sustained word credits that can be accumulated by the flow while BWA F_0 burst limit 78 specifies the maximum number of burst word credits that can be accumulated by the flow. Either of the sustained or burst limits may be fixed values. While the allocation of two different types of credits to each flow is considered as being highly advantageous in and by itself, more detailed features of the present invention, concerning allocation and use of these different types of credits, are considered to provide still further advantages, as will be described.

Referring to Figure 5, a diagram of a BWA flow state machine is illustrated. Consistent with previous examples, the present example illustrates F_0 BWA state machine 80. All other BWA state machines are identical with the exception of storing parameter sets for different flows. Moreover, the BWA state machines for all of the flows operate in parallel. The flow 0 parameter configuration values are taken from flow configuration 70 (Figure 4) and used only in this state machine. BWA F_0 sustained fraction 72 and sustained limit 74 are used as inputs to an ALU_{SC} while BWA F_0 burst fraction 76 and burst limit 78 are used as inputs to an ALU_{BC}. The use of these parameters will be described more fully below.

Still referring to Figure 5, the reader is initially reminded of the present invention's highly advantageous use of variable length cells in combination with the advanced priority allocation scheme under description. In accomplishing the latter, a sustained priority register 130 stores a concatenated value in which SC (Sustained Credit) 132 indicates a number of maximum sized cell credits that have been accumulated for sustained transfers. It is noted that the term "sustained transfer" contemplates a transfer that could be continuous over a substantially long period of time. SC Word 134 indicates a fraction of one maximum sized cell credit, in terms of 16 bit words, accumulated for sustained transfers. SC 132 and SC Word 134 cooperatively indicate a total number of sustained word credits that are available to F_0. SC fraction 136 indicates a fractional part of a single word (a 16 bit word is used in the present example, but other word sizes may just as readily be employed) that has been accumulated for sustained transfers. Thus, SC 132, SC Word 134, and SC Fraction 136, concatenated in sustained priority register 130, cooperatively indicate a number of sustained fractional word credits that have been accumulated. It is important to understand that the register tracks not only a fraction of the maximum cell length, but a fraction of a single word from which the cells are made up. In the present example, SC fraction 136 is made up of 7 bits and is, therefore, capable of indicating fractional word amounts in units of 1/128th of a word. The significance of this capability will be made evident in continuing discussions. For the moment, however, it is noted that a fraction of 1 (128/128) is not permitted since this is equivalent to allocating the entire link to one flow.

Now considering burst elements of the Flow 0 state machine, a burst priority register 140 stores a concatenated value in which BC (Burst Credit) 142 indicates a number of maximum sized cell credits that have been accumulated for burst transfers. BC Word 144 indicates a number of fractional maximum sized cell burst credits that have been accumulated. BC 142 and BC word 144 together indicate a total number of burst word credits accumulated for use by F_0, noting that the units in BC 142 are maximum sized cells while the units in BC word 144 are words. BC fraction 146 indicates a fractional part of a single word that has been accumulated for burst transfers. BC 142, BC Word 144 and BC Fraction 146 together indicate a total number of burst fractional word credits that are available to F_0. Like the SC fraction, BC fraction 146 indicates a fraction of a word at a resolution of 1/128th of a word. ALU_{SC} performs all computations associated with sustained transfer credits while ALU_{BC} performs all arithmetic computations associated with burst transfer credits. Additional information is provided to these ALU's for use in the subject computations, as will be described at appropriate points below.

Based on sustained and burst credits, two determinations are made by the F_0 state machine. In a sustained priority determination, represented by a sustained priority comparator 150, a sustained priority threshold value is stored in an SC threshold register 152 for comparison with the combined value of SC 132 and SC word 134. The result of this comparison determines whether flow 0 has sustained priority. The sustained priority determination is used internal to the F_0 state machine as well as being supplied to other components on a line 154. The threshold sustained priority value in SC compare register 152 is typically the cell size, in words, for the candidate cell of F_0 that is being considered for transmission. It should be emphasized that this comparison establishes that F_0 has sufficient sustained credits based on the actual length of its candidate cell in words. Alternatively, one optimization for expediting computation is to verify that there are at least sufficient sustained credits to transfer a maximum sized cell, if such a maximum exists. That is, the present invention contemplates an implementation having no upper limit on cell size. In view of these alternatives, it is considered that a wide range of optimizations may be produced by one having ordinary skill in the art. In this regard, one of ordinary skill in the art may utilize any number of the bits present in sustained priority comparator 150 or in the burst priority comparator to be described. Normally, some selection of the upper order bits is used in the comparison. By adding a single lower order bit to the comparison, accuracy in the comparison is increased by a factor of one-half.

Turning now to the burst priority elements of the flow 0 state machine shown in Figure 5, a burst priority threshold value is stored in a BC threshold register 160. In a second, burst priority determination, represented by using a burst priority comparator 162, the burst priority threshold value is compared with the combined value of BC 142 and BC word 144. As described above with regard to the sustained priority comparison, the burst priority threshold may, for example, be the size in words of the F_0 candidate cell or the size in words of a maximum size cell, again, if such a maximum exists. Like the sustained priority determination, this burst priority determination is output on a line 164 for use internally by the state machine and is also supplied to other device components. Similarly, the burst priority comparison can be modified in any number of different ways using various combinations of the bits present in burst priority register 140 such as, for example, in optimizations intended to increase the accuracy of the comparison by adding more lower order bits to the comparison. Each state machine further uses indicators including: a "word transmitted flow k" received on a line 166 (derived from select line 124 in Figure 3), which indicates that a word has been transmitted in association with any of these flows and wherein the flow identification number is given by the letter "k"; word transmitted had SP (sustained priority) 168 (also derived from select line 124); and transmitted word event on a line 170, which indicates that a word is transmitted from any flow (also derived from select line 124). These indicators comprise inputs to the burst and sustained credit ALU's for use in determinations to be described below. Burst ALU_{BC} additionally uses a Flow Queue k Not Empty 172 indication in consideration of adjustment of the value in BC register 140, also to be described at an appropriate point.

Referring to Figures 5 and 6, attention is now directed to certain functional operations that occur within the BWA state machines of each flow, generally indicated by the reference number 200. More particularly, updates to sustained priority register 130 and burst priority register 140 will be described. Initially, it is noted that all BWA flow state machines operate in parallel such that the state for each flow is brought up to current on every clock cycle. Beginning at step 201, an indication arrives that a word has been transmitted for any flow via line 170 in Figure 5. In other words, the steps shown in Figure 6 are only performed responsive to transmission of a word. Following step 201, step 204 determines whether the word was for the flow associated with the particular BWA component and the whether sustained priority was associated with that transmitted word. These indications are received on lines 166 and 168 of Figure 5. If the transmitted word was from this flow and possessed sustained priority at the time it was selected, step 208 is performed to debit this flow's credits amounting to the cost of transmitting a word of data. The credits in sustained priority register 130 and burst priority register 140 are each debited by one word of credit. These values are not permitted to go negative. If the transmitted word was not from this flow on step 206 and/or did not have sustained priority, step 208 is not performed. In this regard, it should be appreciated that this process is performed on each clock cycle and that one word of data is transmitted on each clock cycle. It is for this reason that the flow is debited only for one word; the amount of bandwidth that was actually used. In the instance of transmitting a maximum size cell, the flow credits are decremented incrementally, as the words that make up the cell are transmitted. In the present example, a maximum size cell is made up of 64 words. By carrying out the method of the present invention in this manner, transfer of a cell having less than the maximum cell length results in decrementing the credits accumulated for its associated flow by no more than the number of credits corresponding to the length in words of that cell.

Referring to Figures 4-6, following step 208, computation continues at step 209 wherein it is determined if any additional cells are present in the flow queue for this flow (i.e., queue 80 in the example of F_0). Alternatively, step 209 is performed immediately after step 206 if the transmitted word was from a different flow. If there is at least one pending cell in the F_0 flow queue, step 210 adds BWA burst fraction 76 to the number of available credits in burst credit register 140. Step 210 is performed irrespective of the originating flow of the transmitted word. On the other hand, if the F_0 flow queue is empty, F_0 is not allocated additional burst credits and the method proceeds immediately to step 212. At this latter step, F_0 is credited by an amount equal to BWA sustained fraction 72 in sustained credit register 130. It should be appreciated that receiving the sustained credit allocation is based solely on the fact that a word was transmitted from any flow, as detected in step 202. That is, each flow receives the sustained credit fraction allocation, even the flow which transferred the word that initiated step 201. It should be mentioned that the sustained and burst fractions may be different for each flow, as is readily specifiable per the separate entries available for each flow in Figure 4. On this basis, flows may accumulate either sustained or burst credits at very different rates. Specifically, either the burst or sustained credit fraction may be specified in the range from 1/128 of a word up to 127/128 of a word, creating significant control in the rate at which credits accumulate (i.e., in increments of approximately 0.8 %), as will be described in further detail. It should be appreciated that additional bits can be added and relied on to further reduce the size of the smallest available fractional increment of a word, thereby allowing still further accuracy in specifying the allocation fractions disclosed herein.

At step 214, it is determined whether the burst temporary value is less than F_0 burst limit 78, shown in Figure 4. If not, the burst temporary value is set to the burst limit and the latter is stored in burst credit register 140 in step 216. If the burst temporary value is less than F_0 burst limit 78, the burst temporary value is stored in burst credit register 140 as the number of sustained credits available to F_0 in step 218. That is, the prior number of burst credits is incremented by F_0 BWA burst fraction 76. At step 220, it is determined whether the sustained temporary value is less than F_0 BWA sustained limit 74, shown in Figure 4. If not, the sustained temporary value is set to the sustained limit and the latter is stored in sustained credit register 130 by step 222. If the sustained temporary value is less than F_0 sustained limit 74, the sustained temporary value is stored by step 224 in sustained credit register 130 as the number of sustained credits available to F_0. With step 226, monitoring is returned to step 201. It should be appreciated that the foregoing discussion utilizes temporary values at least in part for purposes of clarity. A particular embodiment may or may not rely on the use of such temporary values, as desired.

In view of the foregoing, it is worth emphasizing that a flow is debited by a word's worth of credit for the word that is transferred from that flow on a particular clock cycle. Irrespective of which flow transmits a word, all flows receive their respective sustained and burst credit fractions as a result of transfer of the word, although a flow with an empty flow queue does not receive burst credit. In the instance of sustained or burst credits, the amount of credit that is allocatable to each flow on one clock cycle is specified as a fraction, since the allocation amount is a fraction of the maximum amount of credit (i.e., one word) that a flow can actually use on one clock cycle. In fact, the allocation fraction in the present implementation is specified in increments of 1/128th of a word. Therefore, the effect of the sustained credit allocation is to permit an over time average allocation to a flow while the burst credit allocation serves to limit or control short term dominance of a network link by a bursty flow. Examples of flows which exhibit various behaviors are illustrated hereinafter for purposes of demonstrating at least some appropriate settings for the burst and sustained fractions as well as the corresponding limits.

Turning to Figure 7 and having described the BWA allocation method of the present invention in terms of maintaining allocated bandwidth, attention is now directed to the active packet priority process of the present invention. This figure illustrates APP_0 (Figure 3) including a representative cell having the information described with regard to Figure 3a contained by cell header 32. Of particular interest at this juncture is packet info 34. This information identifies the cell as one of either start of packet, middle of packet or end of packet. Accordingly, a packet is considered as being active if the indication is middle or end of packet. While a start of packet indication may signal the beginning of a multi-cell packet, such that additional cells are to follow, it is also possible that the packet consists only of a single "start" cell. The present invention implements an APP (Active Packet Priority) state identification process, for example, within each of the APP sections. It should be appreciated that the APP section associated with each flow is essentially identical. The resulting states are indicated for each flow to bandwidth allocation arbiter 90 as part of lines 92, 94 and 98 by augmenting cell header information as it passes through the APP section of each of the flows. In the instance of F_0 line 92 is used.

Referring to Figures 7 and 8, the latter illustrates steps performed in conjunction with the APP process of the present invention that are generally indicated by the reference number 302 and which is implemented within an APP calculator 304 shown in Figure 7. Again, the present example refers to F_0. At step 310, the APP cell type is extracted from cell header data 34 as the cell header data passes through APP_0. The APP calculator 304 determines the APP cell type for use in determining the APP value to be presented to arbiter 90 by augmenting line 92.

Still referring to Figures 7 and 8, following extraction of the APP cell header type at step 310, the APP value is determined immediately following step 310. To that end, step 312 determines if the cell type is "start of packet." If this is the case, it is implied that a packet is not in progress because the start of packet indication is given only to the first cell of a packet. In this regard, it should be appreciated that the cell being considered has not yet been transmitted. Therefore, a middle or end of packet indication represents a cell that is needed to complete transmission of a packet that is already underway. Conversely, a start of packet indication is the first cell of that packet and, hence, transmission of the entire packet is pending. For this reason, monitoring only for start of packet is necessary in step 314 to set the active packet value to false. Otherwise, step 316 sets the APP value to true. Step 318 then returns monitoring to step 310. It should be appreciated that there are a number of alternative ways in which to establish a start, middle, end indication for packets.

For example, a cell could be "marked" by an external agent as being a cell that should be given APP. That is, a cell is marked with APP at some upstream point. The APP, as utilized by BWA 90 will function nonetheless. As another example, an integer may represent the packet offset from the first word of data in a particular cell. In this case, APP is true when the offset is equal to zero. That is, the first word of a packet is pending transmission. All of these modifications are considered as being within the scope of the present invention so long as APP is determinable in accordance with the teachings herein. Specific examples illustrating the use of the APP value will be given below.

Referring to Figures 3 and 9, having described the bandwidth allocation method of the present invention as well as the APP priority process, one implementation of bandwidth allocation arbiter 90, as well as relevant portions of cooperating components within network switch 12 will now be described. For this purpose, each flow cell header queue supplies cell information to bandwidth arbiter 90 by way of its corresponding APP section. For example, APP_0 provides packet priority information on line 92 which passes through the APP section and which is augmented by raw cell header information in accordance with the active packet priority process described with reference to Figures 7 and 8. Line 92 is provided to a compare information multiplexer 400 along with other information from the BWA State component of the corresponding flow comprising a sustained priority (SP) and a burst priority (BP) indication. The origins of the SP and BP indications are shown in Figure 5 on lines 154 and 164, respectively. On each clock cycle, a flow counter 402 stores a flow count that is incremented on a per clock cycle basis so as to sequentially select the information from each flow for routing to a priority encode comparator 404. The latter includes a Candidate Cell input from multiplexer 400 and a Current Cell input. Output of the priority encode comparator is indicated as Selected Cell Update Result. As flow information passes through compare info multiplexer 400, a flow ID 405 for the selected flow is added to the priority indications. The flow ID is derived from factors such as, for example, if a cell is to be grouped (for bandwidth allocation purposes) with other cells such that all cells in that group have the same numerical flow value or if a cell is not to be grouped with any other cells. In this instance, a new (unused) flow ID is assigned to start another "group of cells." The latter is typically all cells "owned" by some single paying customer or all cells of a particular characteristic/type (e.g., all internet telephone calls have similar sustained rate and burst rate characteristics and can be grouped so long as the sustained fraction is increased for every "call" added to the particular flow). It is noted that these considerations reflect the parameters recited for use in flow control described above with regard to Figures 3 and 4 since per flow differences are defined in terms of those 4 parameters.

The information for a particular flow F is selected whenever the following equation is true (where N is the number of flows):

(1) $$F = \text{Flow Count} \bmod N$$

Other components in bandwidth allocation arbiter 90 include a current selected cell register 406 and a round robin group 408, indicated within a dashed line. The current selected cell register receives input from the Selected Cell Update Result output of priority encode comparator 404 and provides an output, in the form of the stored values, to the Current Cell input of the priority encode comparator. The round robin group is connected to priority encode comparator 404 and itself includes a sustained credit round robin (SRR) 410, a burst credit round robin (BRR) 412 and a no credit round robin (NCRR) 414. It should be appreciated that priority encode comparator 404 performs a comparison on every clock cycle during which a maximum of one word is transmitted on the network link. The objective of the selection process is to select the next cell to be transmitted while the transmission of a previously selected or released cell is already underway. In this regard, one of the previously selected cells being transmitted is likely to be composed of a plurality of words such that its transmission extends over a series of clock cycles. Therefore, during transmission of such a previously selected cell, the selection of the next cell to be transmitted is continuously updated on each clock cycle. It should be mentioned that, when a cell is selected for transmission, the selected cell is not immediately transmitted onto the network link, but is immediately transferred or released into transmit header queue 110 of Figure 3 from current selected state register 406. Insofar as the handling of a cell for priority purposes, however, the cell is, for practical purposes, considered as being transmitted upon its release into the transmit header queue. Accordingly, on every clock cycle, priority encode comparator 404 reads the Candidate Cell input from compare info mux 400 and also reads the Current Cell input from current selected cell register 406. The priority encode comparator then selects between the two inputs based on the inputs themselves and potentially based upon round robin arrangement 408. The input which wins the comparison, either the flow data read from the compare information multiplexer 400, determined by flow counter 402, or the current selected cell read from current selected cell register 406 is then saved in the current selected cell register for use at least in the next comparison. Potentially, the current selected cell may change on each clock cycle. Over the duration of transmission of a previously selected cell, therefore, the current selected cell in register 406 should represent the candidate cell having the highest priority, as selected from among the various flows.

Still referring to Figures 3 and 9, transmitter 100 provides select line 124 signals immediately upon de-queue of a cell header from transmit header queue 110, indicating that the cell identified at that instant by current selected cell register 406 is to be transmitted on its output lines connected to transmit header queue 110. This arrangement is considered as being advantageous since it serves to maintain the transmit header queue in a relatively full state. In response, header information for that cell is released into transmit header queue 110, under the requirement that the data is valid. Released data remaining in current selected cell register 406 is immediately invalidated, as will be described. A de-queued header will remain in cell data fetcher 122 until a sufficient portion of data has been retrieved from cell data storage section 42 so as to insure uninterrupted transfer of the cell. It should be appreciated that the use of the current selected cell register contents by transmit header queue, as described immediately above, is considered as being advantageous for several reasons. For example, this implementation is relatively fast and inexpensive to implement. At the same time, the contents of the register may be slightly stale. That is, one of the other flow inputs to compare info mux 400 may have a higher priority. Therefore, as an alternative implementation, an additional test can be performed to compare current selected cell register contents 406 to the flow values present at the inputs of compare info mux 400 such that the flow possessing the highest priority can be selected. Of course, the contents of that highest priority flow must be sent to transmit header queue 110 in an appropriate manner. In this light, any number of alternatives are possible with regard to keeping the contents of selected cell register 406 as up to date as is desired. Such alternative become more significant when additional flows are added.

Referring to Figure 10 in conjunction with Figures 3 and 9, one method, generally referred to by the reference number 500, will be described for updating current selected cell register 406. Accordingly, step 502 receives an internal clock signal 504. On each clock cycle, flow counter 402 (Figure 9) is updated using equation 1 to increment modulo the flow indicated by the counter based on the number of flows. Compare info mux 400 receives the updated flow indication from flow counter 402 and promotes the candidate cell for the indicated flow at the candidate cell input of priority encode comparator 404. Step 506 monitors select line 124 signals to determine if a cell transmission is initiated for any flow. If a cell is transmitted, step 508 is performed to determine if the data in current selected cell register 406 is valid. If the data is invalid, step 510 is performed, as will be described below. If, on the other hand, the data is valid, step 512 is performed in which data for the current selected cell is released into transmit header queue 110 and to round robin arrangement 408. Step 514 then marks the contents of current selected cell register 406 as invalid.

Another task performed responsive to detection of cell transmission step 506 is to update round robin arrangement 408, as will be described below. In addition, a released Flow ID 515 is used by a Select Queue Decode section 516 to provide each flow queue an indication on lines 518, 520 and 522 for flows 0, 1 and N-1, respectively. A positive indication is given on one of these lines to only one of the flows when its cell has been selected/released from current selected cell register 406. As still another task, when a word is transmitted, a transmit decoder 524 receives information related to the transmitted word on a line 526 which includes information provided from select line 124 of Figure 3. This information includes the flow queue ID associated with the transmitted word and an indication of the sustained priority state of the flow at the time the cell was selected. Transmit decoder 524 provides each flow queue with data indicating whether a word was transmitted for that flow and if that word had sustained priority using lines 530, 532 and 534 for the respective flows. That is, transmit decoder 518, receives flow ID and priority information on line 526 to produce a unique signal on lines 530, 532 and 534, indicating to one flow that a word from that flow was transmitted with sustained priority. These unique signals are derived from the combination of lines 166 and 168 of Figure 3, while data from selected queue decoder 516 and transmit decoder 524 is transferred from arbiter 90 to the respective flow queues on lines 536, 538 and 540.

In the alternative instances wherein step 506 does not detect a transmit selection on a particular clock cycle , step 510 is performed. In this step, the contents of current selected state register 406 and compare info mux 400 are fed into priority encode comparator 404 at its Candidate Cell and Current Cell inputs, respectively. In step 542, the two priority encode comparator inputs are compared to select the cell/flow having the best or highest priority data. Step 544 determines if the Candidate Cell is selected. In this case, the contents of the current selected cell register are to be replaced. Step 546 is then performed to load the Candidate Cell data into current selected cell register 406 via the Selected Cell Update Result output. This information includes cell information, priorities and Flow ID. If step 544 determines that the current cell data, already contained by current selected cell register 406, has a relatively higher priority, step 548 loads the Current Cell data back into current selected cell register 406 via the Selected Cell Update Result output. Following either of steps 546 or 548, step 550 validates the data now stored in current selected cell register 406. At step 552, execution moves to the next clock cycle. If, alternatively, step 508 determines that the current selected cell data in register 406 is invalid, the contents of register 406 are replaced with the candidate cell in step 553 and the method then moves to step 542, continuing as described above.

Attention is now directed to Figure 11 which illustrates, in flow diagram form, one process, indicated generally by the reference number 600, for use in updating round robin arrangement 408. The latter forms part of the componentry illustrated in Figure 9. Process 600 follows step 512 of Figure 10. Initially, step 602 verifies that a cell was selected. If transmit selection step 602 indicates "NO" on a particular clock cycle, step 604 follows and represents that no change is made in the priority allocation states. Step 605 therefore follows step 604, to return the process to step 602. That is, Current Selected Cell register 406, SRR 410, BRR 412, and NCRR 414 remain unchanged. If, however, transmit selection step 602 indicates that a cell has been selected, step 608 then determines if the released selected flow/cell has sustained credit. If the flow has no sustained credit, then no credit round robin NCRR 411 is updated in step 610. If the step 608 test indicates that the selected flow has been transmitted with sustained credit, sustained credit round robin, SRR 410 is updated in step 612. Following step 612, step 614 determines if the flow has burst credits. If so, burst credit round robin, BRR 412 is updated in step 616. When a round robin is updated, it is set to the flow number just selected plus one. For example, if a round robin indicates flow 0, but flow 1 was selected since flow 1 did not have the appropriate priority, the round robin would be set to flow 2. If flow 2 were selected in the same scenario, the round robin would be set to flow 0. On the other hand, if the round robin points to the selected flow, the round robin is set to the next highest flow number. After steps 610, 616 or after a negative determination on step 614, operation returns to step 602 via step 605, to continue monitoring for a transmit selection on a clock cycle. Specific details as to the use of round robin arrangement 408 in making priority determinations will be described below.

Referring to Figure 12, one embodiment of priority encode comparator 404 is diagrammatically illustrated, generally indicated by the reference number 404a. The reader is referred to Figure 9 for component interconnections illustrated therein. On each clock cycle, a different set of data is presented on the Candidate Cell input from compare information multiplexer 400. This data is to be compared with the data present on the Current Selected Cell input from register 406. The latter input is used to construct a current integer 640, for example, within an appropriate register. A candidate integer 650 is constructed, for example, in another register based on the candidate cell input from multiplexer 400. Thus, the comparison is made between current integer 640 and candidate integer 650. Details relating to the construction of these values will be provided at an appropriate point below. Once created, however, the two values are compared in a way which determines the greater of the values as represented by a comparator 652. The greater of the values is indicated to a flow comparison section 653 along with the candidate and current register information.. The flow comparison section tests the flow ID's from the registers. If the flow ID's are identical, a signal is sent to a current/candidate multiplexer 654 on a line 655 indicating that candidate cell information is to be used. If the Flow ID's are not equal, the winner of the comparison made by comparator 652 is to be used by multiplexer 654. Accordingly, multiplexer 654 includes a candidate cell input 656, a current cell input 658 and a selection or result output 660. The data from the input selected in accordance with the indication on line 655 is presented at selection output 660. The selection is then routed to current selected cell register 406 (Figure 9).

Referring to Figures 9 and 12, current integer 640 is computed as the bitwise concatenation of data presented at current cell input 658 from priority encode comparator 404. In the present example, current integer 640 is made up of a cell info valid field 642a; a current sustained priority indicator 642b, comprising a single bit; a current burst priority field indicator 642c, also comprising a single bit; a current active packet priority indicator 642d; and a distance (DS) indicator, 642e. Cell info valid field 642a is manipulated by steps 514 and 550 of Figure 10. The current sustained and burst priority indicators are originally obtained from lines 154 and 164, respectively, of Figure 5. APP indicator 642d is derived from the APP computation section of each flow as shown in Figure 3 and described with regard to Figures 7 and 8. The value for distance indicator, DS, is produced using, for instance, a current flow distance calculator section 662 in conjunction with a current round robin selector 664. Round robin selector 664 receives inputs from each round robin within round robin group 408 of Figure 9 and also receives the SP and BP indications forming portions of the current selected cell input data. The SP round robin (SRR) input is received on a line 666, the BP round robin (BRR) input is received on a line 668 and the no credit round robin (NCRR) input is received on a line 670. The particular round robin used in determining the value of DS is established in view of SP and BP as received on lines 672 and 674, respectively. If the SP value is set, indicating that the current selected cell has sustained priority while the BP value is zero, indicating that the current selected cell has sustained priority, SP round robin 410 is used in contributing to the value of DS. If SP and BP are both set, indicating that the current selected cell has both sustained and burst priority, round robin selector 664 chooses the burst priority round robin value to be passed to calculator 662. If SP and BP are both zero, indicating that the selected cell has no credit, round robin selector 664 chooses the no credit round robin value to pass to calculator 662.

Referring to Table 1 in conjunction with Figures 9 and 12, DS calculator 662 determines the distance value in accordance with Table 1, using the round robin pointer value passed from round robin selector 664 along with the flow ID provided on a line 680 from current integer register 640.

FlowID	RR Pointer Value	DS Value	FlowID	RR Pointer Value	DS Value
0	0	0	2	0	2
0	1	3	2	1	1
0	2	2	2	2	0
0	3	1	2	3	3
1	0	1	3	0	3
1	1	0	3	1	2
1	2	3	3	2	1
1	3	2	3	3	0

TABLE 1
Distance (DS) Determination

It should be mentioned that Table 1 contemplates the presence of four flows, however, any number of flows may be utilized in accordance with these teachings. Essentially, the DS value is determined based on the flow ID of the cell being represented by current integer 640 and the flow pointed to by an appropriate one of the round robins, as selected by round robin selector 664. For example, if the flow ID is 1, the selected cell has only sustained priority and the sustained round robin points to flow 2, a DS value of 3 is used. A relative priority boost is provided with an increase in the value of DS, as a result of the addition of DS to current integer 640. Candidate register 650 includes essentially identical components for use in establishing the DS value for the candidate cell, consistent with the foregoing discussion relating to current integer 640. These components have been identified using identical reference numbers having an appended prime (') mark and have not been described for purposes of brevity. The underlying purpose of round robin arrangement 408 and associated round robin pointers and distance calculators is to break priority ties between the current cell and the candidate cell. That is, with all other data being equal as between current integer 640 and candidate integer 650, DS will break ties between the flows in a fair manner alternating from one flow to the next. In this regard, it should be appreciated that the order of the round robin pointer values that appear in the round robin column of Table 1 for each flow ID may be reordered so long as each of the round robin pointer values is included for each of the flow ID's.

It should be noted that the encoded priority scheme shown in Figure 12 relies on the use of APP on a low order priority basis. That is, if the current and candidate registers are tied in all positions above the APP position, packet priority will break the tie in the event that the represented cells possess different APP values, the formulation of which has been described above. In this manner, latency as to transferring a packet for which a first cell has already occurred can be improved.

As mentioned above, the round robins are updated only when a cell is transmitted. In some instances, for example during transmission of a previously selected cell having a length in words that is greater than the total number of flows, it is possible to compare a current selected cell in current register 640 to itself in candidate register 650. In this event, the contents of candidate cell register 650 always replace the contents of candidate register 640 since the priority information relating to that cell may have changed subsequent to the comparison which previously placed its information into current cell register 640.

Referring now to Figure 13, it should be appreciated that alternative encodings of the data that is used in the composition of current integer 640 and candidate integer 650 will result in the implementation of alternate priority arrangements. Figure 13 illustrates a first alternative encoding of the candidate and current integers which is generally indicated by the reference numbers 640a/650a. As in Figure 12, the leftmost bit is the most significant bit. Accordingly, this first alternative encoding prioritizes active packet selection above all other priority factors. That is, active packet priority, APP, cell selection is given higher priority than bandwidth allocation represented as the combination of SP and BP. This encoding is useful, for example, when minimizing latency is more important than maintaining actual bandwidth allocation goals.

Figure 14 represents a second alternative encoding which is generally indicated by the reference numbers 640b/650b. This encoding uses a per cell priority (PCP) field 700 in place of the DS field. Using this encoding, ties are broken in accordance with a strict priority scheme. That is, fairness in breaking ties is given essentially no weight. The priority is per cell and can either be extracted from the cell itself or the priority may be inferred from the flow number (e.g., flow 0 is the lowest priority while flow N-1 is the highest priority). The present invention contemplates that an unlimited number of alternative encodings are possible in view of this overall disclosure and that all of these alternative encodings are within the scope of the present invention. For example, the encoding can be arranged to do best effort active packet priority while absolutely maintaining minimum bandwidth allocations. A suitable encoding for this is: [Valid, SP, BP, APP, DS] arranged in the manner of figures 13 and 14. Another example is an encoding which does best effort bandwidth allocation but absolutely maintains minimum packet latency using an encoding: [Valid, APP, SP, BP, DS].

Having described implementation of the present invention in the foregoing discussions, attention is now directed to specific examples which are intended to illustrate the operation of the present invention as well as pointing out its various attendant advantages. For purposes of clarity, certain constraints and/or assumptions may be imposed in the examples which are not required in an actual implementation, but which are intended only to foster the reader's initial understanding.

Attention is now directed to Figure 15 which is a sustained priority transfer example generally indicated by the reference number 800. The horizontal or X axis of the figure measures time in units labeled successively from 0-9 and A-Z with the letters "O" and "Q" omitted from the series for the purpose of avoiding confusion. Each time unit may be, for example, the amount of time required to transmit a single word, the amount of time required to transmit a single cell of a given size, or the amount of time required to transmit some arbitrary amount of data such as 1500 bytes which could comprise multiple cells of variable individual length. For present purposes, the time units are illustrated as being of equal width across the horizontal axis, but this constraint is not a requirement. The vertical or Y axis denotes units of bandwidth. Only fully consumed blocks of bandwidth are considered. For example, if the X axis units are "time for 128 bytes of data to arrive" then the Y axis unit value indicates the "number of 128 byte portions of data to arrive." A partially filled bandwidth block is not considered. It should be emphasized that this constraint is imposed for illustrative purposes only and that such a constraint is, likewise, not required in an actual implementation. For example, if a bandwidth block is a byte, less than 8 bits could be transferred, since such a transfer may be required at times.

Figure 15 contemplates three flows F_0-F_2 passing through switch 12 of the present invention. The upper portion of the figure illustrates ingress data 802 representing the arrival of the data from each flow at the network switch. The lower portion of the figure illustrates queued data 804 for each flow awaiting transmission along with a lowermost egress data map 806 showing the order of transmission of the data obtained from queued data 804. As mentioned, it is assumed that, on either an input port or an output port of the network switch, a block of data takes up the entire link for that unit of time. That is, for the ingress data, a single block indicates that one of the input ports was fully utilized for that time period by the corresponding flow's data or, for the egress data, that one of the output ports was completely utilized for that time period. Ingress data may represent data inbound from any of the input ports of switch 12, the combination of which determines the depth of data blocks stacked vertically along the Y axis. For an 8 port switch, as an example, the maximum stack depth is therefore 8. Each data block is labeled with the notation <Flow ID (IngressTime) >.<Time(Ingress#)>.<Packet #(Channel ID) > where <Time > is the Y position of the block in the ingress stack. Packet # is not relevant to the present example, but will be considered in conjunction with at least one subsequent example. Therefore, every block/data unit is labeled uniquely for purposes of tracking its progress through the example. It should be appreciated that a block arriving on a flow at time 0 will be queued at time 1. This same data unit will then be delayed by at least one additional time period prior to being transmitted on the egress data map. See, for example, block 0.0.0 in time period 0 on ingress data F_0, time period 1 on F_0 queued data 804 and time period 2 on egress data map 806. Hence, the blocks are skewed to the right in moving down the figure from an ingress position, to a queued position, to an egress map position. Sustained priority ties are broken round robin in the order 0,1,2,0,1,2... in cases where all three flows are able to consistently maintain sustained priority. Generally, where one flow has priority greater than the other flows, but has no queued data for transmission and the other flows are tied at a lower priority, the round robin corresponding to the tied priority level is relied on to break the tie and is subsequently updated.

Table 2: Sustained Priority Allocation Example

	Sustained Fraction	Sustained Limit	Burst Fraction	Burst Limit
Flow 0	25%	Unused	Unused	Unused
Flow 1	25%	Unused	Unused	Unused
Flow 2	50%	Unused	Unused	Unused

Referring to Table 2 in conjunction with Figure 15, the former illustrates sustained fraction assignments invoked in the present example. It is noted that the sustained limit, burst fraction and burst limits are not used since the purpose of this example is to facilitate an understanding of the sustained fraction. The latter appears in Figure 4 designated by the reference number 72. Accordingly, F_0, F_1 and F_2 are respectively assigned sustained fractions of 50%, 25% and 25%. All data units arrive for the three flows in succession, as shown. Due to skewing, the first data unit to be transmitted at time 2 is 0.0.0.

Considering Figure 16 in conjunction with Figure 15 and Table 2, the former presents ingress data 802, queued data 804 and egress data 806 in a form which is intended to aid the reader's understanding of the subject priority decisions, accompanied by other information which contributes, in certain instances, to these decisions including the specific amount of Sustained Credit available to each flow and the values of the no credit and sustained round robin pointers for each time period. For purposes of this example, each flow begins with 1.0 SP credits. The SP priority level is set to 2.0 credits. It should be appreciated that this is the value specified in SC threshold register 152 (see Figure 5) of the BWA state machine for each flow. The NCRR and SP pointers values are also shown and are both initially set to point to flow 0 at time 0.

At time 1, no transmission takes place due to the aforementioned skewing effect. Accordingly, all priority states are unmodified. Priority comparison of the three flows results in selecting flow 0 for transmission at time 2 since there is a three-way no credit tie which is broken by the NCRR. Therefore, data unit 0.0.0 is transmitted at time 2. The SP flow credits at time 2 are all incremented by the respective sustained fractions. Flow 0 is not decremented as a result of its "no ticket" transmission and hence also moves closer to sustained priority. The NCRR is updated while the SRR is not updated. Based on these new values at time 2, a three-way no credit tie still obtains in the priority determination performed on the time 2 information with selection of flow 1, based on the NCRR. At time 3, the flows all receive their designated SP credits, flow 1 is not decremented and the NCRR is updated to point to flow 2. At the same time, only flow 0 has SP. Therefore, at time 4, data unit 0.1.0 is transmitted. Sustained credit values for flows 1 and 2 are incremented by their sustained fractions. Flow 0 credits are incremented by the flow 0 sustained fraction of 0.5 while being debited by 1 credit for transmitting a word resulting in a net reduction of in the amount of 0.5 credits. The SRR is set to flow 1. This process continues with subsequent time periods. It is of interest, however, to note certain events which unfold. For example, at time 7, flow 2 wins arbitration by having SP. Accordingly, the SRR is set to flow 0 in the time 8 row. Then at time 8, flow 0 again wins the arbitration. The SRR is updated by setting it to the ID number of the flow transmitted plus one. That is, the SRR is again set to flow 1. In the arbitration at time D, the flows are all tied at SP. Because the SRR points to flow 0, its data unit should be transmitted at time E; however, the flow 0 queue is empty. Therefore, the next higher flow with SP (flow 1) is permitted to transmit at time E. Accordingly, the SRR is set to the value 2 in the time E row. In the determination at time F, flow 0 still has nothing to transmit while flows 1 and 2 are in a no credit tie. In this case, the NCRR is used to break the tie, permitting flow 1 to transmit data unit 1.4.0 at time G. The NCRR is then set to point to flow 2 in the time G row. Over times J-M only flow 2 has data to transmit. The first transmit, at time J, is accomplished with SP causing the SRR to be set to 0 in the time J row. The times K and L transmits are made with no credit causing the NCRR to be set to 0 in each of these rows. Flow 2 regains sustained priority for the transmission of 2.8.0 at time M, but the SRR is once again re-set to flow 0. At time N, all flows are out of queued data such that idles occur (no transmission). Throughout these idles, the priority state of each flow is maintained. The maintenance of the states would also occur in situations where the flow queues still contain data but, for some reason, no transmissions are taking place. Such a situation could result, for example, if a downstream link has failed or due to downstream flow control which can cause flow constriction via clock control.

Table 3: Sustained Priority with Unallocated Available Bandwidth

	Sustained Fraction	**Sustained Limit**	**Burst Fraction**	**Burst Limit**
Flow 0	33%	Unused	Unused	Unused
Flow 1	25%	Unused	Unused	Unused
Flow 2	20%	Unused	Unused	Unused

Turning now to Table 3 in conjunction with Figure 17, it is noted that in Figure 15, as well as in similar remaining figures, the conventions used in Figure 15 are repeated. Figure 17 illustrates an available priority bandwidth allocation example performed in accordance with the present invention and generally indicated by the reference number 820. Ingress data is designated by the reference number 822, queued data is designated by the reference number 824 and the egress data map is designated as 826. Figure 17 illustrates flows F_0-F_2 passing through switch 12 of the present invention under the SP configuration parameters specified by table 3. It is noted that F_0 is assigned a sustained allocation fraction of 33%. This value is satisfied by the average arrival of a data unit in one out of every three time periods. However, based on the ingress data shown, two extra data units 0.6.1 and 0.P.1 arrive in time periods 6 and P, respectively. It is noted that these extra data units, in the present example, have the channel ID 1 while the other F_0 data units have the channel ID 0 such that the channel 0 and 1 data units may comprise data initially derived from different packets. At the same time, F_1 is assigned an SP allocation of 25%, thereby allowing an average receipt of one data unit over each four time periods. In this instance, F_1 receives ingress data exactly consistent with its SP allocation. Similarly, F_2 has an SP allocation of 20% and receives ingress data in one out of each five time periods. The total allocation to the three flows is 78% such that a portion of bandwidth is expected to go unused.

Referring to Figures 17 and 18, each flow once again begins with 1 SP credit and the SP credit level is set at 2 credits. Figure 18 is similar to previously described Figure 16, showing sustained credits and pointer values. Other similar figures will be provided, as appropriate. Over time periods 2-4, the NCRR decides the data unit to be transmitted since no flow has reached the SP level of 2 credits. At time 5, data unit 0.3.0 is transmitted with SP. At time 7, it is of interest to note that data units 0.6.0 and 0.6.1 arrive on the flow 0 queue. Flow 0, at this time has SP and, therefore, data unit 0.6.0 is immediately transmitted at time 8. Data unit 0.6.1 remains in the flow 0 queue. The credit situation among the three flows at time 8 is such that there is a three-way no credit tie. Accordingly, if other flows have queued data, the three-way tie is broken using the NCRR, which happens to be pointing to flow 0. It is noted, however, that only the flow 0 data unit is ready to transmit and would be selected for transmit at time 9, irrespective of the NCRR value. At time A, data unit 1.8.0 is transmitted with SP as operation continues. It is of further interest to note that, at time R, data units 0.P.0 and 0.P.1 arrive in the flow 0 queue. A two way SP tie exists between flows 1 and 2; however, the SRR pointer indicates flow 1. Therefore, data unit 1.P.0 is transmitted at time S and the SRR is updated. At time S, flows 0 and 2 have SP with the SRR now pointing to flow 2. Therefore, data unit 2.R.0 is transmitted at time T and the SRR is again updated. Also at time T, the priority determination finds flow 0 alone having SP. Accordingly, data unit 0.P.0 is transmitted at time U. Data unit 0.P.1 is then transmitted at time V, since this is the only data unit ready to transmit in the flow queues.

Still considering the example of Figures 17 and 18, it is emphasized that the excess incoming bandwidth on flow 0 has been accommodated in a highly advantageous way using available bandwidth which is unused by other flows. The extra data units, 0.6.1 and 0.P.1, are transferred with no credit at times 9 and V, respectively, keeping in mind that the actual decision to transmit is made in the preceding time period. At the same time, flow 0 is not penalized for using the available network bandwidth. The decision to transmit the "extra data" is made instantaneously without the need to monitor any particular aspect of operations. That is, for example, no priority tables need be updated. More importantly, no available bandwidth is permitted to go idle pending any sort of decision to transmit. Leftover bandwidth is instantaneously and immediately allocatable to any flow. Any flow having queued data may transmit so long as no other flow has superior priority and in accordance with selection using the NCRR, if some other flow is tied at no priority. A flow receiving the leftover bandwidth is not debited either type of priority credit (sustained or burst) as a result of making this leftover bandwidth transmission. In essence, such a flow has been given a "free ticket." It is submitted that prior art priority allocation schemes are devoid of this scheme. For example, while ATM provides an Available Bit Rate (ABR) concept that is intended to account for leftover bandwidth, ABR is not capable of such instantaneous response. In effect, ATM must readjust allocation parameters to "after the fact" account for leftover bandwidth. Some of the leftover bandwidth is inherently wasted in this manner. Moreover, it should be appreciated that the present invention permits alternate transfer of data with allocated priority and used leftover bandwidth as part of the same flow.

Attention is now directed to Figure 19 which illustrates a reduced network bandwidth example carried forth in accordance with the present invention and generally indicated by the reference number 830. Ingress data is designated by the reference number 832, queued data is designated by the reference number 834 and the egress data map is designated as 836. It is noted that this example is identical to aforedescribed Figure 15 with respect to ingress of the data on the three flows. The present example is also identical to that of Figure 15 over the first 7 time periods. In this regard, the data units within queued data 834 are identical to those in queued data 804 of Figure 15. At time period 8, however, it is assumed that the available bandwidth on the output network link is reduced by 50%, as is represented by doubling the clock period for egress data map 836 at time 8. This condition persists until time P.

Referring to Figures 15, 16 and 19, comparison of egress data 806 (Figure 15) and egress data 836 (Figure 19) reveals that the order of egress of the data units is identical in both examples. For this reason, the series of priority allocation decisions in Figure 19 is the same as those in Figure 15 except that from times 8-N, the decisions in Figure 19 are subjected to a delay as a result of the reduction in network bandwidth. The only real difference between the two examples resides in the conclusion of all transfers at time X for the Figure 19 data rather than at time N for the Figure 17 data. The delay of eight time periods corresponds exactly to the eight (i.e., A-H) time periods over which the output bandwidth was cut in half. It should be appreciated that virtually any change in available output bandwidth, either up or down, will result in a proportional difference with respect to the overall time required to complete all of the data transfers. In the instance of reduction in output bandwidth, the present invention tolerates severe constriction or even a complete inability to transfer data without adverse consequences to the data awaiting transfer. Data transfer resumes unaffected when bandwidth again becomes available. This tolerance of bandwidth constriction is enabled through the use of percentages in the flow configuration parameters.

In sharp contrast, ATM uses flow configuration parameters typically in a strict data rate sense such as, for example, a particular Flow A gets 10000 bytes/second. In any system using this type of configuration parameter it is not possible to maintain the desired flow configurations for all flows in the face of a bandwidth constriction reduced below the sum of the per-flow data rates for all of the active flows. In a highly advantageous way, the present invention maintains the desired flow percentages even when the bandwidth approaches constriction at 0 bytes/second.

Table 4: Sustained Credit Counting Example

	Sustained Fraction	***Sustained* Limit (words)**	**Burst Fraction**	***Burst* Limit (words)**
Flow 0	25%	128	Unused	Unused
Flow 1	25%	128	Unused	Unused
Flow 2	50%	65	Unused	Unused

Turning to Figures 20-21 and Table 4, a more detailed, sustained credit counting example is illustrated generally indicated by the reference number 840. In this example, both the sustained fraction and sustained limit values are specified, for instance, in units of words. These parameters appear in Figure 4 designated by the reference numbers 72 and 74, respectively. Only queued data is illustrated due to drawing space constraints and because illustration of ingress data is not necessary to an understanding of the present example. As indicated by Table 4, the sustained limit for flows 0 and 1 is set to 128 words while the sustained limit for flow 2 is set to 65 words. The sustained priority/no-priority boundary in the example is 64 word credits, specified in SC threshold register 152 of Figure 5. Therefore a flow having a number of sustained credits greater than or equal to 64 is given sustained priority. As mentioned previously, this limit may correspond to the number of credits required to transmit a full sized cell (i.e., 64 words transmitted in 64 consecutive time periods). The notational convention within the data blocks of Figure 22 is <FLOW>.<CELL>.<WORD#>. It should be appreciated that all words in a cell must be transferred without any other intervening words of data. As in previous examples, tie breaking is performed using the SP and NC Round Robins. It is further assumed that flow 2 never has any data to transmit. Therefore, flow 2 will reach its 65 word burst limit and remain there. Since this flow never has data to transmit it can essentially be ignored for purposes of determining the next cell to be transmitted upon completion of transmission of a previous cell.

Referring to Figures 20-22, at time 0, flow 0 has sustained credit (since it's SP credit value ≥ 64) while flow 1 does not have sustained priority at 63.75 credits. Accordingly flow 0 is selected for transmission (flow 0, cell 0). As flow 0 is transmitting, having sustained priority when the flow was selected, SP credits are being debited by 1 – 0.25= 0.75 word worth of credits for every word transmitted. Figure 22 plots the sustained credit available to each flow showing the credit count of flow 0 diminishing over times 0 and 1. The other active flow, flow 1, is not transmitting so, like flow 2, it is gaining 0.25 word worth of credit for every word transmitted by flow 0. It should be appreciated that the slopes at which the flows accumulate and lose credits is a function of the sustained priority fraction assigned to each flow. Transmission of word 0.0.1 at time 2 completes the transmission of flow 0, cell 0 since this cell consists of these two words of data. A decision is made at each time period as to which flow would have priority on the next time period. For example, such a decision is made at time 1 concerning which cell is to be transmitted at time 2. Since a word is pending for the cell under transmission and must be transmitted at time 2, this flow-to-flow decision will not be relied on at time 1. However, it should be appreciated that the decision process at each time period accommodates changing the balance of credits amongst the various flows for use in a subsequent decision at the completion of a cell. That is, the present invention continuously updates the current selected cell (stored in 406, Figure 9) which is to be transmitted next. It should also be mentioned that flow 0 was only debited for credits equal to the amount of credit required to transmit two words rather than a full sized cell.

At time 2, the next cell is selected for transmission. Flows 1 and 2 are tied at SP with the SRR pointing to flow 1. Therefore, flow 1, cell 0 is selected for transmission. This cell consists of four words that are transmitted over times 3-6. Consequently, the flow 1 credit plot in Figure 22 decreases in value across times 3-6 while flow 0 accumulates credits. Flow 0 hits its sustained credit limit of 65 at time 4. At time 6, flow 0 has sustained priority and flow 1 does not have sustained priority as a result of being debited for transmission of four successive words. Therefore, flow 0, cell 1 is selected for transmission at time 6. This cell is made up of four words that are transmitted at times 7 through A. For the priority decision at time A, flows 1 and 2 are tied with no credit with the NCRR pointer indicates flow 0. Therefore, flow 0, cell 2 is selected and its 2 cells are transmitted at times B and C. The NCRR is set to flow 1 at time B. At time C, another no credit arbitration is performed between flows 0 and 2 with flow 1 winning since the NCRR indicates flow 1. Flow 1, cell 1 is made up of five cells that are transferred over times D-H. At time H, flow 1 has regained SP and is selected to transmit a two word cell, 1.2.0 and 1.2.1 at times I and J, respectively. At time J, flows 0 and 1 have no priority. Because the NCRR indicates flow 2, flow 0 is selected to transmit a three word cell: 0.3.0, 0.3.1 and 0.3.2. Since flow 0 was selected the NCRR is set to flow 1 at time K. The network link is idle during times N, P and R since no data is available to transmit. At time R, flow 1 has SP and is selected to transmit a three word cell: 0.4.0, 0.4.1 and 0.4.2 at times S, T and U, respectively. At time U, flow 1 is the only flow with data to transmit and is selected to transmit its third cell made up of four words: 1.3.0, 1.3.1, 1.3.2 and 1.3.3.

Briefly considering flow 2 in the present example, it is noted that the sustained credit limit is reached at time 4 and grows no more after this point. The limit prevents a flow from accumulating a great number of sustained credits which would translate into a long period of sustained priority. If no limit were provided, there is a potential, unwanted burst of traffic (from flow 2) that could be quite long. The sustained limit is set to limit the accumulation of sustained credits allowing only manageable burst lengths. In this example, the manageable burst tolerance is very low. Hence, if flow 2 were to become active, its cells would be highly interleaved with the other active flows.

Table 5: Burst Fraction Example

	Sustained Fraction	Sustained Limit	Burst Fraction	Burst Limit
Flow 0	50%	Unused	100%	Unused
Flow 1	25%	Unused	0%	Unused
Flow 2	25%	Unused	0%	Unused

Referring to Figure 23 and Table 5, a burst fraction example is illustrated, generally indicated by the reference number 850. In this example, the sustained and burst fractions are both utilized. These parameters appear in Figure 4 designated by the reference numbers 72 and 76, respectively. Per the sustained fraction allocations of the three active flows in Table 5, the network link output is 100% utilized such that there is always queued data in order to transmit. Flows 1 and 2 have completely uniform arrival patterns. Flow 0 is fully bursty. That is, flow 0 fully provides its 50% egress bandwidth but does so by bursting at 4x the configured egress bandwidth allocation, but only bursts one quarter of the time. Further, in the egress data, data originating from adjacent bursts are separated by a fixed amount of time present between the last cell of the earlier burst and the first cell of the latter burst. Accordingly, it should be appreciated that any additional delay in completing the transmission of all the cells of a burst will impose a corresponding additional delay in starting transmission of the next burst. Delay notations 857 in the example of Figure 23 show that a subsequent burst follows after the end of the transmission of the last cell in of the current burst by 1 time unit. Therefore, the goal in using the burst fraction here is to configure the burst fraction so that whenever there is a sustained rate tie, the tie is always broken in favor of flow 0. In the present example, this goal is accomplished by setting the burst fraction for flow 0 to 100%. That is, whenever flow 0 has sustained credit, it will always have burst credit which, in effect, insures that there is never an SP tie between flow 0 and any other flow; flow 0 will always win in the instance where it has credit. At the same time, flows 1 and 2 have their burst fractions set to zero so that they never accumulate burst credits. It is emphasized that the need to use the burst fraction in this example resides in the requirement that all cells on flow 0 must be transmitted as soon as possible on the output bus, prior to cells of flows 1 or 2. If this requirement is violated, the subsequent ingress of cells on flow 0 will ultimately be delayed such that the upstream device is unable to supply data at the 50% rate. In this event, the network switch executing the present example will not be able to sustain the output bandwidth of 50%.

Referring to Figures 23 and 24, the latter serves in understanding the priority allocation decisions made at each time period in the manner of previous Figures 16, 18 and 21. Descriptions of the priority allocation decisions are limited to those instances in which burst priority comes into play since all other priority decisions in the present example are consistent with previous descriptions. In this example, the sustained priority threshold is set to 2 credits. The reader is reminded that flow 0 will have burst priority at all times that it has sustained priority. Flow 0 first achieves sustained/burst priority at time 3. Flows 0 and 1 are still at no credit, below the 2 credit threshold. Therefore, flow 0 transmits cell 0.0.1 at time 4. Flow 0 again reaches sustained/burst credit at time 5 in what would be a three-way SP tie without the burst priority held by flow 0. Accordingly, flow 0 transmits cell 0.0.2 at time 6. Flow 0 then transmits cell 0.0.3 at time 8 with burst/sustained priority, completing transmission of the first burst of four cells initially received at time 0 in the flow 0 ingress data. Flow 0 accumulates burst/sustained credits over times 9, A and B, since no queued data is available to flow 0. Thus, at time B, when queued data is available to flow 0, 3.0 burst/sustained credits are available such that the flow transmits cells 0.A.0, 0.A.1 and 0.A.2 at times C, D and E, respectively, with burst/sustained priority. The last cell of the second burst, 0.A.3 is transferred at time G following flow 0 reaching burst/sustained priority at time F. This pattern essentially repeats in egress data 856 at times K and U for the remaining two bursts which appear in the ingress data at times I and S. It is noted that the burst round robin value is not shown in Figure 24 for the reason that flow 0 will always be selected if it has burst/sustained credit since no other flow attains burst credit. Each time flow zero transmits, the burst round robin is reset to point to flow 1. The BRR value is not relied on when only one flow has burst credit. However, the BRR operates in essentially the same manner as the SRR in cases where more than one flow has burst credit. Specifically, the flow pointed to by the BRR is selected if that flow has burst credit. If the pointed to flow does not have burst credit, the next highest flow with burst credit is selected. The BRR is then set to the selected flow ID plus one.

Still considering the burst priority example of Figures 22 and 23, the use of two types of credits assigned to a single flow is considered to be highly advantageous. In the present example, the use of burst credits permits a priority configuration that boosts the priority of flow 0 in a way that assures transmission of flow 0. Looking at egress data 856 and remembering that flow 0 is assigned a sustained fraction of 0.5 or 50%, it is observed that over the eight time periods from time 2 to time 9, flow 0 is allocated the output bandwidth on four of these time periods. In the next eight time periods from A through H, one-half of the output bandwidth is again given to flow 0. Over time periods I through R and S through Z, one-half of the output bandwidth is dedicated to flow 0. It is also observable that the priority configuration of this example provides one extra time period between completion of one flow 0 burst and starting the next flow 0 burst, implying that a smaller burst fraction will also work. Since there is one extra time period, the burst fraction can be reduced to allow the one additional intervening cell. Accordingly, a burst fraction of 0.667 or greater will accomplish the contemplated goals in this example. In other words, even at a setting of 0.667, burst credits will accumulate to a degree that is sufficient to insure timely transmission of the flow 0 data. The following example serves to illustrate the result when burst credits are not used, as in the immediately foregoing example.

Table 6

	Sustained Fraction	Sustained Limit	Burst Fraction	Burst Limit
Flow 0	50%	Unused	0%	Unused
Flow 1	25%	Unused	0%	Unused
Flow 2	25%	Unused	0%	Unused

Turning to Figures 25 and 26 along with Table 6, the example of previously described Figures 23 and 24 is repeated without the use of burst credits. That is, the burst fraction is set to zero for all three flows. With regard to ingress data 862, it is noted again that any delay in transmitting the last cell of one of the flow 0 bursts translates directly into additional delay which is imposed prior to transmission of the first cell of the next burst. Inasmuch as development of queued data 864 and egress data 866 is consistent with previous application of the SP allocation process of the present invention, a detailed discussion of the development of the egress data is not provided. With regard to the sole use of the SP fraction, one might expect that setting the sustained fraction to 50% is sufficient to insure proper transfer of bursty flow 0. A closer examination of this expectation follows immediately hereinafter.

Referring to Figures 23 and 25, comparison of the ingress and egress data between these two figures illustrates the value of the burst fraction concept of the present invention. Specifically, in egress data 866 (Figure 25), the sustained fraction is not achieved. While over the eight time periods from 2 through 9 four flow 0 cells are transferred, subsequent groups of eight time periods each contain only three flow 0 cells. This is the case for time periods A through H, I through R, and S through Z. It is important to understand that the gaps or idles in egress data 866 (Figure 6) at times C and L are a direct result of the ingress device (not shown) being unable to supply the data at the required rate in view of the imposed transmit delays for reasons described above. Delay is imposed with receipt of the flow 0 ingress bursts at times

B, K and V in Figure 25 as compared with times A, I and S, respectively, for the same bursts in Figure 23. Ultimately, the delay in receiving the ingress data bursts on flow 0 could be attributed to a number of factors.

One factor, as an example, is a switching device that aggravates an upstream device throughput issue by virtue of having limited buffering. That is, only a limited amount of data buffering is available in the switch for ingress traffic. Once the buffers have been filled, the switch must apply some sort of flow control back to the upstream device to prevent that device from sending more data for which no storage buffers exist, if that data were to arrive prior to a buffer being freed (typically via transmission). This flow control may cause the upstream device to incur additional latency once a buffer does free up in the switch and the upstream device receives the "ok to send" message.

Still referring to Figures 23 and 25, considering an upstream device that has a fixed latency between the time it is informed that it is "ok to send" and the time at which it is ready send, this "startup" latency will be incurred any time the switching devices buffers are completely filled. In this situation, it is advantageous for the bandwidth allocation configuration to be flexible enough prevent buffer overruns. If this is accomplished, the contemplated latency will occur only once (at the beginning of the burst transmission). Once the burst transmission has started, the latency is avoided by cooperation between the switching device and the upstream device. Accordingly, the upstream device may be bursty, avoiding the latency as often as possible by bursting out enough data to fill the switching device's buffers and then delaying further transmission to allow the switching device buffers to clear. In this regard, the present invention contemplates matching the network switch or device to an input device such that data is bursted out at the same rate as the incoming burst. This match is achieved using the combined sustained and burst fraction settings.

In considering a bursty upstream device that is not tolerant of latency, the present invention contemplates setting up an inbound burst for transmission upon receipt with a higher priority than the priority at which the originating, upstream device transferred the burst. This is, in essence, achieved by using the burst fraction as illustrated in connection with Figure 23. The benefits of this arrangement reside in reduced latency through maintaining flow of the burst in order to avoid the aforedescribed startup latency. At the same time, hardware costs may be reduced since there is a corresponding reduction in the amount of buffering that is required. That is, a buffer capable of storing a full sized packet is not necessary. Upstream device requirements are satisfied by efficiently buffering only portions of packets in conjunction with the highly advantageous priority allocation concepts of the present invention. In contrast, the typical prior art approach is to provide ever more buffering.

Referring to Figure 27, attention is now directed to the active packet priority feature of the present invention. The figure e illustrates an active packet priority example performed in accordance with the present invention and generally indicated by the reference number 870. The present example contemplates the application of the priority encoding described above with regard to Figure 13. That is, the active packet priority data is given seniority for the use of output network link bandwidth above all other considerations. Three packets are illustrated, one of which arrives on each of flows 0-2. The packets are all of the same length and arrive simultaneously for illustrative purposes. The packets are scheduled such that, as long as there is data for an active packet to be sent, that data is prioritized over sending any data from a new (non-active) packet. In an initial arbitration, the flow 0 packet is given the use of the network link to transfer cell 0.0.0, for example, as a result of the NCRR pointing to flow 0. Each successive cell of the flow 0 packet then has priority for transfer based on APP. Therefore, the flow 0 packet is transferred over times 2-7. At time 7, an arbitration is performed based on either the sustained or no credit round robin, depending upon the sustained fraction setting. The application of burst credits and the burst round robin may also be an arbitration factor, but is not considered in the present example for purposes of clarity.

Continuing to refer to Figure 27, as a result of the arbitration at time 7, the flow 1 packet is selected for transfer, for example, if the SRR indicates flow 0 and both flows 1 and 2 have gained sustained priority during the transfer of the flow 0 packet, flow 1 would be selected as having the next highest flow ID. The SRR would then be set to flow 2. Alternatively, if flows 1 and 2 do not have sustained priority, the NCRR would indicate flow 1 as a result of selection of cell 0.0.0 at time 1. Therefore, flow 1 would be selected using the NCRR. In either instance, the flow 1 packet is transferred over times 8-D. The flow 2 packet is then given the use of the network link since the flow queues of the remaining flows are empty. The flow 2 packet is transferred over times E-J. The last cell of the flow 0 packet appears in egress data 874 at time 7; the last cell of the flow 1 packet appears in the egress data at time D; and the last cell of the flow 2 packet appears in the egress data at time J. Transfer latency is indicated for each of the flow packets in terms of the total time taken to complete packet transfer as measured from time 0. Specifically, the latency for packet 1 is 8 time units while the latencies for the flow 1 and 2 packets are 14 and 20 time periods, respectively. Specific advantages of the APP feature will be discussed in conjunction with the next example.

Turning to Figures 28 and 29, an SP example, indicated by the reference number 880, is presented using ingress data 872 of Figure 27 without the use of APP. Because queued data 884 and egress data 886 are developed in accordance with the aforedescribed SP allocation technique of the present invention, a detailed description of the development of this data will not be provided. It is mentioned, however, that upon the flows achieving sustained priority at two credits, a succession of 0 to 1 to 2 counts is seen in the settings of SRR. It is noted that the sustained fraction of each flow is set to 1/3 so that each flow has equal access to the output bandwidth. As in the previous example, the latency of each flow is indicated. The latency of flow 0 is 18 time periods while the latencies corresponding to flows 1 and 2 are 19 and 20 time periods, respectively.

Comparing Figures 27 and 28, it is noted that the latency of the flow 2 packet is at 20 time periods in both figures. The advantage of the APP feature is evident, however, in view of the relative latencies of the flow 0 and flow 1 packets. The latter experiences a latency of 19 time periods in Figure 28 and a latency of only 14 time units in Figure 27: a reduction of 5 time periods. Remarkably, however, the latency of the flow 0 packet is reduced from 18 to only 8 time periods: a reduction of 10 time periods. Accordingly, it is submitted that the APP feature of the present invention is highly advantageous in environments where it is desired to transfer at least certain packets with the lowest possible latency. Average latency has been significantly reduced while worst case latency is unaffected.

One skilled in the art may devise many alternative configurations for the system and method disclosed herein. For example, it should be appreciated that the foregoing examples were selected for illustrating the features of the present invention in a way which is thought to best aid the understanding of the reader. However, in view of the foregoing disclosure, it is to be understood that the multiple credit type features, the APP allocation feature and remaining teachings of the present invention may be combined in any suitable manner. Moreover, a number of different types of devices may incorporate one or more features described in the foregoing discussions. These devices include, but are not limited to network switches, bus arbiters of any sort requiring any of the features described in the foregoing discussions and telephone switches. Therefore, it should be understood that the present invention may be embodied in many other specific forms without departing from the spirit or scope of the invention and that the present examples and methods are to be considered as illustrative and not restrictive, and the invention is not to be limited to the details given herein, but may be modified within the scope of the appended claims.

What is claimed is:

1. In a network link configured for transferring a plurality of data flows, each of which requires a minimum bandwidth allocation on the network link, a method for transferring said data flows on the network link, said method comprising the steps of:

receiving the data flows at a first end of the network link;

dividing the data flows into cells such that at least some of the cells vary in length with respect to one another;

transferring the cells on said link in a predetermined way such that cells of different data flows are multiplexed and the minimum bandwidth allocation of each data flow is maintained on the network link; and

at a second end of the network link, receiving the cells transferred on the network link, including the cells which vary in length, and combining the cells so as to reconstruct the plurality of data flows at the second end of the link.

2. The method of Claim 1 wherein the data flows are divided into cells such that the length of each cell does not exceed a maximum cell length and at least a portion of the cells include a length that is less than said maximum cell length.

3. The method of Claim 2 wherein the step of transferring the cells in said predetermined way includes the steps of:

allocating credits to each data flow associated with the transfer of the flows; and

before any one data flow is given the use of the network link in transferring a particular cell having a particular length, using accumulated ones of the credits to verify that the data flow requesting to transfer the particular cell has accumulated credits equal to at least a certain allocation.

4. The method of Claim 3 wherein each cell includes a length of at least one word and wherein the step of using the accumulated credits compares the length of the particular cell in words to the accumulated credits of the particular flow.

5. The method of Claim 3 wherein the data flows are divided into cells such that the length of each cell does not exceed a maximum cell length and at least a portion of the cells include a length that is less than said maximum cell length and wherein the particular length of the particular cell is less than the maximum cell length and the certain allocation is sufficient to transfer the maximum cell length.

6. The method of Claim 3 wherein the data flows are divided into cells such that the length of each cell does not exceed a maximum cell length and at least a portion of the cells include a length that is less than said maximum cell length and wherein the particular length of the particular cell is less than the maximum cell length and the certain allocation is sufficient to transfer the particular cell length, but is insufficient to transfer the maximum cell length.

7. The method of Claim 3 further comprising the step of:

upon transferring the particular cell which includes a length that is less than the maximum cell length, decrementing the accumulated number of credits available to the particular flow by an amount of credits corresponding to no more than the length of the particular cell.

8. The method of Claim 7 wherein the length of each cell is made up of a number of one or more words and wherein the total number of credits available to the particular flow is decremented by the number of words which make up the length of the particular cell by incrementally decrementing the total number of credits available to the particular flow by one word at a time as each word is transferred on the network link.

9. The method of Claim 3 wherein the step of allocating credits includes the steps of allocating at least two different types of credits to each data flow for use in an assignment of priority to each data flow.

10. The method of Claim 9 including the steps of making a priority decision based on a particular one of the different types of allocated credits such that a group of flows are tied at one priority and, thereafter, breaking the tie by using a pointer selected from a group of pointers having at least one pointer associated with each of the different types of credits such that the selected pointer is associated with the particular type of allocated credits.

11. The method of Claim 9 wherein each cell is made up of one or more words having a predetermined length and wherein the step of allocating credits to each flow incrementally allocates a fraction of one word.

12. The method of Claim 11 wherein the network link operates over a series of clock cycles and wherein credits are incrementally allocated to each flow on each clock cycle.

13. The method of Claim 9 wherein a sustained credit allocation, serving as a first type of said credits, is used to control at least for selected ones of the data flows the minimum bandwidth allocations on the network link specified for the selected data flows.

14. The method of Claim 13 wherein at least one of the selected data flows intermittently bursts data on the network link such that the sustained credit allocation for that data flow is exceeded on the network link during bursting and such that the sustained credit allocation controls average utilization of the network link over time for that data flow by balancing at least one burst time period during which that data flow bursts data on the network link with at least one quiescent time period during which that data flow does not utilize the network link.

15. The method of Claim 14 wherein a burst credit allocation, serving as a second type of said credits, is used to establish at least for the selected ones of the data flows a maximum utilization of the network link during bursting.

16. The method of Claim 15 wherein priority for the use of said network link is assigned to each flow based, at least in part, on an accumulated number of sustained credits and an accumulated number of burst credits.

17. The method of Claim 16 wherein a sustained priority is assigned to flows with accumulated sustained credits greater than a predetermined number.

18. The method of Claim 17 wherein more than one flow is assigned sustained priority such that a group of flows is tied at the sustained priority and wherein the tie is broken in a way which results in alternate transmission of those tied flows which form the group of flows.

19. The method of Claim 16 wherein a burst priority is assigned to flows with accumulated burst credits greater than a predetermined number.

20. The method of Claim 19 wherein more than one flow is assigned burst priority such that a group of flows is tied at the burst priority and wherein the tie is broken in way which results in alternate transmission of the group of flows.

21. The method of Claim 16 wherein the network link transfers a selected cell over a series of clock cycles based on a prior priority determination and wherein a current selected flow, identifying a candidate cell to next be transferred on the network link is updated on each clock cycle based, at least in part, on the accumulated number of sustained credits and the accumulated number of burst credits.

22. The method of Claim 21 including the step of updating the accumulated number of sustained and burst credits available to each flow on each clock cycle such that the identified candidate cell is continuously updated based upon changing numbers of accumulated sustained and burst credits during transfer of the selected cell.

23. The method of Claim 21 wherein sustained priority is assigned to any flow having an accumulated number of sustained credits greater than at least a first predetermined number and burst priority is assigned to any flow having an accumulated number of burst credits greater than at least a second predetermined number of burst credits and the current selected flow having sustained priority is replaced by a new current selected flow having burst priority.

24. The method of Claim 23 wherein more than one flow is assigned burst priority such that a group of flows is tied at the burst priority and wherein the tie is broken in way which results in alternate transmission of those tied flows which form the group of flows.

25. The method of Claim 24 including the step of providing a group of flow pointers including at least a sustained flow pointer associated with the sustained priority and a burst flow pointer associated with the burst priority such that one of the flow pointers is selected in association with the burst priority for use in breaking the tie at the burst priority.

26. The method of Claim 21 wherein transfer of the candidate cell of the current selected flow is initiated on the network link immediately upon completion of the transfer of the selected cell without further decision as to updating of the candidate cell.

27. The method of Claim 21 wherein the candidate cell forms a portion of a packet received on the current selected flow and wherein the current selected flow is updated on each clock cycle based, at least in part, on a packet priority which identifies a relative position of the candidate cell within the packet.

28. The method of Claim 27 wherein the relative position of the candidate cell is selected as a start of packet, a middle of packet or an end of packet.

29. In a network link configured for transferring packets originating from a plurality of data flows, a method for transferring said packets on the network link, said method comprising the steps of:

receiving the packets at a first end of the network link;

for at least specific ones of the packets, each of which includes a length that is greater than a maximum cell length, dividing each specific packet into a series of cells such that each of the cells is identifiable as forming a part of that specific packet;

transmitting at least the specific ones of the packets on said link incrementally by alternating each series of cells;

at a second end of the network link, receiving the cells thereon; and

identifying the specific packets from which each cell originated and, thereafter, using that identification to reconstruct the series of cells for each specific packet to produce the specific packets at the second end of the network link.

30. The method of Claim 29 wherein said transmitting step includes the step of multiplexing cells from different ones of the specific packets on the network link and said identifying step includes the step of demultiplexing the cells using the identification.

31. The method of Claim 29 wherein said transmitting step incrementally transmits the cells on the network link based, at least in part, on identification of the specific packet from which a particular cell originated.

32. The method of Claim 31 wherein a particular one of the cells is selected to be transmitted in order to reduce the latency of transfer of the specific packet which the particular cell identifies after having transmitted at least one other cell which identifies that specific packet.

33. The method of Claim 29 wherein each cell forming part of a particular one of the specific packets is identified as a start of packet, a middle of packet or an end of packet.

34. The method of Claim 33 wherein the transmitting step includes the step of designating the cells originating from the particular one of the specific packets using a particular one of a number of different channel identifications such that the cells from the particular packet are identified at the second end of the network link using the particular channel identification.

35. The method of Claim 34 wherein the particular channel identification is dedicated to the particular packet for the duration of the transfer of the particular packet on the network link.

36. In a network link configured for transferring packets originating from a plurality of data flows in which, for at least specific ones of the packets having a length that is greater than a minimum cell length, each specific packet is divided into a series of cells, the improvement comprising the steps of:

identifying a particular one of the specific packets from which each cell originated and, after transferring the packets across the network link, using that identification to reconstruct the series of cells for each specific packet and, thereby, the specific packets at a receiving end of the link.

37. The improvement of Claim 36 wherein the packet identification is used in a way which controls latency in the transfer of the specific packets across the network link.

38. In a network link configured for transferring a plurality of data flows, each of which requires a priority allocation of at least a minimum bandwidth on the network link, a method for transferring said data flows on the network link, said method comprising the steps of:

receiving the data flows at a first end of the network link;

forming cells using the data flows such that each data flow is made up of a series of one or more cells;

initiating the transmission of the data flows on said network link by transferring one or more cells on the network link which form portions of the flows and consistent with the established priority allocation; and

at a time when the use of the network link is not allocatable to any of the data flows consistent with the established priority allocations of the data flows, transferring at least one additional cell forming part of one of the flows irrespective of the priority allocation assigned to that flow.

39. The method of Claim 38 including the step of receiving certain ones of the cells transmitted consistent with the established priority allocation for a particular flow along with at least one cell transmitted irrespective of the established priority allocation for the particular flow such that the certain cells include a cell configuration and the one cell transmitted irrespective of the established priority allocation for the particular flow includes a configuration that is identical to the cell configuration of the certain cells.

40. The method of Claim 38 wherein said minimum bandwidth for each data flow is defined using an accumulated number of credits assigned to each flow such that a request for the use of the network link by a particular one of the flows is granted when the particular flow has more than a predetermined minimum number of accumulated credits and, when the particular flow requests the use of the network link with less than the predetermined minimum number of accumulated credits, maintaining the accumulated number of credits assigned to the particular flow when the use of the network link is granted to the particular flow irrespective of the priority allocation assigned to the particular flow.

41. The method of Claim 38 wherein a plurality of data flows request the use of the network link at a time when the requesting flows have no allocatable priority consistent with their priority allocations for the use of the network link and the one additional cell is selected forming part of one of the requesting flows in a predetermined way.

42. The method of Claim 41 wherein the additional cell is selected in said predetermined way from a particular one of the requesting flows using a no credit round robin which alternately points to one of the data flows.

43. The method of Claim 42 wherein the requesting flows have assigned flow identifications and wherein the particular requesting flow is selected in said predetermined way based, at least in part, on the assigned flow identification of the particular requesting flow in relation to the flow which is pointed to by the no credit round robin.

44. In a network link having an available network link bandwidth which potentially varies over time and which is configured for incrementally transferring a plurality of data flows, for at least specific ones of which preferred minimum bandwidth allocations are specified, a method for transferring the data flows on the network link, said method comprising the steps of:

receiving the data flows at a first end of the network link for transfer thereon to a second end of the network link;

based on the preferred minimum bandwidth allocations, establishing sustained percentages of the available network link bandwidth to be assigned, respectively, to each of the specific data flows for use during incremental transfer of the specific data flows on the network link; and

responsive to variation, over time, of the available network link bandwidth, controlling actual use of the available network bandwidth by the specific data flows during the incremental transfer of the specific data flows across the network link such that each of the specific data flows is allocated approximately its respective sustained percentage of the available network bandwidth over time.

45. The method of Claim 44 wherein each of the specific data flows is transferred over a series of clock cycles and wherein said sustained percentages are specified as sustained fraction allocations that are credited to each of the specific flows on each clock cycle such that, during each clock cycle, each of the specific flows is potentially maintainable at a rate corresponding to its sustained fraction allocation.

46. The method of Claim 45 wherein data which makes up each specific flow is incrementally transferred on the network link in intermittent bursts such that, at some points in time, no data associated with at least a particular one of the specific flows is present on the network link and, at other points in time, data associated with the particular flow is present on the network link and wherein the step of controlling actual use of the available network bandwidth includes the step of using a burst fraction allocation which limits bandwidth allocatable at least to the particular flow only during the other points in time when data associated with the particular flow is present on the network link such that an average allocation to the particular flow over time on the network link is approximately equal to its sustained fraction allocation.

47. The method of Claim 46 wherein the burst fraction allocation is assigned to the particular flow such that at least a predetermined burst percentage of the available network link bandwidth is available to the particular flow on the network link during the incremental burst transfers of the data associated with the particular flow.

48. The method of Claim 45 wherein said flows are transferred on the network link using cells which vary in length up to a predetermined maximum length and which contain data associated with the flows.

49. The method of Claim 48 wherein the sustained fraction allocations are each equal to a fractional portion of the predetermined maximum length of one of the cells.

50. The method of Claim 48 wherein each cell is made up of words having a length in words and wherein the sustained fraction allocations are each specified as being less than the length in words such that no more than one word is credited to each flow on one of the clock cycles.

51. The method of Claim 50 wherein each of the sustained fraction allocations is specifiable as being less than one percent of the length in words.

52. The method of Claim 48 including the step of tracking the credits available to each of the specific flows such that the available credits for each of the specific flows accumulate in increments which are less than the maximum predetermined length of one cell.

53. The method of Claim 52 wherein each cell is made up of words having a length in words and wherein the sustained fraction allocation for each of the specific flows is specified as being less than the length in words such that no more than one word is credited to each of the specific flows on one of the clock cycles and such that the available credits for each of the specific flows accumulate in increments per clock cycle which are less than the length of one word.

54. The method of Claim 52 wherein at least a certain cell transmitted on the network link, including data associated with a particular one of the specific flows, includes a certain length that is less than the predetermined maximum cell length and wherein the available credits for the particular flow are decremented by an amount equal to the certain length when that certain cell is transferred on the network link.

55. The method of Claim 52 wherein each cell is transmitted over a series of clock cycles and wherein at least a certain cell transmitted on the network link, including data associated with a particular one of the specific flows, includes a certain length that is less than the predetermined maximum cell length and wherein the credits available to the particular flow are decremented on each clock cycle by an amount that is equal to a maximum amount of data that is transferable on each clock cycle.

56. The method of Claim 44 wherein each specific flow incrementally transfers data on the network link in intermittent bursts such that, at some points in time, no data associated with certain ones of the specific flows is present on the network link and, at other points in time, data associated with the certain ones of the specific flows is present on the network link and wherein the actual use of the available network bandwidth is controlled, at least in part, by using a burst fraction allocation which limits bandwidth allocatable to the certain ones of the specific flows only at times when data associated with the those flows is present on the network link.

57. The method of Claim 44 wherein the actual use of the network link is controlled by allocating credits to each of the specific flows and, before any one flow is given the use of the network link in transferring a particular cell having a particular length, using accumulated ones of the credits to verify that the data flow requesting to transfer the particular cell has accumulated credits equal to at least a certain allocation.

58. The method of Claim 57 wherein at least two different types of credits are allocated to each data flow for use in an assignment of priority to each data flow.

59. The method of Claim 58 wherein a sustained credit allocation, serving as a first type of said credits, is used to control at least for the specific ones of the data flows the minimum bandwidth allocation on the network link specified for the specific data flows.

60. The method of Claim 59 wherein at least a particular one of the specific data flows intermittently bursts data on the network link such that the sustained credit allocation for that particular data flow is exceeded on the network link during bursting and such that the sustained credit allocation controls average utilization of the network link over time for that particular data flow by balancing at least one burst time period during which the particular data flow bursts data on the network link with at least one quiescent time period during which the particular data flow does not utilize the network link.

61. The method of Claim 60 wherein a burst credit allocation, serving as a second type of said credits, is used to establish at least for the specific ones of the data flows a maximum utilization of the network link during bursting.

62. The method of Claim 61 wherein more than one of the specific flows is assigned sustained priority at one point in time such that a group of the specific flows is tied at the sustained priority and wherein the tie is broken in a way which results in alternate transmission of the group of flows.

63. The method of Claim 61 wherein burst priority is assigned to each of the specific flows with accumulated burst credits greater than a predetermined number.

64. The method of Claim 58 wherein the assignment of priority to each of the specific flows is based upon an accumulated number of each different type of credit such that, at a point in time, a group of flows is tied in priority based on having one or both of the different types of credits and the tie is broken in a way which results in alternate transmission of each specific flow within the group of tied flows.

65. The method of Claim 64 wherein a group of pointers is provided having at least one pointer associated with each type of credit and wherein the tie is broken by using a specific one of the pointers selected from the group of pointers such that the specific pointer is associated with the type or types of credits held by the group of flows that are tied in priority.

66. The method of Claim 65 wherein the specific flows are characterized by bursting data over the network link such that, at some points in time, data for a particular one of the specific flows is not present on the network link and, at other points in time data for the particular one of the specific flows is present on the network link and wherein the two different types of credits include at least sustained credits for use in controlling average use over time of the network link by the specific flows and burst credits for use in controlling actual use of the network link by the specific flows during bursting and the group of flow pointers includes at least a sustained flow pointer associated with the sustained credits and a burst flow pointer associated with the burst credits.

67. The method of Claim 64 wherein another group of the specific flows is tied at a priority based on having less than a predetermined amount of both types of credits and wherein the no credit tie is broken in a way which results in alternate transmission of the no credit priority tied group of specific flows when the network link would otherwise be idle.

68. The method of Claim 67 wherein the no credit tie is broken by using a no credit flow pointer selected from a group of flow pointers including other flow pointers that are associated with the different types of credits.

69. A network arrangement configured for transferring a plurality of data flows on a network link, each of which data flows requires a minimum bandwidth allocation on the network link, said network arrangement comprising:

a processing arrangement at a first end of the network link configured for receiving the data flows and for dividing the data flows into cells, each of which cells includes a length that does not exceed a maximum cell length and at least some of which cells include a length that is less than said maximum cell length;

a transfer arrangement at the first end of the network link for transferring the cells on said network link, including cells having lengths less than said maximum cell length, such that cells of different data flows are multiplexed and the minimum bandwidth allocation of each data flow is maintained on the network link; and

a second, receiving end of the network link which receives the cells and, thereafter, combines the cells so as to reconstruct the plurality of data flows at the second end of the network link.

70. The network arrangement of Claim 69 further comprising:

a credit allocation arrangement forming part of the processing arrangement for allocating credits to each flow in a predetermined way associated with the transfer of the credit flows and for tracking an accumulated number of credits associated with each one of the data flows; and

an arbitration configuration forming another part of the processing arrangement which uses accumulated ones of the credits to verify that, before any one data flow is given the use of the network link in transferring a particular cell having a particular length, the data flow requesting to transfer the particular cell has accumulated credits equal to at least a certain allocation.

71. The network arrangement of Claim 70 wherein the particular length of the particular cell is less than the maximum cell length and the certain allocation is sufficient to transfer the maximum cell length.

72. The network arrangement of Claim 70 wherein the particular length of the particular cell is less than the maximum cell length and the certain allocation is sufficient to transfer the particular cell length, but is insufficient to transfer the maximum cell length.

73. The network arrangement of Claim 70 wherein the credit allocation arrangement includes a series of credit registers, one of which is associated with each flow, for storing accumulated credits associated with each of the flows such that upon transfer of the particular cell which includes a length that is less than the maximum cell length, the accumulated number of credits available to the particular flow in an associated one of the credit registers is decremented by an amount of credits corresponding to no more than the length of the particular cell.

74. The network arrangement of Claim 73 wherein the cell length is made up of a number of one or more words and wherein the total number of credits available to the particular flow is decremented by the number of words which make up the length of the particular cell in allowing the particular flow to use the network link.

75. The network arrangement of Claim 73 wherein the allocation arrangement is configured for allocating at least two different types of credits to each data flow and each credit register is configured having a number of fields for storing and accumulating each of the different types of credits for use in an assignment of priority to each data flow.

76. The network arrangement of Claim 75 wherein each cell is made up of one or more words having a predetermined length and wherein the credit allocation arrangement is designed to allocate credits to each flow incrementally as a fraction of one word.

77. The network arrangement of Claim 76 wherein the network link operates over a series of clock cycles and wherein the credits are incrementally allocated to each flow on each clock cycle.

78. The network arrangement of Claim 75 wherein the credit allocation arrangement allocates sustained credits as one of the two different types of credits and wherein each of the credit registers is configured with a sustained credit field for accumulating and tracking a stored number of sustained credits for use by the arbitration configuration in controlling, at least for selected ones of the data flows, the minimum bandwidth allocations on the network link specified for the selected data flows.

79. The network arrangement of Claim 78 wherein the credit allocation arrangement allocates burst credits as the other one of the two different types of credits and wherein a second one of the credit registers is configured with a burst credit field for accumulating and tracking a stored number of burst credits for use by the arbitration configuration in controlling, at least for selected ones of the data flows, a maximum utilization of the network link during bursting.

80. The network arrangement of Claim 79 wherein the network link transfers a selected cell over a series of clock cycles based on a prior priority determination and wherein the arbitration configuration includes a current selected flow register identifying a candidate cell to next be transferred on the network link.

81. The network arrangement of Claim 80 wherein the arbitration configuration is configured for updating the current selected flow register and thereby the candidate cell on each clock cycle based, at least in part, on the accumulated number of sustained credits and the accumulated number of burst credits.

82. The network arrangement of Claim 80 wherein the credit allocation arrangement is configured to increment the accumulated number of sustained and burst credits available to each flow on each clock cycle such that the identified candidate cell is continuously re-selected based upon changing numbers of accumulated sustained and burst credits during transfer of the selected cell.

83. The network arrangement of Claim 82 wherein the arbitration configuration is configured for assigning a sustained priority to any flow having an accumulated number of sustained credits greater than at least a first predetermined number and for assigning a burst priority to any flow having an accumulated number of burst credits greater than at least a second predetermined number such that a group of flows is tied at a particular one of the sustained or burst priorities and wherein the arbitration configuration includes a group of flow pointers having at least a sustained flow pointer associated with the sustained priority and a burst flow pointer associated with the burst priority such that one of the flow pointers is selected as being associated with the particular one of the priorities for use in breaking the tie at the particular priority.

84. The network arrangement of Claim 83 wherein said transfer arrangement cooperates with the arbitration configuration such that transfer of the candidate cell of the current selected flow is initiated on the network link immediately upon completion of the transfer of the selected cell without further decision as to selection of the candidate cell.

85. The network arrangement of Claim 84 wherein the candidate cell forms a portion of a packet received on the current selected flow and wherein the arbitration arrangement is configured for updating the current selected flow on each clock cycle based, at least in part, on a packet priority which identifies a relative position of the candidate cell within the packet.

86. The network arrangement of Claim 85 wherein the arbitration arrangement identifies the relative position of the candidate cell as a start of packet, a middle of packet or an end of packet.

87. A network arrangement configured for transferring packets originating from a plurality of data flows on a network link, said network arrangement comprising:

a processing arrangement at a first end of the network link configured for receiving the data flows and packets therein and for dividing specific ones of the packets, having a length that is greater than a minimum cell length, into a series of cells such that each of the cells is identifiable as forming a part of one of the specific packets;

a transfer arrangement at the first end of the network link for transferring at least the specific ones of the packets on said link incrementally by alternating each series of cells; and

a receiving arrangement at a second, receiving end of the network link which receives the cells and identifies the specific packets from which each cell originated and then uses that identification to reconstruct the series of cells for each specific packet and, thereby, the specific packets at the second end of the network link.

88. The network arrangement of Claim 87 wherein said transfer arrangement multiplexes cells from different ones of the specific packets on the network link and said receiving arrangement demultiplexes the cells using the identification.

89. The network arrangement of Claim 87 wherein said transfer arrangement transmits the cells on the network link based, at least in part, on identification of the specific packet from which a particular cell originated.

90. The network arrangement of Claim 89 wherein the transfer arrangement is configured to such that a particular one of the cells is selected to reduce the latency of transfer of the specific packet which the particular cell identifies after having transmitted at least one other cell which identifies that packet.

91. The network arrangement of Claim 87 wherein each cell forming part of a particular one of the specific packets is identified as a start of packet, a middle of packet or an end of packet.

92. The network arrangement of Claim 91 wherein the transfer arrangement designates the cells originating from the particular one of the specific packets using a particular one of a number of different channel identifications such that the cells from the specific packet are identified by the receiving arrangement using the particular channel identification.

93. The network arrangement of Claim 92 wherein the particular channel identification is dedicated to the particular packet for the duration of the transfer of the particular packet on the network link.

94. A network arrangement configured for transferring a plurality of data flows using a network link, each of which flows requires a priority allocation of at least a minimum bandwidth on the network link, said network link comprising:

a processing arrangement at a first end of the network link for forming cells using the data flows such that each data flow is made up of a series of one or more cells; and

a transfer arrangement for initiating the transmission of the data flows on said network link according to the priority allocation of each flow and thereafter transferring one or more cells on the network link which cells form portions of the data flows and configured for cooperating with the processing arrangement such that, at a time when the use of the network link is not allocatable to any of the data flows consistent with the priority allocations of the data flows, transferring at least one additional cell forming part of one of the flows irrespective of the priority allocation assigned to that flow.

95. The network arrangement of Claim 94 wherein said processing arrangement defines said minimum bandwidth for each data flow using an accumulated number of credits assigned to each flow such that a request for the use of the network link by a particular one of the flows is granted when the particular flow has more than a predetermined minimum number of accumulated credits and, when the particular flow requests the use of the network link with less than the predetermined minimum number of accumulated credits, maintaining the accumulated number of credits assigned to the particular flow when the use of the network link is granted to the particular flow, irrespective of the priority allocation assigned to the particular flow.

96. The network arrangement of Claim 94 wherein a plurality of data flows request the use of the network link at a time when the requesting flows have no allocatable priority consistent with their priority allocations for the use of the network link and wherein the processing arrangement chooses the one additional cell which forms part of one of the requesting flows in a predetermined way.

97. The network arrangement of Claim 96 wherein the processing arrangement includes a no credit round robin register and the additional cell is selected in said predetermined way from a particular one of the requesting flows using the no credit round robin which alternately points to one of the data flows.

98. The network arrangement of Claim 97 wherein the requesting flows have assigned flow identifications and wherein the processing arrangement chooses the particular one of the requesting flows based, at least in part, on the assigned flow identifications of the requesting flows in relation to the one of the flows which is currently pointed to by the no credit round robin.

99. A network arrangement configured for using a network link having an available bandwidth which varies over time and which is configured for incrementally transferring a plurality of data flows, for at least specific ones of which preferred minimum bandwidth allocations are specified, said network arrangement comprising:

a processing arrangement which receives the data flows at a first end of the network link for transfer thereon to a second end of the network link and which, based on the preferred minimum bandwidth allocations, establishes sustained percentages of the available network link bandwidth to be assigned, respectively, to each of the specific data flows for use during incremental transfer of the specific data flows on the network link; and

a transfer arrangement which, responsive to variation, over time, of the available network link bandwidth, cooperates with the processing arrangement to control actual use of the available network bandwidth by the specific data flows during the incremental transfer of the specific data flows across the network link such that each of the specific data flows is allocated approximately its respective sustained percentage of the available network bandwidth over time.

100. The network arrangement of Claim 99 wherein said transfer arrangement transfers each of the specific data flows over a series of clock cycles and wherein said sustained percentages are specified as sustained fraction allocations that are credited to each of the specific flows by the processing arrangement on each clock cycle such that, during each clock cycle, each of the specific flows is potentially maintainable at a rate corresponding to its sustained fraction allocation.

101. The network arrangement of Claim 100 wherein each specific flow is incrementally transferred on the network link by the transfer arrangement in intermittent bursts such that, at some points in time, no data associated with at least a particular one of the specific flows is present on the network link and, at other points in time, data associated with the particular flow is present on the network link and wherein said processing arrangement is configured to use a burst fraction allocation at least for the particular flow which limits bandwidth allocatable to the particular flow only during the other points in time when data associated with the particular flow is present on the network link such that an average allocation to the particular flow over time on the network link is approximately equal to its respective sustained fraction allocation.

102. The network arrangement of Claim 100 wherein said flows are transferred on the network link using cells which vary in length up to a predetermined maximum length and which contain data associated with the flows and wherein the sustained fraction allocations credited by the processing arrangement to the flows on each clock cycle are each equal to a fractional portion of the predetermined maximum length of one of the cells.

103. The network arrangement of Claim 102 wherein each cell is made up of words having a length in words and wherein the sustained fraction allocations are each specified as being less than the length in words such that the processing arrangement credits no more than one word to each flow on one of the clock cycles.

104. The network arrangement of Claim 103 wherein each of the sustained fraction allocations is specifiable as being less than one percent of the length in words.

105. The network arrangement of Claim 102 wherein the processing arrangement is configured for tracking sustained allocation credits available to each of the specific flows such that the available sustained allocation credits for each of the specific flows accumulates in increments which are less than the maximum predetermined length of one cell.

106. The network arrangement of Claim 105 wherein each cell is made up of words having a length in words and wherein the sustained fraction used by the processing arrangement for each of the specific flows is specified as being less than the length in words such that no more than one word is credited to each of the specific flows by the processing arrangement on one of the clock cycles and such that the available sustained allocation credits for each of the specific flows accumulate in increments per clock cycle which are less than the length of one word.

107. The network arrangement of Claim 105 wherein at least a certain cell transferred on the network link by the transfer arrangement, including data associated with a particular one of the specific flows, includes a certain length that is less than the predetermined maximum cell length and wherein the sustained allocation credits available to the particular flow are decremented by the processing arrangement in an amount equal to the certain length as a result of transfer of that certain cell on the network link.

108. The network arrangement of Claim 107 wherein each cell is made up of words having a length in words and wherein the sustained allocation credits available to the particular flow are decremented by the processing arrangement in the amount of a fraction of one word length.

109. The network arrangement of Claim 105 wherein each cell is transmitted by the transfer arrangement over a series of clock cycles and wherein at least a certain cell transmitted on the network link, including data associated with a particular one of the specific flows, includes a certain length that is less than the predetermined maximum cell length and wherein the sustained allocation credits available to the particular flow are decremented on each clock cycle by the processing arrangement in an amount equal to a maximum amount of data that is transferable on each clock cycle.

110. The network arrangement of Claim 99 wherein said transfer arrangement incrementally transfers data on the network link in intermittent bursts for each of the specific flows such that, at some points in time, no data associated with certain ones of the specific flows is present on the network link and, at other points in time, data associated with the certain ones of the specific flows is present on the network link and wherein the actual use of the available network bandwidth is controlled by the processing arrangement, at least in part, by using a burst fraction allocation which limits bandwidth allocatable to the certain ones of the specific flows only at times when data associated with those flows is present on the network link.

111. The network arrangement of Claim 99 wherein the processing arrangement controls the actual use of the network link by allocating credits to each of the specific flows and, before the transfer arrangement is given the use of the network link in transferring a particular cell having a particular length, the processing arrangement uses accumulated ones of the credits to verify that the data flow requesting to transfer the particular cell has accumulated credits equal to at least a certain allocation.

112. The network arrangement of Claim 111 wherein the processing arrangement allocates at least two different types of credits to each data flow for use in an assignment of priority to each data flow.

113. The network arrangement of Claim 112 wherein a sustained credit allocation, serving as a first type of said credits, is used by the processing arrangement to control at least for the specific ones of the data flows the minimum bandwidth allocation on the network link specified for the specific data flows.

114. The network arrangement of Claim 113 wherein at least a particular one of the specific data flows intermittently bursts data on the network link such that the sustained credit allocation for that particular data flow is exceeded on the network link during bursting and wherein the processing arrangement cooperates with the transfer arrangement for using the sustained credit allocation to control average utilization of the network link over time for that particular data flow by balancing at least one burst time period during which the particular data flow bursts data on the network link with at least one quiescent time period during which the particular data flow does not utilize the network link.

115. The network arrangement of Claim 114 wherein a burst credit allocation, serving as a second type of said credits, is used by the processing arrangement to establish at least for the specific ones of the data flows a maximum utilization of the network link during bursting.

116. The network arrangement of Claim 115 wherein more than one of the specific flows is assigned sustained priority at one point in time such that a group of the specific flows is tied at the sustained priority and wherein the processing arrangement is configured for breaking the tie in a way which results in alternate transmission of the group of flows.

117. The network arrangement of Claim 115 wherein the processing arrangement assigns a burst priority is to the specific flows with accumulated burst credits greater than a predetermined number.

118. The network arrangement of Claim 112 wherein the processing arrangement assigns priority to each of the specific flows based upon an accumulated number of each different type of credit such that, at a point in time, a group of flows is tied in priority based on having one or both of the different types of credits and the tie is broken in a way which results in alternate transmission of each specific flow within the group of flows.

119. The network arrangement of Claim 118 wherein the processing arrangement includes a group of pointers having a specific pointer associated with each type of credit and the processing arrangement breaks the tie by using the specific pointer selected from a group of flow pointers in which the specific pointer is associated with the type or types of credits held by the group of flows that are tied in priority.

120. The network arrangement of Claim 119 wherein the specific flows are characterized by bursting data over the network link such that, at some points in time, data for a particular one of the specific flows is not present on the network link and, at other points in time data for the particular one of the specific flows is present on the network link and wherein the two different types of credits include at least sustained credits for use in controlling average use over time of the network link by the specific flows and burst credits for use in controlling actual use of the network link by the specific flows during bursting and the group of flow pointers includes at least a sustained flow pointer associated with the sustained credits and a burst flow pointer associated with the burst credits.

121. The network arrangement of Claim 118 wherein another group of the specific flows is tied at a no credit priority based on having less than a predetermined amount of both types of credits and wherein said processing arrangement breaks the no credit priority tie in a way which results in alternate transmission of the no credit priority tied group of specific flows when the network link would otherwise be idle.

122. The network arrangement of Claim 121 wherein the processing arrangement includes a group of priority pointers at least one of which is associated with each different type of priority and having a no credit priority pointer and the processing arrangement selects the no credit priority pointer for use in breaking the no credit priority tie.

123. In a network link configured for transferring a plurality of data flows, each of which requires a minimum bandwidth allocation on the network link and which flows include bursts having assigned priorities, a method for transferring said data flows on the network link, said method comprising the steps of:

receiving at least one burst as part of one of said flows at a first end of the network link with a particular assigned priority; and

transferring the burst to the second end of the network link with a new priority that is higher than the particular assigned priority.

124. The method of Claim 123 wherein said new priority is based on the use of at least two different types of priority allocation credits.

125. In a network link configured for transferring a plurality of data flows, each of which requires a minimum bandwidth allocation on the network link and which flows include bursts having assigned priorities, an arrangement for transferring said data flows on the network link, said arrangement comprising:

a first configuration which receives at least one burst as part of one of said flows at a first end of the network link with a particular assigned priority; and

a second configuration for transferring the burst to the second end of the network link with a new priority that is higher than the particular assigned priority.

126. The arrangement of Claim 125 wherein said new priority is based on the use of at least two different types of priority allocation credits.

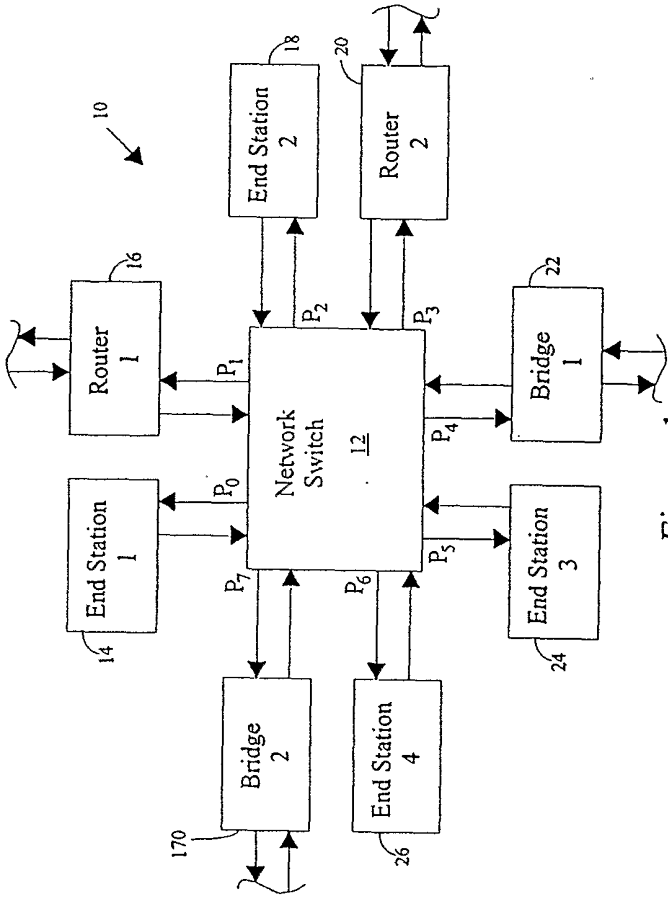

10
End Station 2
18
Router 2
20
Router 1
16
Bridge 1
22
Network Switch
12
P_0
P_1
P_2
P_3
P_4
P_5
P_6
P_7
End Station 1
14
End Station 3
24
Bridge 2
170
End Station 4
26

Figure 1

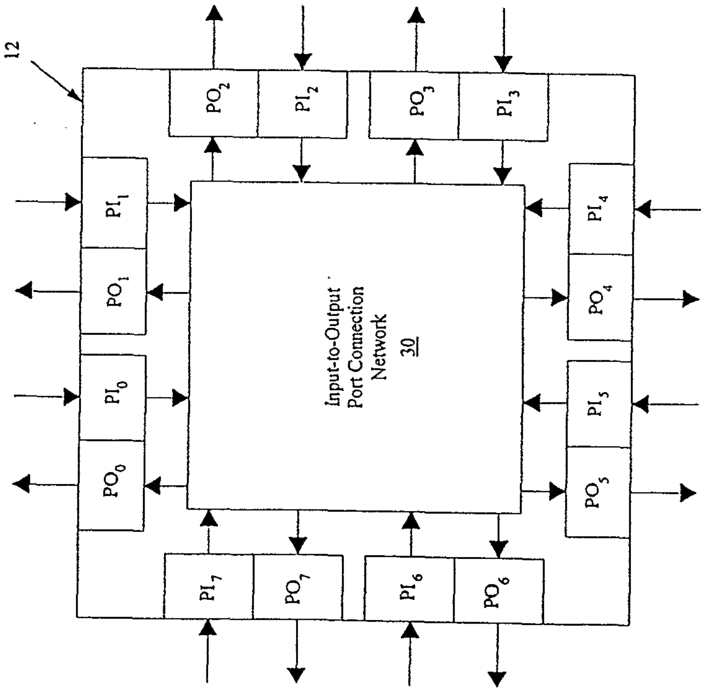

12
PO$_2$
PI$_2$
PO$_3$
PI$_3$
PI$_1$
PO$_1$
PI$_0$
PO$_0$
PI$_4$
PO$_4$
PI$_5$
PO$_5$
PI$_7$
PO$_7$
PI$_6$
PO$_6$
Input-to-Output Port Connection Network 30

Figure 2

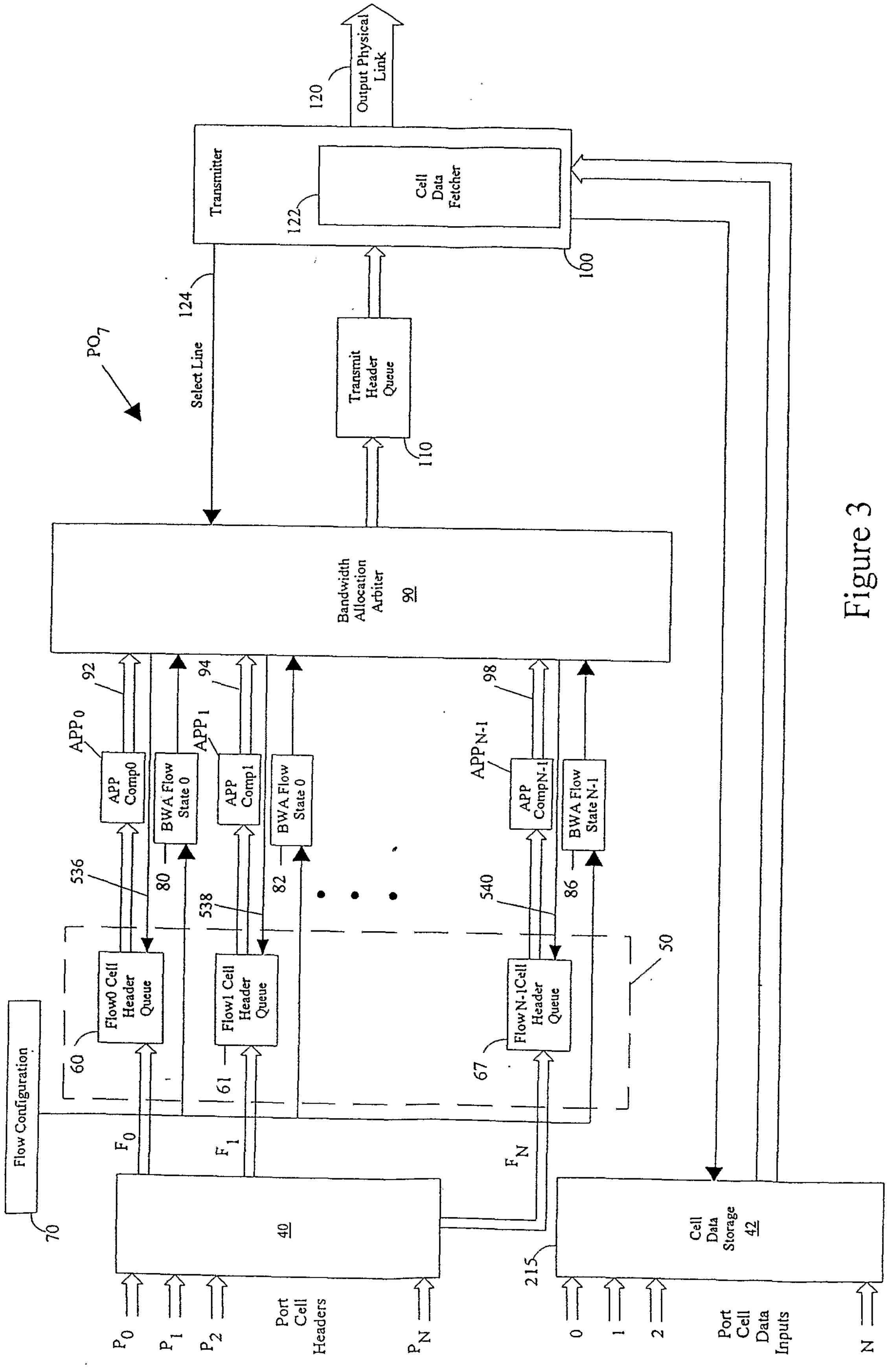

Output Physical Link
120
Transmitter
Cell Data Fetcher
122
100
124
Select Line
PO7
Transmit Header Queue
110
Bandwidth Allocation Arbiter 90
92
94
98
APP0
APP1
APPN-1
APP Comp0
APP Comp1
APP CompN-1
BWA Flow State 0
BWA Flow State 0
BWA Flow State N-1
80
82
86
536
538
540
Flow0 Cell Header Queue
Flow1 Cell Header Queue
FlowN-1Cell Header Queue
60
61
67
50
Flow Configuration
70
F0
F1
FN
40
Cell Data Storage 42
215
P0
P1
P2
Port Cell Headers
PN
0
1
2
Port Cell Data Inputs
N

Figure 3

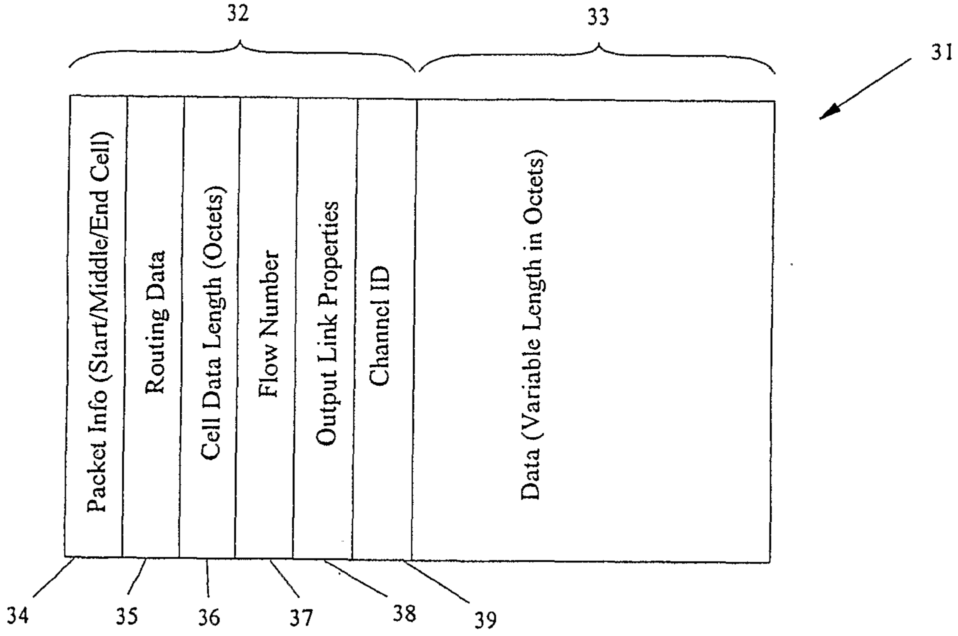

32
33
31
Packet Info (Start/Middle/End Cell)
Routing Data
Cell Data Length (Octets)
Flow Number
Output Link Properties
Channel ID
Data (Variable Length in Octets)
34
35
36
37
38
39

Figure 3a

Flow Configuration Parameters

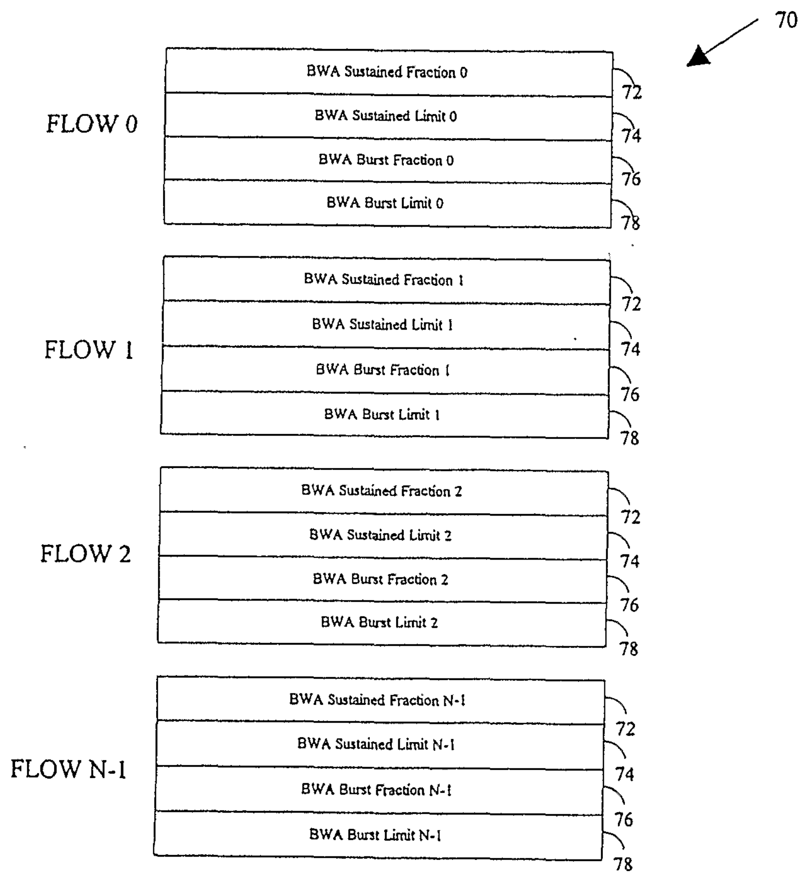


Figure 4

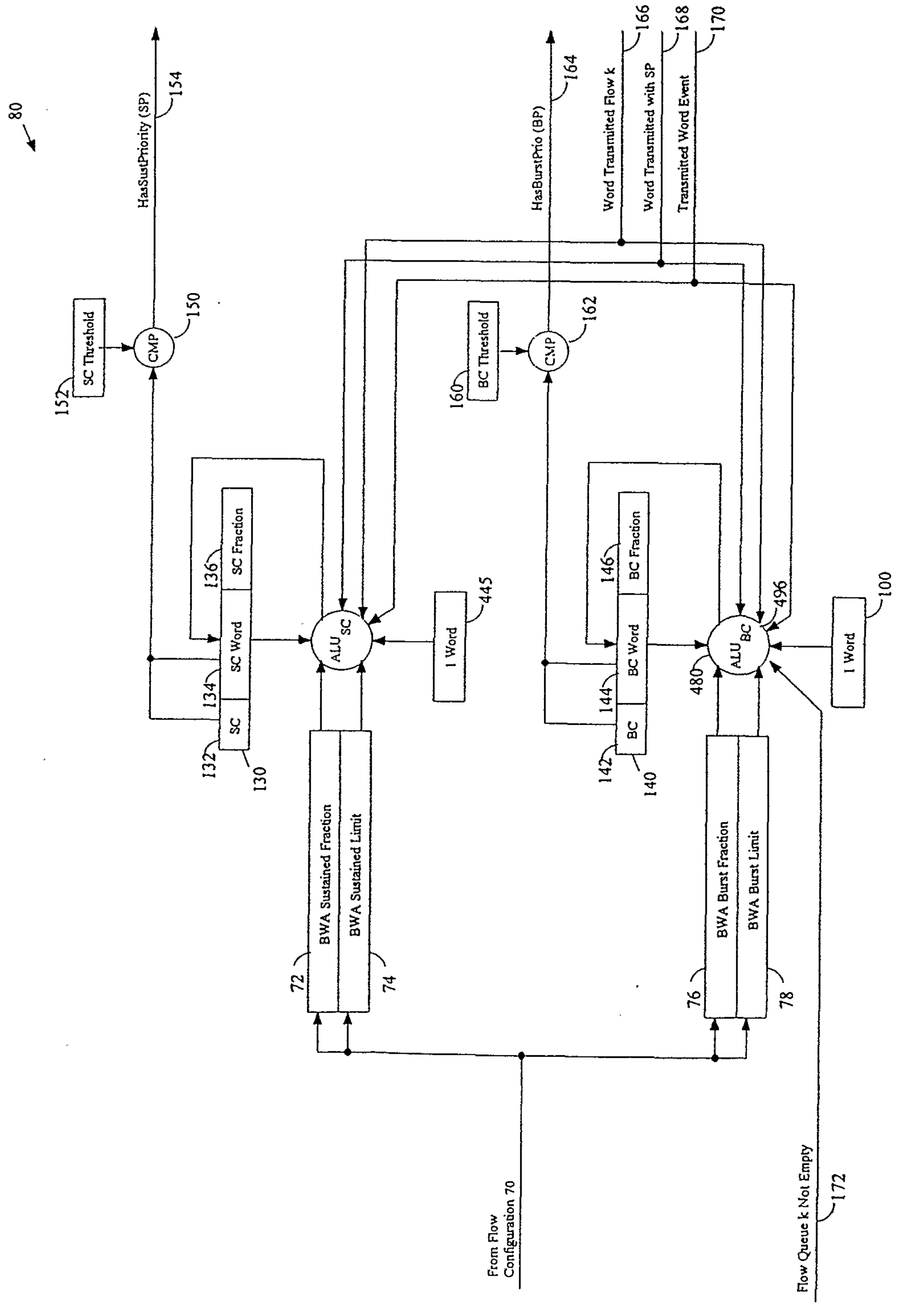

80
HasSustPriority (SP)
154
SC Threshold
152
CMP
150
SC Fraction
136
SC Word
134
SC
132
130
ALU SC
1 Word
445
BWA Sustained Fraction
BWA Sustained Limit
72
74
HasBurstPrio (BP)
164
Word Transmitted Flow k
166
Word Transmitted with SP
168
Transmitted Word Event
170
BC Threshold
160
CMP
162
BC Fraction
146
BC Word
144
BC
142
140
ALU BC
480
496
1 Word
100
BWA Burst Fraction
BWA Burst Limit
76
78
From Flow Configuration 70
Flow Queue k Not Empty
172

Figure 5

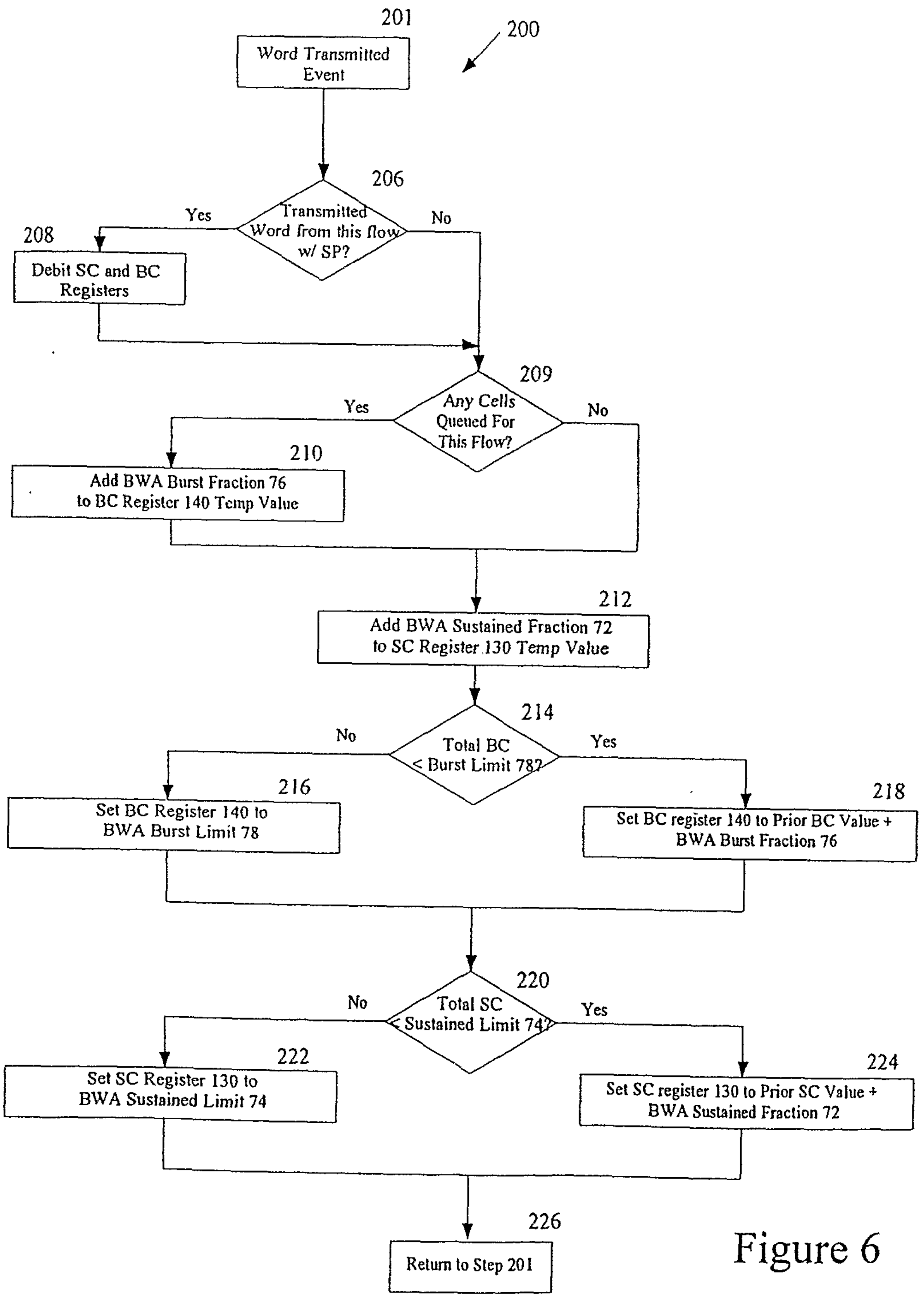

200
201
Word Transmitted Event
206
Transmitted Word from this flow w/ SP?
Yes
No
208
Debit SC and BC Registers
209
Any Cells Queued For This Flow?
Yes
No
210
Add BWA Burst Fraction 76 to BC Register 140 Temp Value
212
Add BWA Sustained Fraction 72 to SC Register 130 Temp Value
214
Total BC < Burst Limit 78?
No
Yes
216
Set BC Register 140 to BWA Burst Limit 78
218
Set BC register 140 to Prior BC Value + BWA Burst Fraction 76
220
Total SC < Sustained Limit 74?
No
Yes
222
Set SC Register 130 to BWA Sustained Limit 74
224
Set SC register 130 to Prior SC Value + BWA Sustained Fraction 72
226
Return to Step 201

Figure 6

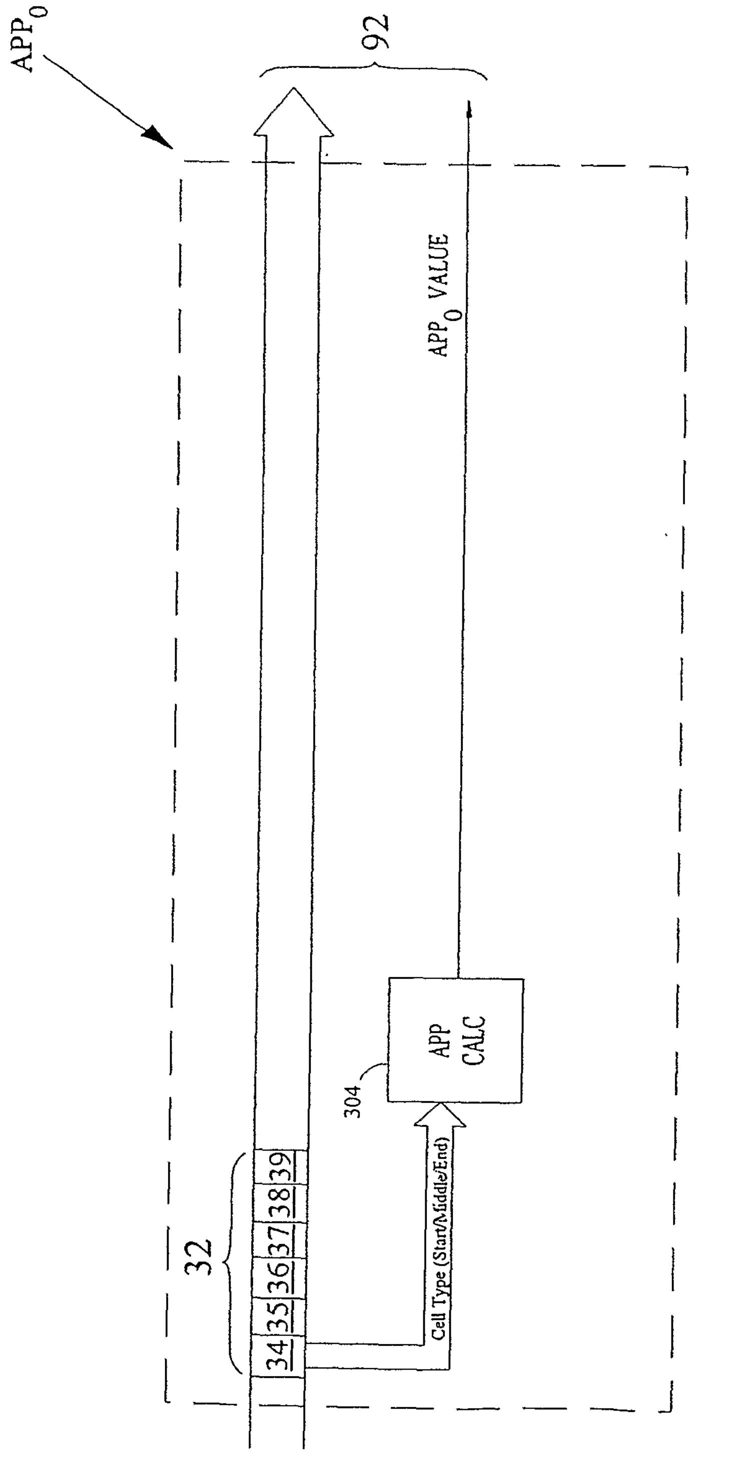

APP$_0$
92
APP$_0$ VALUE
APP CALC
304
Cell Type (Start/Middle/End)
32
34
35
36
37
38
39

Figure 7

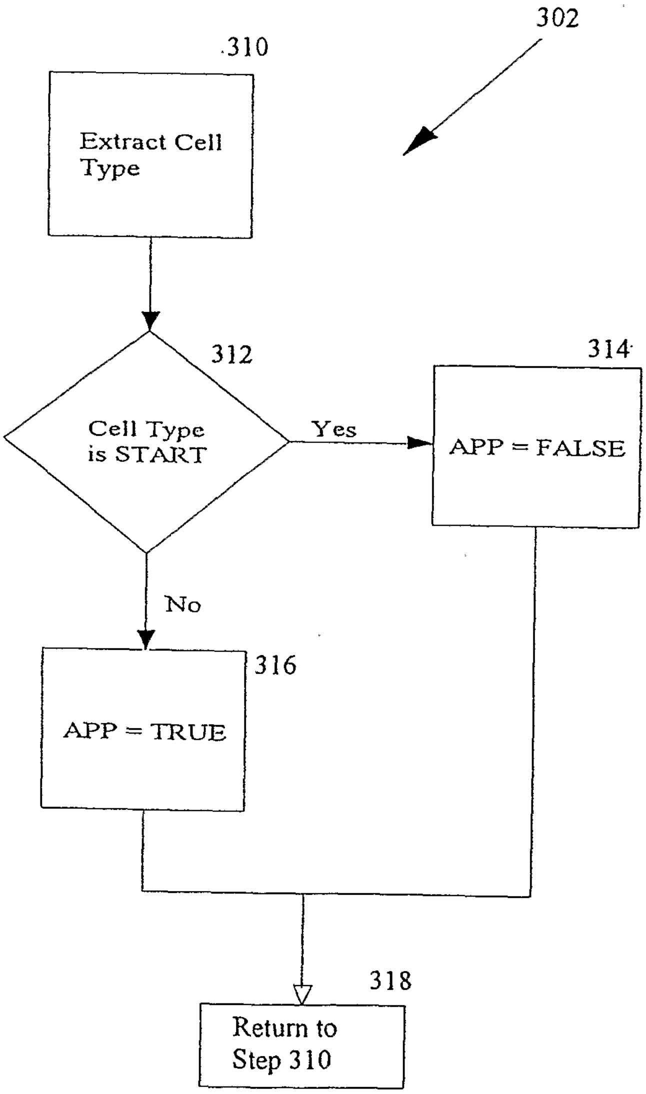

302
310
Extract Cell Type
312
Cell Type is START
Yes
314
APP = FALSE
No
316
APP = TRUE
318
Return to Step 310

Figure 8

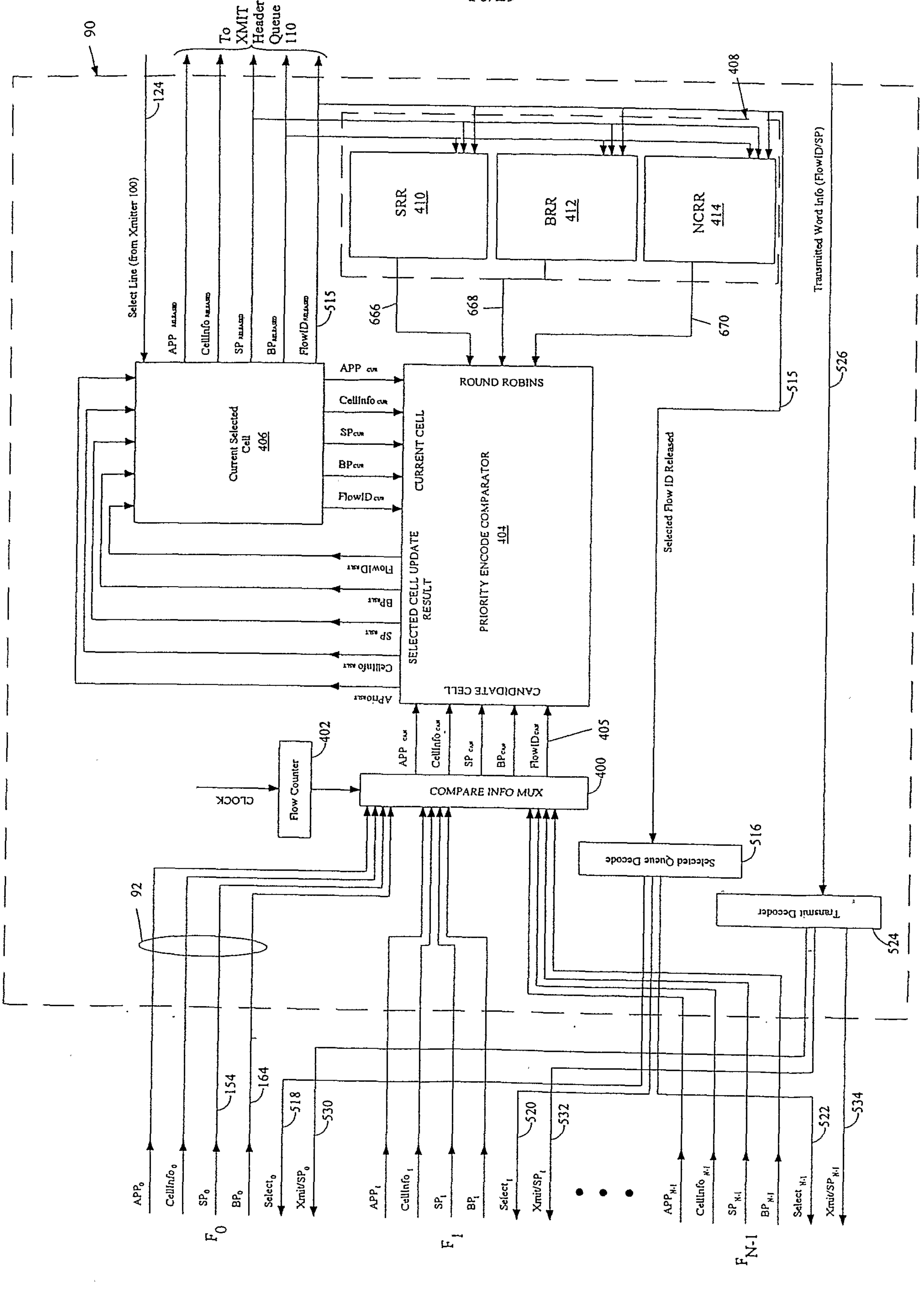

90
To XMIT Header Queue 110
124
Select Line (from Xmitter 100)
APP RELEASED
CellInfo RELEASED
SP RELEASED
BP RELEASED
FlowID RELEASED
515
408
SRR 410
BRR 412
NCRR 414
666
668
670
Transmitted Word Info (FlowID/SP)
526
515
Selected Flow ID Released
Current Selected Cell 406
APP CUR
CellInfo CUR
SP CUR
BP CUR
FlowID CUR
ROUND ROBINS
CURRENT CELL
PRIORITY ENCODE COMPARATOR 404
SELECTED CELL UPDATE RESULT
CANDIDATE CELL
FlowID RSLT
BP RSLT
SP RSLT
CellInfo RSLT
APriorRSLT
APP CAN
CellInfo CAN
SP CAN
BP CAN
FlowID CAN
405
400
COMPARE INFO MUX
Flow Counter
402
CLOCK
Selected Queue Decode
516
Transmit Decoder
524
92
154
164
518
530
520
532
522
534
APP0
CellInfo0
SP0
BP0
Select0
Xmit/SP0
F0
APP1
CellInfo1
SP1
BP1
Select1
Xmit/SP1
F1
APPN-1
CellInfo N-1
SPN-1
BPN-1
Select N-1
Xmit/SP N-1
FN-1

Figure 9

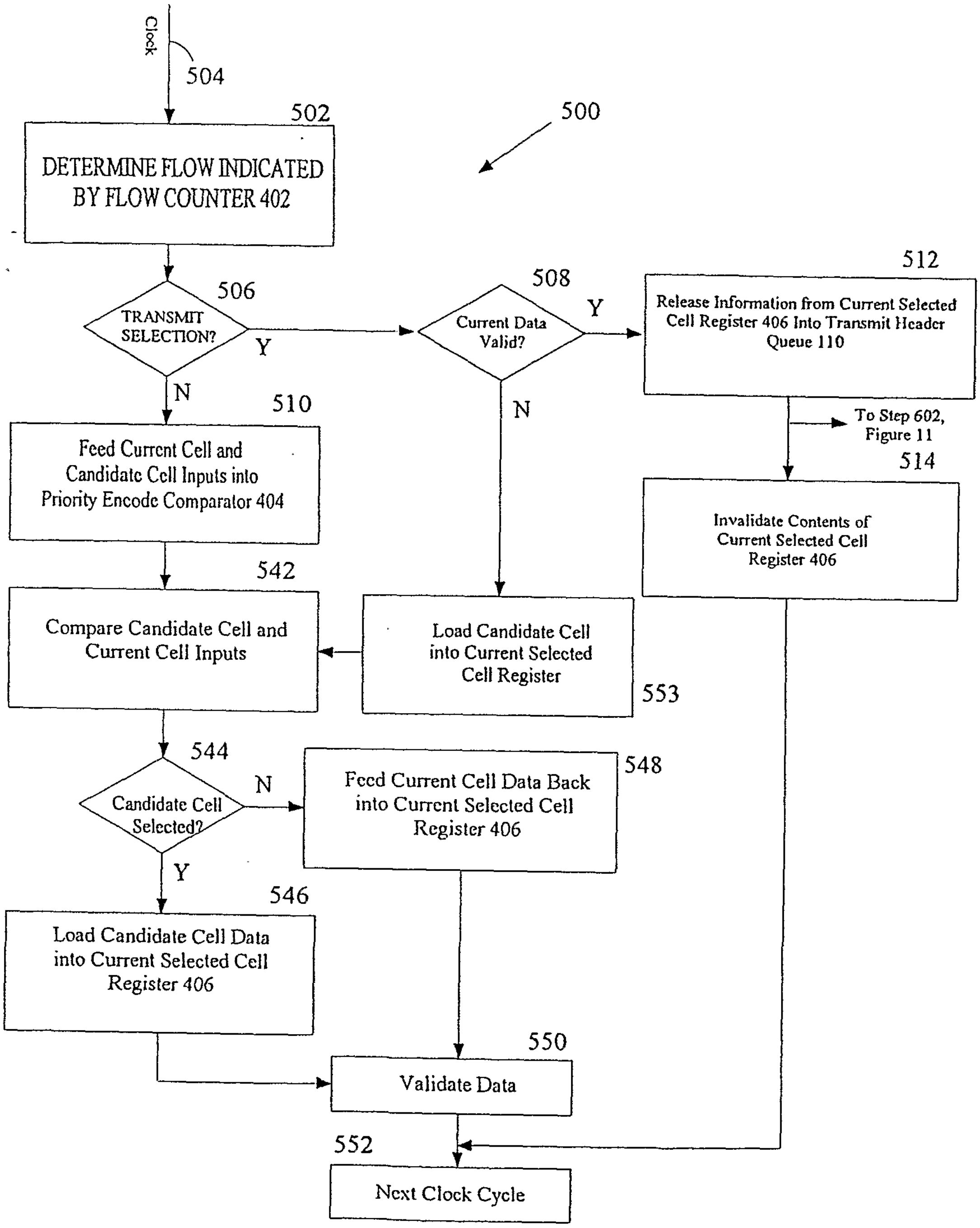

Clock
504
500
502
DETERMINE FLOW INDICATED BY FLOW COUNTER 402
506
TRANSMIT SELECTION?
Y
N
508
Current Data Valid?
Y
N
512
Release Information from Current Selected Cell Register 406 Into Transmit Header Queue 110
To Step 602, Figure 11
514
Invalidate Contents of Current Selected Cell Register 406
510
Feed Current Cell and Candidate Cell Inputs into Priority Encode Comparator 404
542
Compare Candidate Cell and Current Cell Inputs
Load Candidate Cell into Current Selected Cell Register
553
544
Candidate Cell Selected?
N
Y
548
Feed Current Cell Data Back into Current Selected Cell Register 406
546
Load Candidate Cell Data into Current Selected Cell Register 406
550
Validate Data
552
Next Clock Cycle

Figure 10

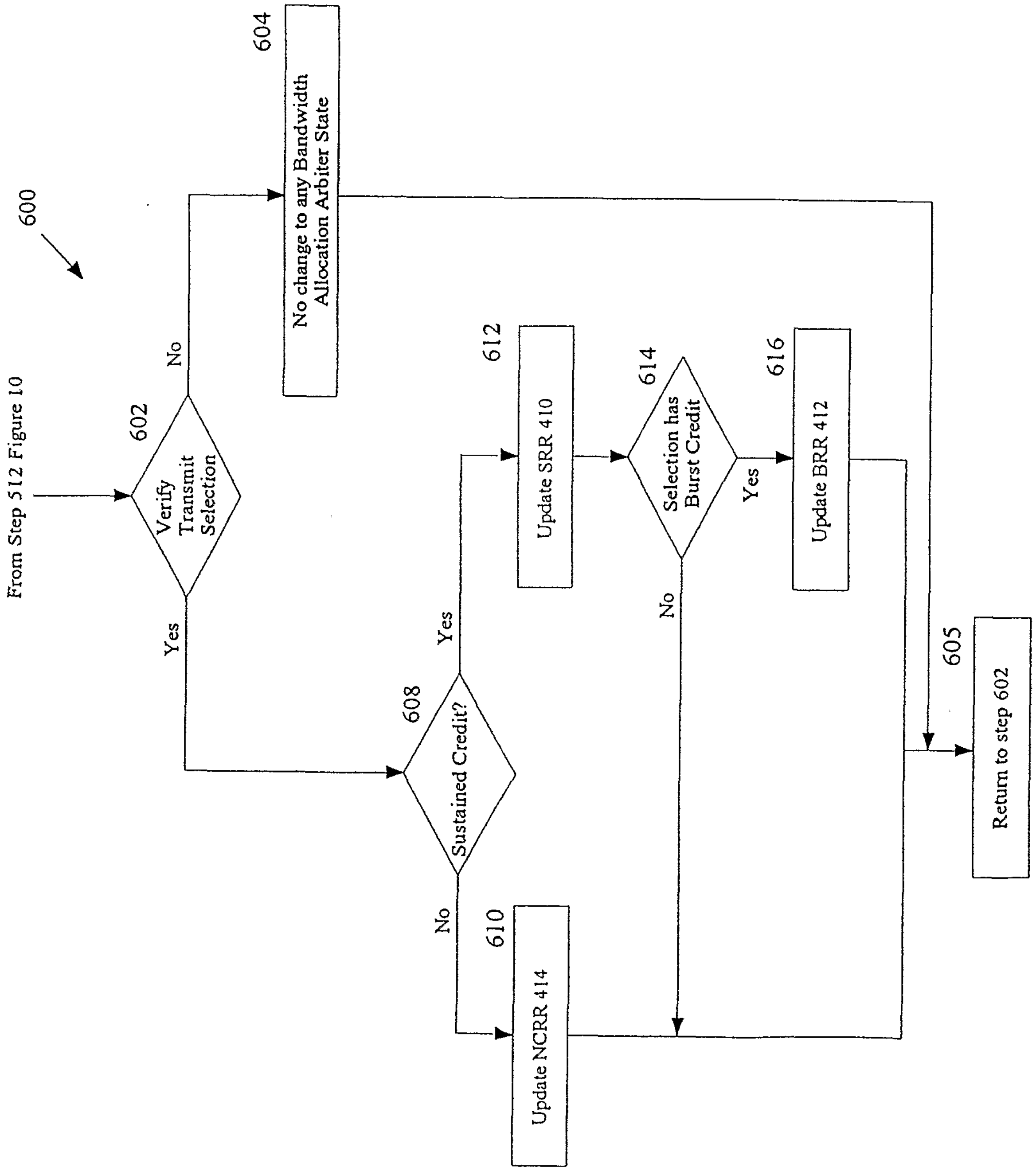

600
From Step 512 Figure 10
602
Verify Transmit Selection
No
Yes
604
No change to any Bandwidth Allocation Arbiter State
608
Sustained Credit?
Yes
No
612
Update SRR 410
610
Update NCRR 414
614
Selection has Burst Credit
Yes
No
616
Update BRR 412
605
Return to step 602

Figure 11

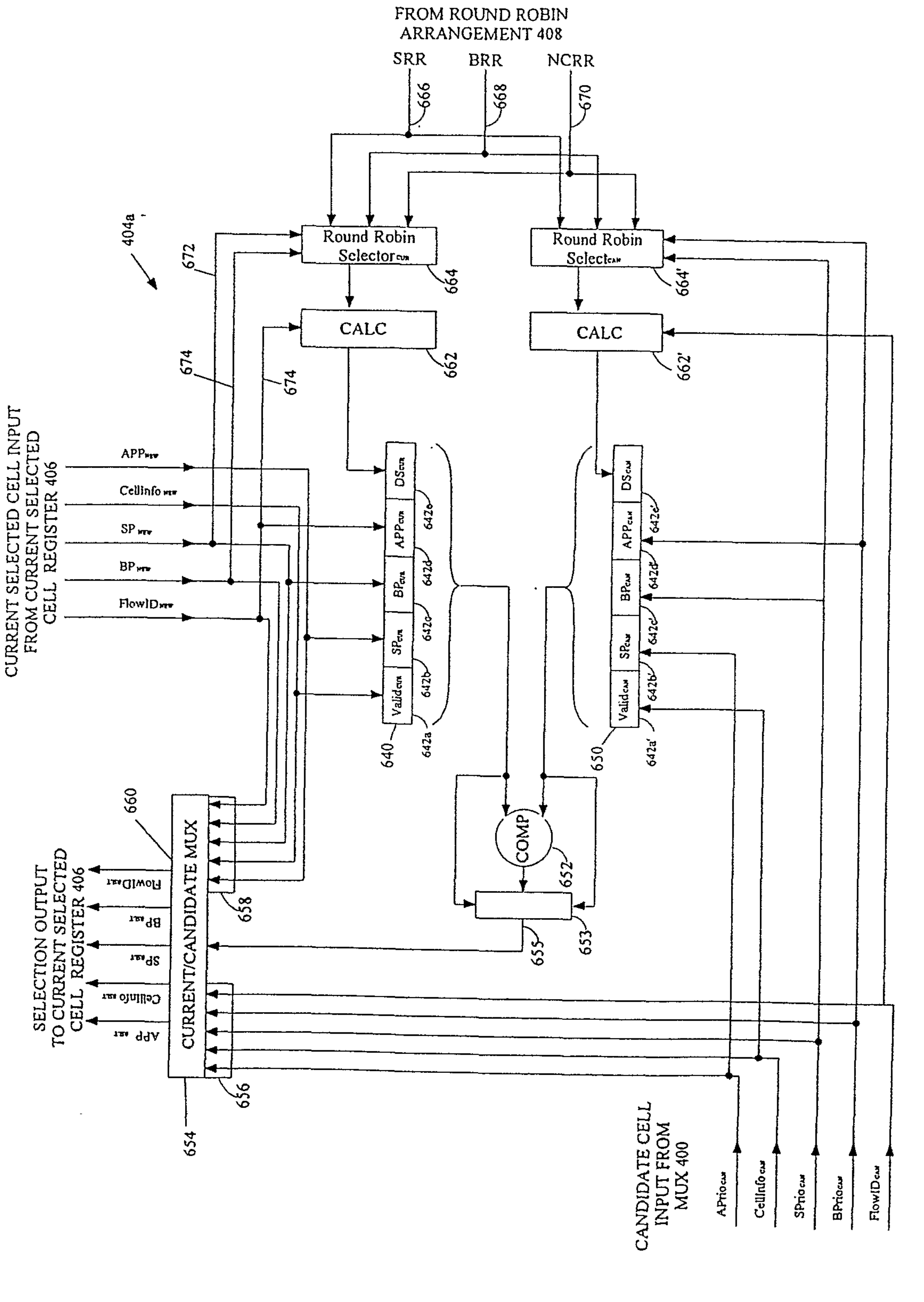

FROM ROUND ROBIN
ARRANGEMENT 408
SRR
BRR
NCRR
666
668
670
404a
672
674
674
Round Robin Selector CUR
664
Round Robin Select CAN
664'
CALC
662
CALC
662'
CURRENT SELECTED CELL INPUT FROM CURRENT SELECTED CELL REGISTER 406
APP NEW
CellInfo NEW
SP NEW
BP NEW
FlowID NEW
DS CUR
APP CUR
BP CUR
SP CUR
Valid CUR
642e
642d
642c
642b
642a
640
DS CAN
APP CAN
BP CAN
SP CAN
Valid CAN
642e'
642d'
642c'
642b'
642a'
650
COMP
652
653
655
660
CURRENT/CANDIDATE MUX
658
656
654
SELECTION OUTPUT TO CURRENT SELECTED CELL REGISTER 406
FlowID SEL
BP SEL
SP SEL
CellInfo SEL
APP SEL
CANDIDATE CELL INPUT FROM MUX 400
APrio CAN
CellInfo CAN
SPrio CAN
BPrio CAN
FlowID CAN

Figure 12

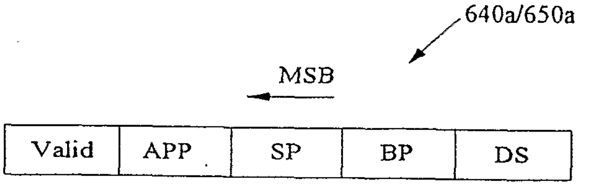


Figure 13

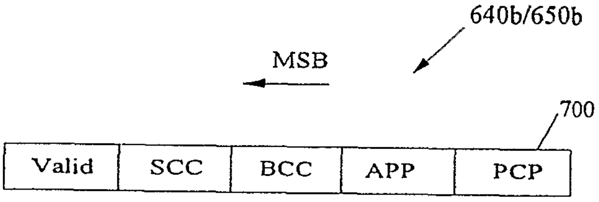


Figure 14

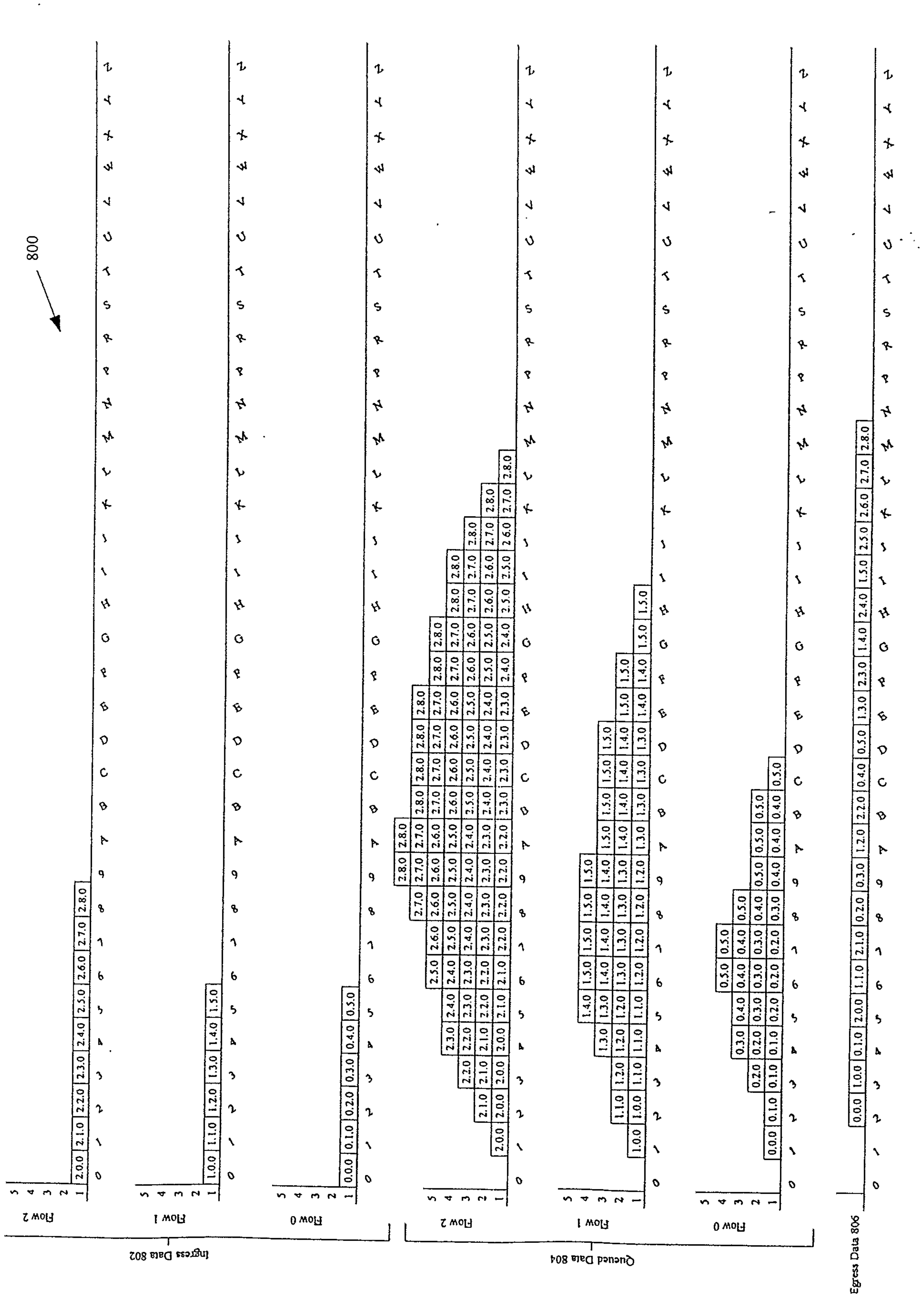

800
Ingress Data 802
Queued Data 804
Egress Data 806
Flow 0
Flow 1
Flow 2
0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N P R S T U V W X Y Z
1 2 3 4 5

FIGURE 15

Figure 16: Sustained Rate Example

	Ingress Data			Queued Data			Egress	SP Credits				
Time	Flow 0	Flow 1	Flow 2	Flow 0	Flow 1	Flow 2	Data	Flow 0	Flow 1	Flow 2	NCRR	SRR
0	0.0.0	1.0.0	2.0.0				IDLE	1.00	1.00	1.00	0	0
1	0.1.0	1.1.0	2.1.0	0.0.0	1.0.0	2.0.0	IDLE	1.00	1.00	1.00	0	0
2	0.2.0	1.2.0	2.2.0	0.1.0	1.0.0/1.1.0	2.0.0/2.1.0	0.0.0	1.50	1.25	1.25	1	0
3	0.3.0	1.3.0	2.3.0	0.1.0/0.2.0	1.1.0/1.2.0	2.0.0/2.1.0/2.2.0	1.0.0	2.00	1.50	1.50	2	0
4	0.4.0	1.4.0	2.4.0	0.2.0/0.3.0	1.1.0/1.2.0/1.3.0	2.0.0/2.1.0/2.2.0/2.3.0	0.1.0	1.50	1.75	1.75	2	1
5	0.5.0	1.5.0	2.5.0	0.2.0/0.3.0/0.4.0	1.1.0/1.2.0/1.3.0/1.4.0	2.1.0/2.2.0/2.3.0/2.4.0	2.0.0	2.00	2.00	2.00	0	1
6			2.6.0	0.2.0/0.3.0/0.4.0/0.5.0	1.2.0/1.3.0/1.4.0/1.5.0	2.1.0/2.2.0/2.3.0/2.4.0/2.5.0	1.1.0	2.50	1.25	2.25	0	2
7			2.7.0	0.2.0/0.3.0/0.4.0/0.5.0	1.2.0/1.3.0/1.4.0/1.5.0	2.2.0/2.3.0/2.4.0/2.5.0/2.6.0	2.1.0	3.00	1.50	1.50	0	0
8			2.8.0	0.3.0/0.4.0/0.5.0	1.2.0/1.3.0/1.4.0/1.5.0	2.2.0/2.3.0/2.4.0/2.5.0/2.6.0/2.7.0	0.2.0	2.50	1.75	1.75	0	1
9				0.4.0/0.5.0	1.2.0/1.3.0/1.4.0/1.5.0	2.2.0/2.3.0/2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	0.3.0	2.00	2.00	2.00	0	1
A				0.4.0/0.5.0	1.3.0/1.4.0/1.5.0	2.2.0/2.3.0/2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	1.2.0	2.50	1.25	2.25	0	2
B				0.4.0/0.5.0	1.3.0/1.4.0/1.5.0	2.3.0/2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	2.2.0	3.00	1.50	1.50	0	0
C				0.5.0	1.3.0/1.4.0/1.5.0	2.3.0/2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	0.4.0	2.50	1.75	1.75	0	1
D					1.3.0/1.4.0/1.5.0	2.3.0/2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	0.5.0	2.00	2.00	2.00	0	1
E					1.4.0/1.5.0	2.3.0/2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	1.3.0	2.50	1.25	2.25	0	2
F					1.4.0/1.5.0	2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	2.3.0	3.00	1.50	1.50	0	0
G					1.5.0	2.4.0/2.5.0/2.6.0/2.7.0/2.8.0	1.4.0	3.50	1.75	1.75	2	0
H					1.5.0	2.5.0/2.6.0/2.7.0/2.8.0	2.4.0	4.00	2.00	2.00	0	0
I						2.5.0/2.6.0/2.7.0/2.8.0	1.5.0	4.50	1.25	2.25	0	2
J						2.6.0/2.7.0/2.8.0	2.5.0	5.00	1.50	1.50	0	0
K						2.7.0/2.8.0	2.6.0	5.50	1.75	1.75	0	0
L						2.8.0	2.7.0	6.00	2.00	2.00	0	0
M							2.8.0	6.50	2.25	1.25	0	0
N							IDLE	6.50	2.25	1.25	0	0
P							IDLE	6.50	2.25	1.25	0	0
R							IDLE	6.50	2.25	1.25	0	0
S							IDLE	6.50	2.25	1.25	0	0
T							IDLE	6.50	2.25	1.25	0	0
U							IDLE	6.50	2.25	1.25	0	0
V							IDLE	6.50	2.25	1.25	0	0
W							IDLE	6.50	2.25	1.25	0	0
X							IDLE	6.50	2.25	1.25	0	0
Y							IDLE	6.50	2.25	1.25	0	0
Z							IDLE	6.50	2.25	1.25	0	0

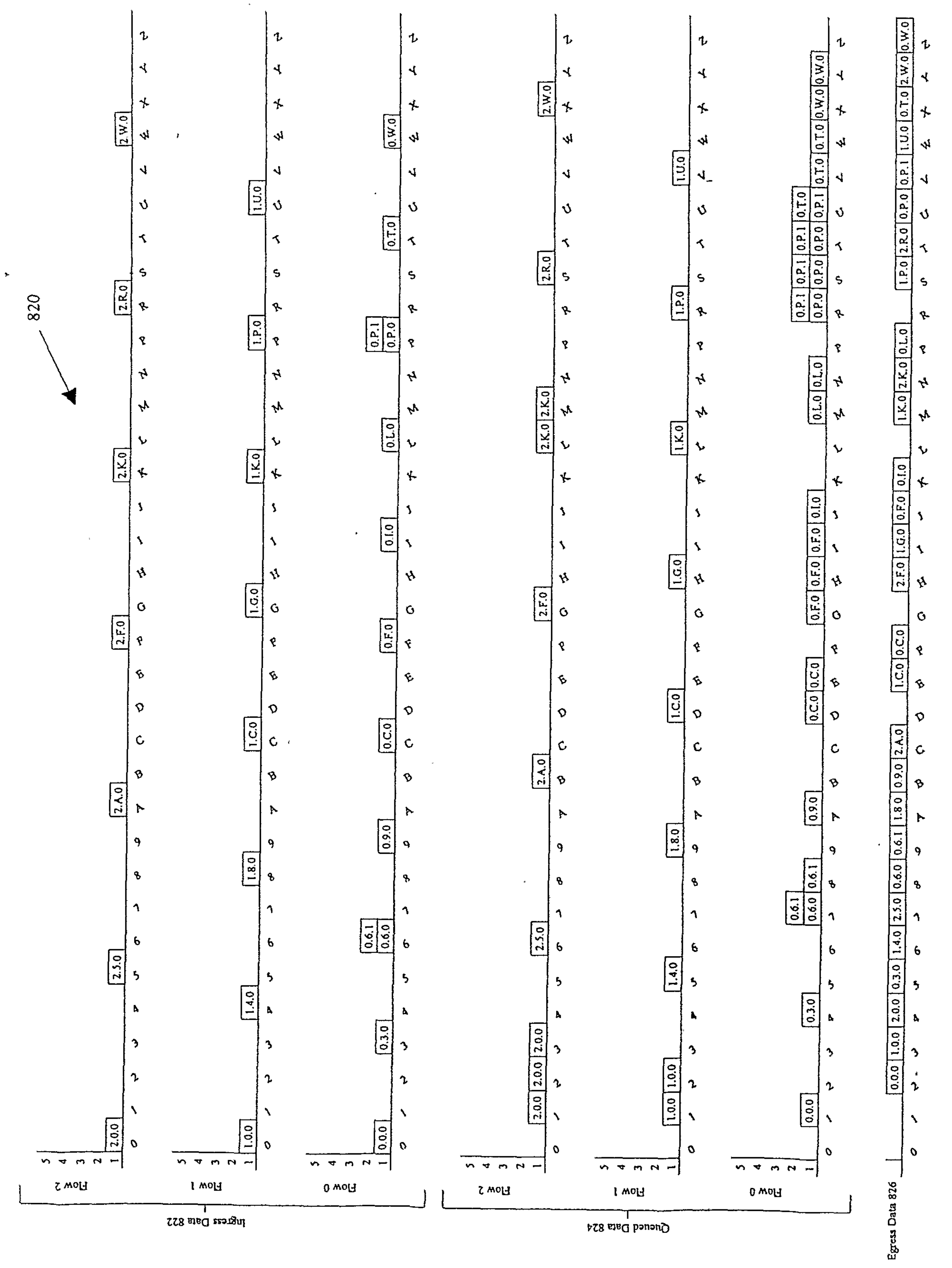

820
Ingress Data 822
Queued Data 824
Egress Data 826
Flow 2
Flow 1
Flow 0
1 2 3 4 5
0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N P R S T U V W X Y Z
2.0.0
2.5.0
2.A.0
2.F.0
2.K.0
2.R.0
2.W.0
1.0.0
1.4.0
1.8.0
1.C.0
1.G.0
1.K.0
1.P.0
1.U.0
0.0.0
0.3.0
0.6.0
0.6.1
0.9.0
0.C.0
0.F.0
0.I.0
0.L.0
0.P.0
0.P.1
0.T.0
0.W.0

FIGURE 17

Figure 18: Available Unused Bandwidth

	Ingress Data			Queued Data			Egress	SP Credits				
Time	Flow 0	Flow 1	Flow 2	Flow 0	Flow 1	Flow 2	Data	Flow 0	Flow 1	Flow 2	NCRR	SRR
0	0.0.0	1.0.0	2.0.0				IDLE	1.00	1.00	1.00	0	0
1				0.0.0	1.0.0	2.0.0	IDLE	1.00	1.00	1.00	0	0
2					1.0.0	2.0.0	0.0.0	1.33	1.25	1.20	1	0
3	0.3.0					2.0.0	1.0.0	1.67	1.50	1.40	2	0
4		1.4.0		0.3.0			2.0.0	2.00	1.75	1.60	0	0
5			2.5.0		1.4.0		0.3.0	1.33	2.00	1.80	0	1
6	0.6.0/0.6.1					2.5.0	1.4.0	1.67	1.25	2.00	0	2
7				0.6.0/0.6.1			2.5.0	2.00	1.50	1.20	0	0
8		1.8.0		0.6.1			0.6.0	1.33	1.75	1.40	0	1
9	0.9.0				1.8.0		0.6.1	1.67	2.00	1.60	1	1
A			2.A.0	0.9.0			1.8.0	2.00	1.25	1.80	1	2
B						2.A.0	0.9.0	1.33	1.50	2.00	1	1
C	0.C.0	1.C.0					2.A.0	1.67	1.75	1.20	1	0
D				0.C.0	1.C.0		IDLE	1.67	1.75	1.20	1	0
E				0.C.0			1.C.0	2.00	2.00	1.40	2	0
F	0.F.0		2.F.0				0.C.0	1.33	2.25	1.60	2	1
G		1.G.0		0.F.0		2.F.0	IDLE	1.33	2.25	1.60	2	1
H				0.F.0	1.G.0		2.F.0	1.67	2.50	1.80	0	1
I	0.I.0			0.F.0			1.G.0	2.00	1.75	2.00	0	2
J				0.I.0			0.F.0	1.33	2.00	2.20	0	1
K		1.K.0	2.K.0				0.I.0	1.67	2.25	2.40	1	1
L	0.L.0				1.K.0	2.K.0	IDLE	1.67	2.25	2.40	1	1
M				0.L.0		2.K.0	1.K.0	2.00	1.50	2.60	1	2
N				0.L.0			2.K.0	2.33	1.75	1.80	1	0
P	0.P.0/0.P.1	1.P.0					0.L.0	1.67	2.00	2.00	1	1
R			2.R.0	0.P.0/0.P.1	1.P.0		IDLE	1.67	2.00	2.00	1	1
S				0.P.0/0.P.1		2.R.0	1.P.0	2.00	1.25	2.20	1	2
T	0.T.0			0.P.0/0.P.1			2.R.0	2.33	1.50	1.40	1	0
U		1.U.0		0.P.1/0.T.0			0.P.0	1.67	1.75	1.60	1	1
V				0.T.0	1.U.0		0.P.1	2.00	2.00	1.80	1	1
W	0.W.0		2.W.0	0.T.0			1.U.0	2.33	1.25	2.00	1	2
X				0.W.0		2.W.0	0.T.0	1.67	1.50	2.20	1	1
Y				0.W.0			2.W.0	2.00	1.75	1.40	1	0
Z							0.W.0	1.33	2.00	1.60	1	1

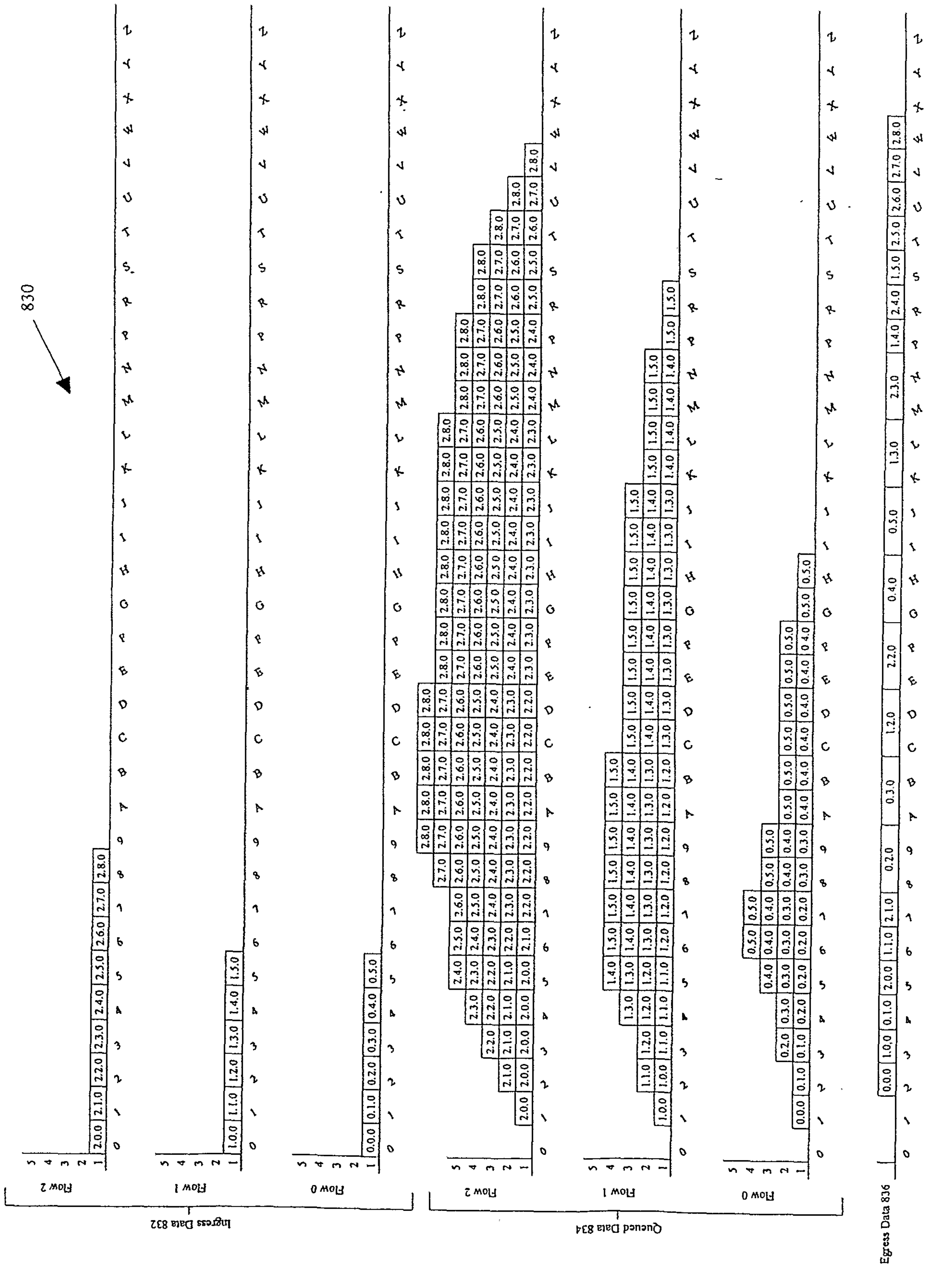

830
Ingress Data 832
Queued Data 834
Egress Data 836
Flow 0
Flow 1
Flow 2
1 2 3 4 5
0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N P R S T U V W X Y Z
2.0.0 2.1.0 2.2.0 2.3.0 2.4.0 2.5.0 2.6.0 2.7.0 2.8.0
1.0.0 1.1.0 1.2.0 1.3.0 1.4.0 1.5.0
0.0.0 0.1.0 0.2.0 0.3.0 0.4.0 0.5.0
0.0.0 1.0.0 0.1.0 2.0.0 1.1.0 2.1.0 0.2.0 0.3.0 1.2.0 2.2.0 0.4.0 0.5.0 1.3.0 2.3.0 1.4.0 2.4.0 1.5.0 2.5.0 2.6.0 2.7.0 2.8.0

FIGURE 19

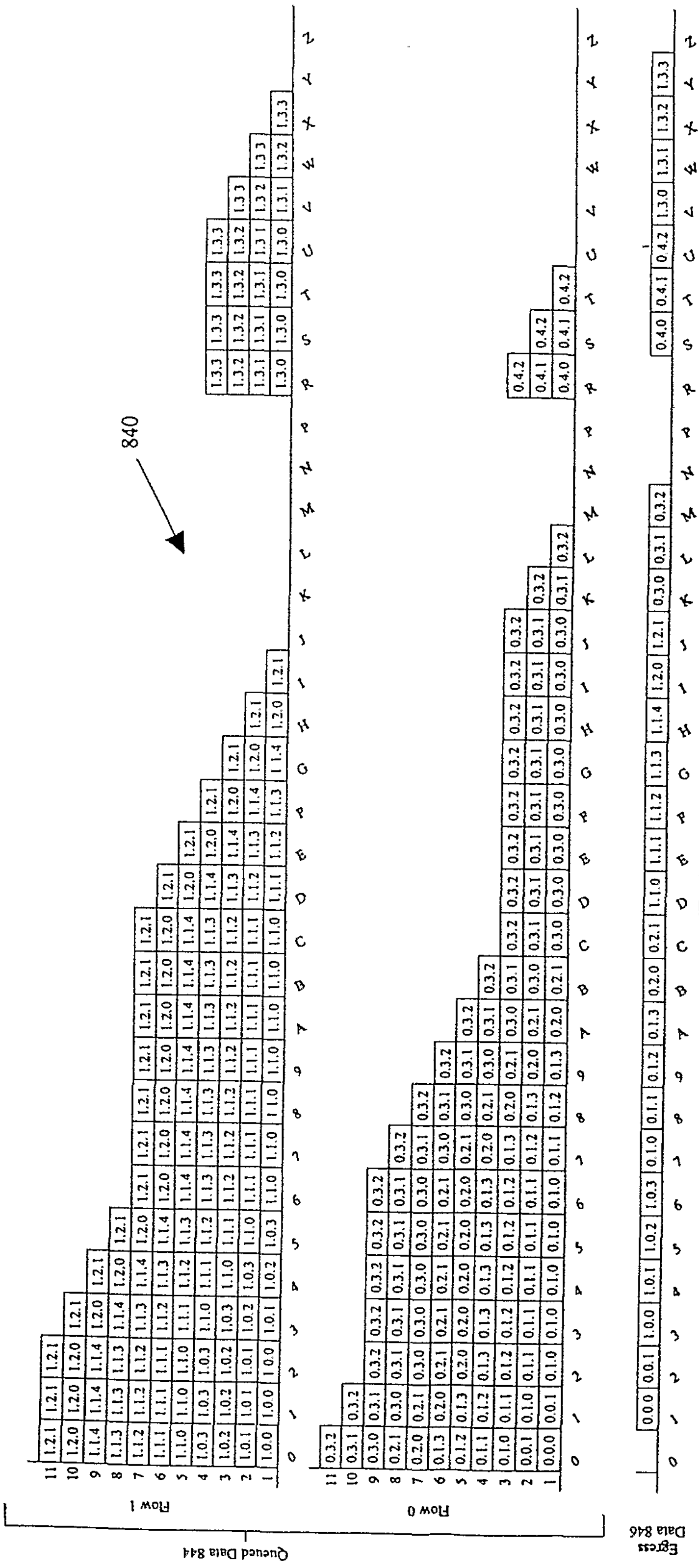

840
Queued Data 844
Flow 1
Flow 0
Egress Data 846
0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N P R S T U V W X Y Z
1 2 3 4 5 6 7 8 9 10 11

FIGURE 20

Figure 21: SP Credit Counting Example

	Ingress Data			Queued Data			Egress	SP Credits				
Time	Flow 0	Flow 1	Flow 2	Flow 0	Flow 1	Flow 2	Data	Flow 0	Flow 1	Flow 2	NCRR	SRR
-1	.2/0.1.3/0.	.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.1					IDLE	65.00	63.75	63.00	0	0
0				0.0.0/0.0.1/0.1.0/0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/	.0.2/1.0.3/1.1.0/1.1.1/1.1.2/1.1.3/1.1.		IDLE	65.00	63.75	63.00	0	0
1				0.0.1/0.1.0/0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3	.0.2/1.0.3/1.1.0/1.1.1/1.1.2/1.1.3/1.1.		0.0.0	64.25	64.00	63.50	0	1
2				0.1.0/0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/	.0.2/1.0.3/1.1.0/1.1.1/1.1.2/1.1.3/1.1.		0.0.1	63.50	64.25	64.00	0	1
3				0.1.0/0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/	2/1.0.3/1.1.0/1.1.1/1.1.2/1.1.3/1.1.4/		1.0.0	63.75	63.50	64.50	0	1
4				0.1.0/0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/	.0.3/1.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.		1.0.1	64.00	62.75	65.00	0	1
5				0.1.0/0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/	3/1.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/		1.0.2	64.25	62.00	65.00	0	1
6				0.1.0/0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/	.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.		1.0.3	64.50	61.25	65.00	0	1
7				0.1.1/0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/0.3	.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.		0.1.0	63.75	61.50	65.00	0	1
8				0.1.2/0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/0.3.2	.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.		0.1.1	63.00	61.75	65.00	0	1
9				0.1.3/0.2.0/0.2.1/0.3.0/0.3.1/0.3.2	.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.		0.1.2	62.25	62.00	65.00	0	1
A				0.2.0/0.2.1/0.3.0/0.3.1/0.3.2	.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.		0.1.3	61.50	62.25	65.00	0	1
B				0.2.1/0.3.0/0.3.1/0.3.2	.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.		0.2.0	61.75	62.50	65.00	1	1
C				0.3.0/0.3.1/0.3.2	.1.0/1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.		0.2.1	62.00	62.75	65.00	1	1
D				0.3.0/0.3.1/0.3.2	1.1.1/1.1.2/1.1.3/1.1.4/1.2.0/1.2.1		1.1.0	62.25	63.00	65.00	2	1
E				0.3.0/0.3.1/0.3.2	1.1.2/1.1.3/1.1.4/1.2.0/1.2.1		1.1.1	62.50	63.25	65.00	2	1
F				0.3.0/0.3.1/0.3.2	1.1.3/1.1.4/1.2.0/1.2.1		1.1.2	62.75	63.50	65.00	2	1
G				0.3.0/0.3.1/0.3.2	1.1.4/1.2.0/1.2.1		1.1.3	63.00	63.75	65.00	2	1
H				0.3.0/0.3.1/0.3.2	1.2.0/1.2.1		1.1.4	63.25	64.00	65.00	2	1
I				0.3.0/0.3.1/0.3.2	1.2.1		1.2.0	63.50	63.25	65.00	2	2
J				0.3.0/0.3.1/0.3.2			1.2.1	63.75	62.50	65.00	2	2
K				0.3.1/0.3.2			0.3.0	64.00	62.75	65.00	1	2
L				0.3.2			0.3.1	64.25	63.00	65.00	1	2
M							0.3.2	64.50	63.25	65.00	1	2
N							IDLE	64.50	63.25	65.00	1	2
P	.0/0.4.1/0.	.3.1/1.3.2/1.3.3					IDLE	64.50	63.25	65.00	1	2
R				0.4.0/0.4.1/0.4.2	1.3.0/1.3.1/1.3.2/1.3.3		IDLE	64.50	63.25	65.00	1	2
S				0.4.1/0.4.2	1.3.0/1.3.1/1.3.2/1.3.3		0.4.0	63.75	63.50	65.00	1	1
T				0.4.2	1.3.0/1.3.1/1.3.2/1.3.3		0.4.1	63.00	63.75	65.00	1	1
U					1.3.0/1.3.1/1.3.2/1.3.3		0.4.2	62.25	64.00	65.00	1	1
V					1.3.1/1.3.2/1.3.3		1.3.0	62.50	63.25	65.00	1	2
W					1.3.2/1.3.3		1.3.1	62.75	62.50	65.00	1	2
X					1.3.3		1.3.2	63.00	61.75	65.00	1	2
Y							1.3.3	63.25	61.00	65.00	1	2
Z							IDLE	63.25	61.00	65.00	1	2

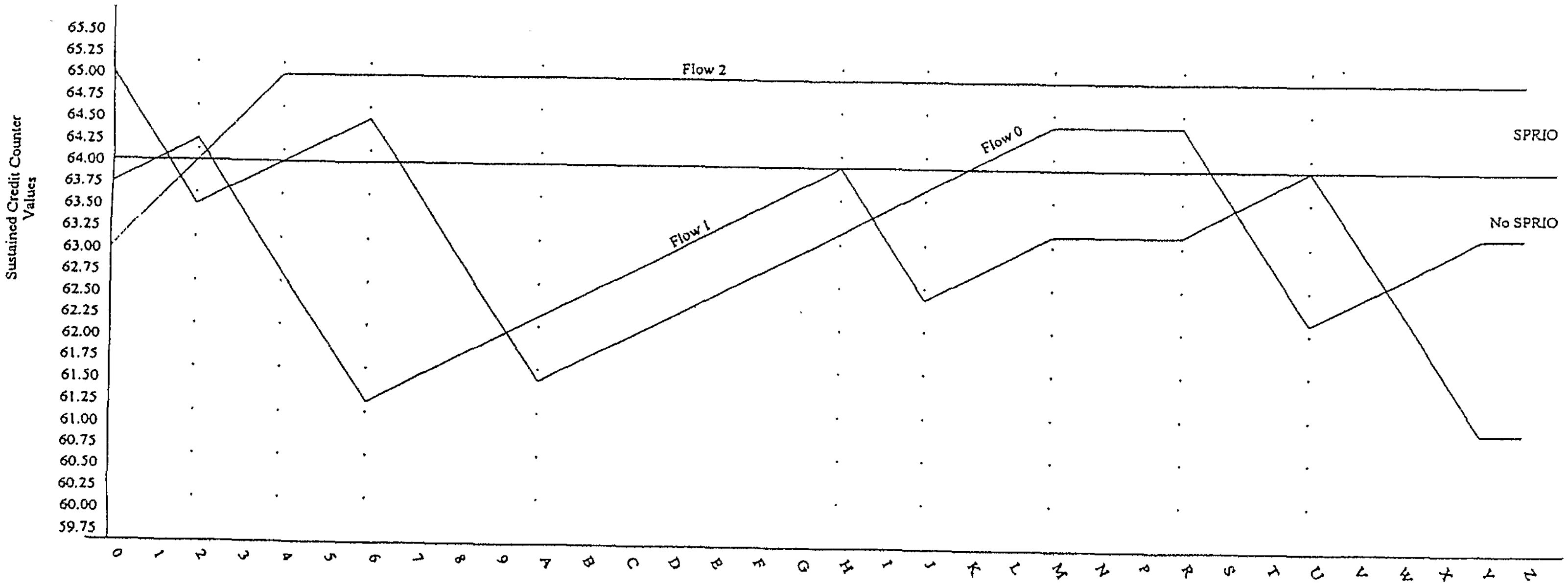

Sustained Credit Counter Values
65.50
65.25
65.00
64.75
64.50
64.25
64.00
63.75
63.50
63.25
63.00
62.75
62.50
62.25
62.00
61.75
61.50
61.25
61.00
60.75
60.50
60.25
60.00
59.75
0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
P
R
S
T
U
V
W
X
Y
Z
Flow 2
Flow 0
Flow 1
SPRIO
No SPRIO

FIGURE 22

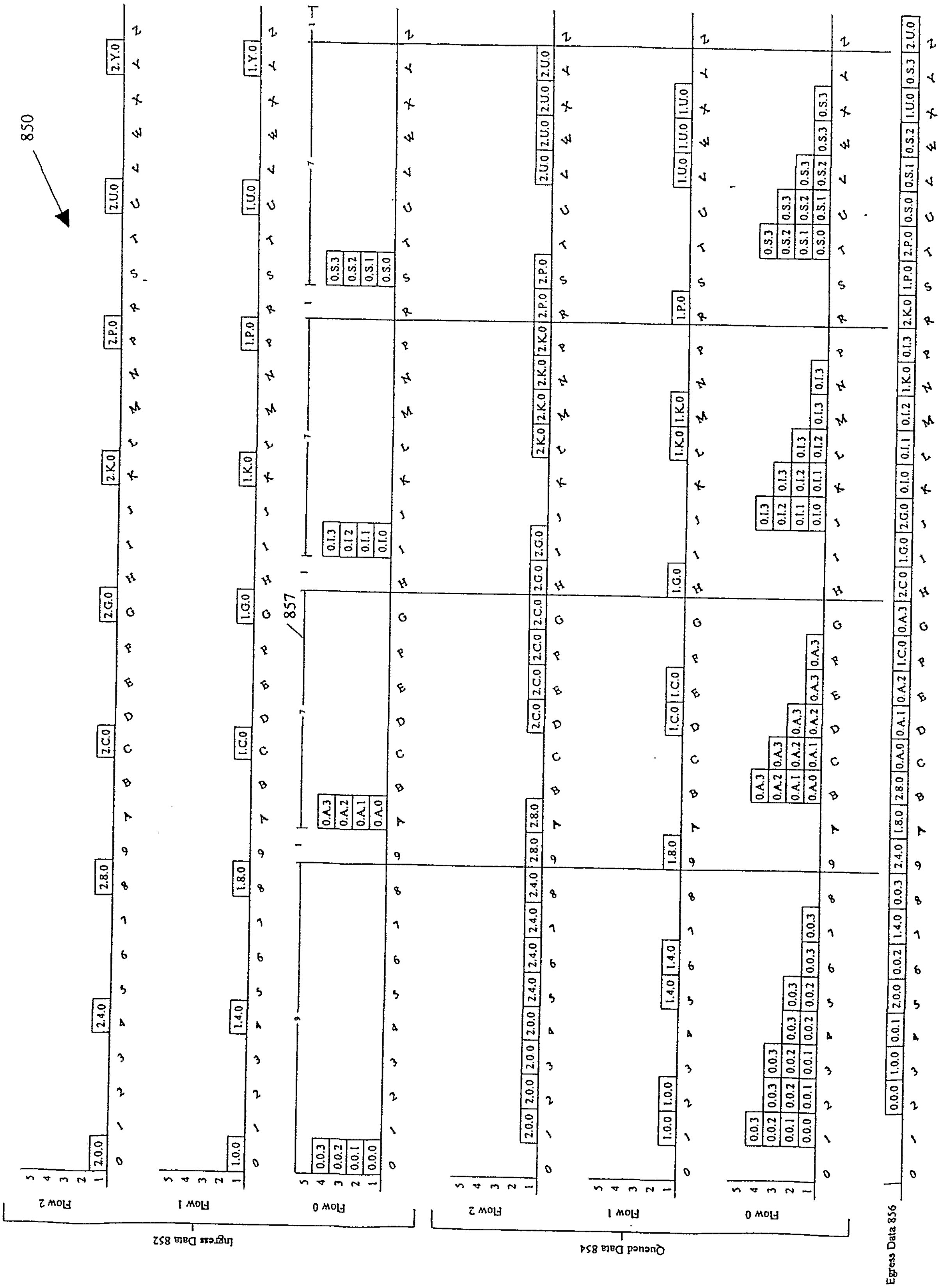

850
857
Ingress Data 852
Queued Data 854
Egress Data 856
Flow 2
Flow 1
Flow 0

FIGURE 23

FIGURE 24: Burst Success Example

	Ingress Data			Queued Data			Egress	SP Credits				
Time	Flow 0	Flow 1	Flow 2	Flow 0	Flow 1	Flow 2	Data	Flow 0	Flow 1	Flow 2	NCRR	SRR
0	0.0.0/0.0.1/0.0.2/0.0.3	1.0.0	2.0.0				IDLE	1.00	1.00	1.00	0	0
1				0.0.0/0.0.1/0.0.2/0.0.3	1.0.0	2.0.0	IDLE	1.00	1.00	1.00	0	0
2				0.0.1/0.0.2/0.0.3	1.0.0	2.0.0	0.0.0	1.50	1.25	1.25	1	0
3				0.0.1/0.0.2/0.0.3		2.0.0	1.0.0	2.00	1.50	1.50	2	0
4		1.4.0	2.4.0	0.0.2/0.0.3		2.0.0	0.0.1	1.50	1.75	1.75	2	0
5				0.0.2/0.0.3	1.4.0	2.4.0	2.0.0	2.00	2.00	2.00	0	0
6				0.0.3	1.4.0	2.4.0	0.0.2	1.50	2.25	2.25	0	0
7				0.0.3		2.4.0	1.4.0	2.00	1.50	2.50	0	2
8		1.8.0	2.8.0			2.4.0	0.0.3	1.50	1.75	2.75	0	2
9					1.8.0	2.8.0	2.4.0	2.00	2.00	2.00	0	0
A	0.A.0/0.A.1/0.A.2/0.A.3					2.8.0	1.8.0	2.50	1.25	2.25	0	2
B				0.A.0/0.A.1/0.A.2/0.A.3			2.8.0	3.00	1.50	1.50	0	0
C		1.C.0	2.C.0	0.A.1/0.A.2/0.A.3			0.A.0	2.50	1.75	1.75	0	0
D				0.A.2/0.A.3	1.C.0	2.C.0	0.A.1	2.00	2.00	2.00	0	0
E				0.A.3	1.C.0	2.C.0	0.A.2	1.50	2.25	2.25	0	0
F				0.A.3		2.C.0	1.C.0	2.00	1.50	2.50	0	2
G		1.G.0	2.G.0			2.C.0	0.A.3	1.50	1.75	2.75	0	2
H					1.G.0	2.G.0	2.C.0	2.00	2.00	2.00	0	0
I	0.I.0/0.I.1/0.I.2/0.I.3					2.G.0	1.G.0	2.50	1.25	2.25	0	2
J				0.I.0/0.I.1/0.I.2/0.I.3			2.G.0	3.00	1.50	1.50	0	0
K		1.K.0	2.K.0	0.I.1/0.I.2/0.I.3			0.I.0	2.50	1.75	1.75	0	0
L				0.I.2/0.I.3	1.K.0	2.K.0	0.I.1	2.00	2.00	2.00	0	0
M				0.I.3	1.K.0	2.K.0	0.I.2	1.50	2.25	2.25	0	0
N				0.I.3		2.K.0	1.K.0	2.00	1.50	2.50	0	2
P		1.P.0	2.P.0			2.K.0	0.I.3	1.50	1.75	2.75	0	2
R					1.P.0	2.P.0	2.K.0	2.00	2.00	2.00	0	0
S	0.S.0/0.S.1/0.S.2/0.S.3					2.P.0	1.P.0	2.50	1.25	2.25	0	2
T				0.S.0/0.S.1/0.S.2/0.S.3			2.P.0	3.00	1.50	1.50	0	0
U		1.U.0	2.U.0	0.S.1/0.S.2/0.S.3			0.S.0	2.50	1.75	1.75	0	0
V				0.S.2/0.S.3	1.U.0	2.U.0	0.S.1	2.00	2.00	2.00	0	0
W				0.S.3	1.U.0	2.U.0	0.S.2	1.50	2.25	2.25	0	0
X				0.S.3		2.U.0	1.U.0	2.00	1.50	2.50	0	2
Y		1.Y.0	2.Y.0			2.U.0	0.S.3	1.50	1.75	2.75	0	2
Z					1.Y.0	2.Y.0	2.U.0	2.00	2.00	2.00	0	0

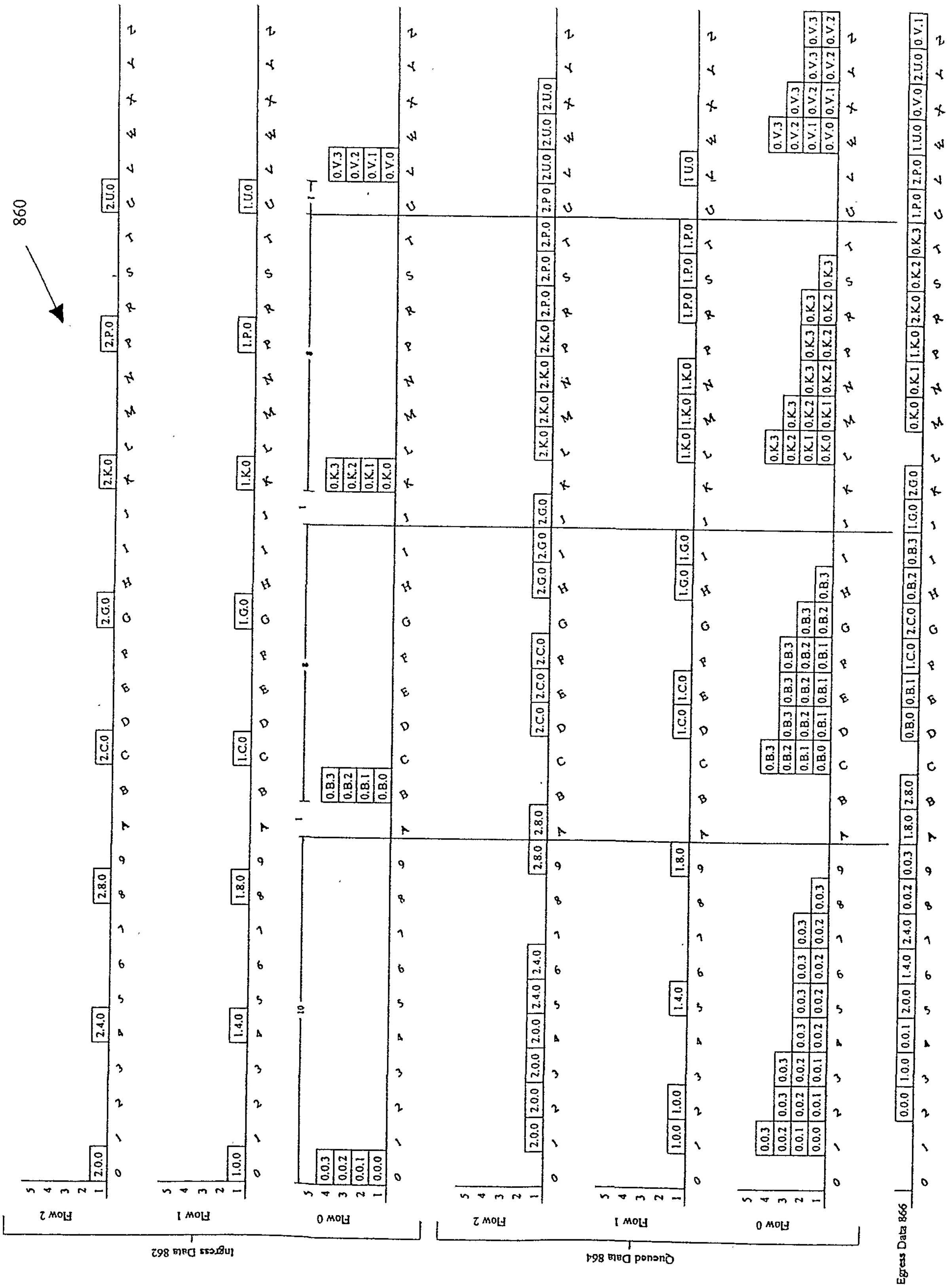

860
Ingress Data 862
Flow 2
Flow 1
Flow 0
Queued Data 864
Flow 2
Flow 1
Flow 0
Egress Data 866
0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N P R S T U V W X Y Z
2.0.0
2.4.0
2.8.0
2.C.0
2.G.0
2.K.0
2.P.0
2.U.0
1.0.0
1.4.0
1.8.0
1.C.0
1.G.0
1.K.0
1.P.0
1.U.0
0.0.0
0.0.1
0.0.2
0.0.3
0.B.0
0.B.1
0.B.2
0.B.3
0.K.0
0.K.1
0.K.2
0.K.3
0.V.0
0.V.1
0.V.2
0.V.3
10
8
8
1
1
1

FIGURE 25

FIGURE 26: Burst Fraction Unused

	Ingress Data			Queued Data			Egress	SP Credits				
Time	Flow 0	Flow 1	Flow 2	Flow 0	Flow 1	Flow 2	Data	Flow 0	Flow 1	Flow 2	NCRR	SRR
0	0.0.0/0.0.1/0.0.2/0.0.3	1.0.0	2.0.0				IDLE	1.00	1.00	1.00	0	0
1				0.0.0/0.0.1/0.0.2/0.0.3	1.0.0	2.0.0	IDLE	1.00	1.00	1.00	0	0
2				0.0.1/0.0.2/0.0.3	1.0.0	2.0.0	0.0.0	1.50	1.25	1.25	1	0
3				0.0.1/0.0.2/0.0.3		2.0.0	1.0.0	2.00	1.50	1.50	2	0
4		1.4.0	2.4.0	0.0.2/0.0.3		2.0.0	0.0.1	1.50	1.75	1.75	2	1
5				0.0.2/0.0.3	1.4.0	2.4.0	2.0.0	2.00	2.00	2.00	0	1
6				0.0.2/0.0.3		2.4.0	1.4.0	2.50	1.25	2.25	0	2
7				0.0.2/0.0.3			2.4.0	3.00	1.50	1.50	0	0
8		1.8.0	2.8.0	0.0.3			0.0.2	2.50	1.75	1.75	0	1
9					1.8.0	2.8.0	0.0.3	2.00	2.00	2.00	0	1
A						2.8.0	1.8.0	2.50	1.25	2.25	0	2
B	0.B.0/0.B.1/0.B.2/0.B.3						2.8.0	3.00	1.50	1.50	0	0
C		1.C.0	2.C.0	0.B.0/0.B.1/0.B.2/0.B.3			IDLE	3.00	1.50	1.50	0	0
D				0.B.1/0.B.2/0.B.3	1.C.0	2.C.0	0.B.0	2.50	1.75	1.75	0	1
E				0.B.2/0.B.3	1.C.0	2.C.0	0.B.1	2.00	2.00	2.00	0	1
F				0.B.2/0.B.3		2.C.0	1.C.0	2.50	1.25	2.25	0	2
G		1.G.0	2.G.0	0.B.2/0.B.3			2.C.0	3.00	1.50	1.50	0	0
H				0.B.3	1.G.0	2.G.0	0.B.2	2.50	1.75	1.75	0	1
I					1.G.0	2.G.0	0.B.3	2.00	2.00	2.00	0	1
J						2.G.0	1.G.0	2.50	1.25	2.25	0	2
K	0.K.0/0.K.1/0.K.2/0.K.3	1.K.0	2.K.0				2.G.0	3.00	1.50	1.50	0	0
L				0.K.0/0.K.1/0.K.2/0.K.3	1.K.0	2.K.0	IDLE	3.00	1.50	1.50	0	0
M				0.K.1/0.K.2/0.K.3	1.K.0	2.K.0	0.K.0	2.50	1.75	1.75	0	1
N				0.K.2/0.K.3	1.K.0	2.K.0	0.K.1	2.00	2.00	2.00	0	1
P		1.P.0	2.P.0	0.K.2/0.K.3		2.K.0	1.K.0	2.50	1.25	2.25	0	2
R				0.K.2/0.K.3	1.P.0	2.P.0	2.K.0	3.00	1.50	1.50	0	0
S				0.K.3	1.P.0	2.P.0	0.K.2	2.50	1.75	1.75	0	1
T					1.P.0	2.P.0	0.K.3	2.00	2.00	2.00	0	1
U		1.U.0	2.U.0			2.P.0	1.P.0	2.50	1.25	2.25	0	2
V	0.V.0/0.V.1/0.V.2/0.V.3				1.U.0	2.U.0	2.P.0	3.00	1.50	1.50	0	0
W				0.V.0/0.V.1/0.V.2/0.V.3		2.U.0	1.U.0	3.50	1.75	1.75	2	0
X				0.V.1/0.V.2/0.V.3		2.U.0	0.V.0	3.00	2.00	2.00	2	1
Y		1.Y.0	2.Y.0	0.V.1/0.V.2/0.V.3			2.U.0	3.50	2.25	1.25	2	0
Z				0.V.2/0.V.3	1.Y.0	2.Y.0	0.V.1	3.00	2.50	1.50	2	1

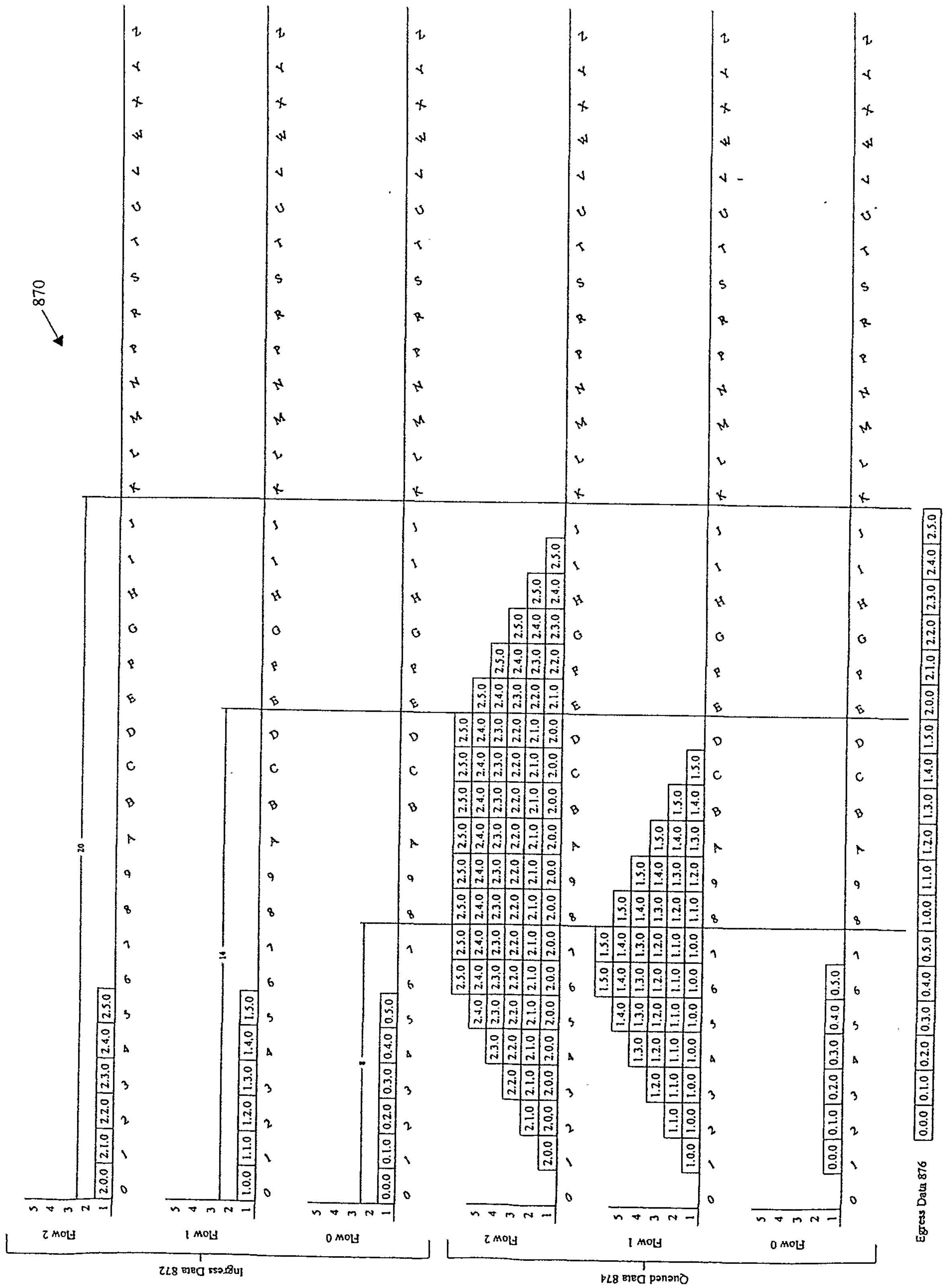

870
Ingress Data 872
Flow 2
Flow 1
Flow 0
20
14
8
Queued Data 874
Flow 2
Flow 1
Flow 0
Egress Data 876
0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N P R S T U V W X Y Z

FIGURE 27

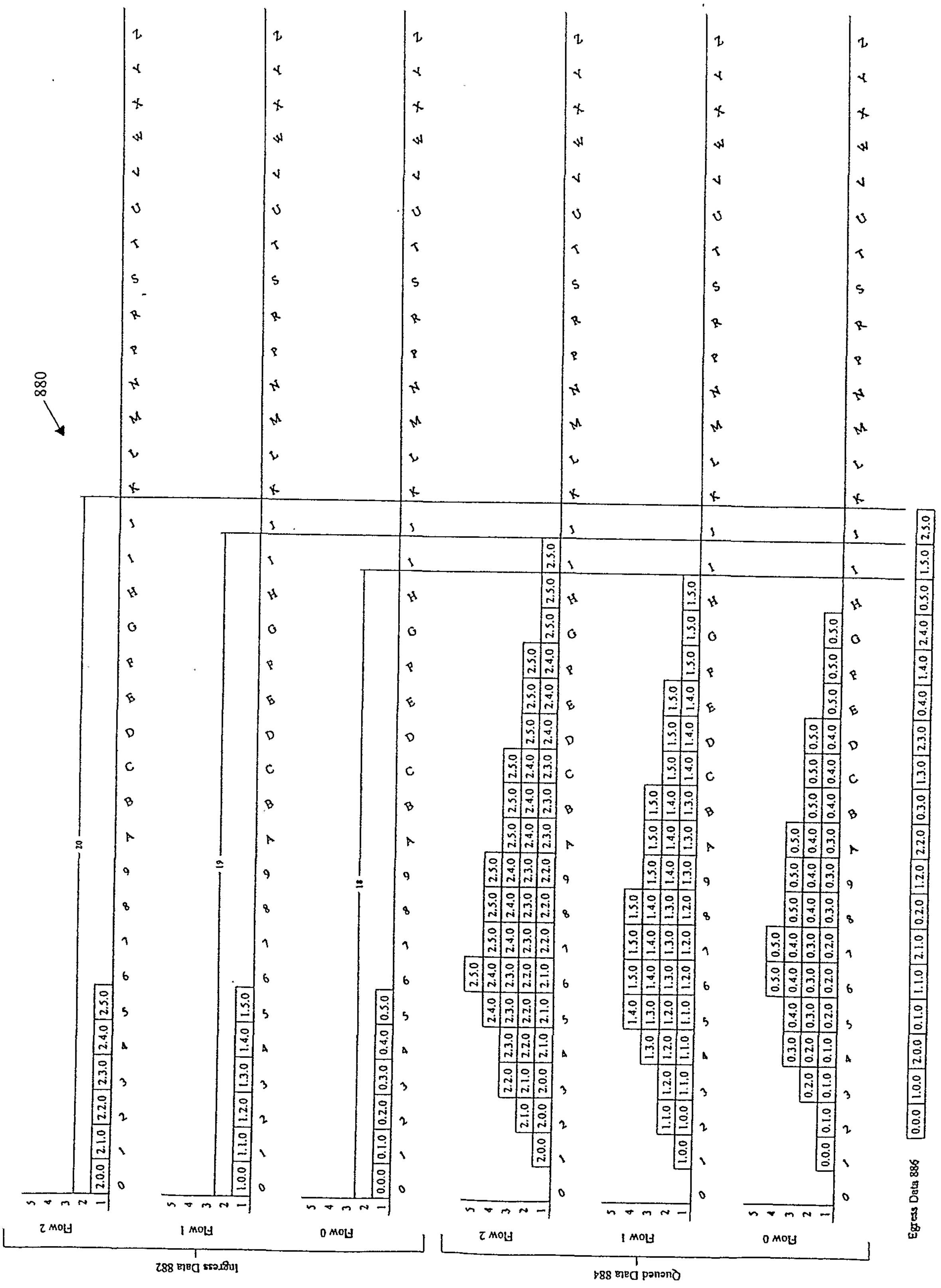

880
Ingress Data 882
Queued Data 884
Egress Data 886
Flow 0
Flow 1
Flow 2
0 1 2 3 4 5 6 7 8 9 A B C D E F G H I J K L M N P R S T U V W X Y Z
18
19
20
2.0.0 2.1.0 2.2.0 2.3.0 2.4.0 2.5.0
1.0.0 1.1.0 1.2.0 1.3.0 1.4.0 1.5.0
0.0.0 0.1.0 0.2.0 0.3.0 0.4.0 0.5.0
0.0.0 1.0.0 2.0.0 0.1.0 1.1.0 2.1.0 0.2.0 1.2.0 2.2.0 0.3.0 1.3.0 2.3.0 0.4.0 1.4.0 2.4.0 0.5.0 1.5.0 2.5.0

FIGURE 28

Figure 29: APP. Unused (SP Allocation)

							Egress					
	Ingress Data 872			Queued Data 884			Data	SP Credits				
Time	Flow 0	Flow 1	Flow 2	Flow 0	Flow 1	Flow 2	886	Flow 0	Flow 1	Flow 2	NCRR	SRR
0	0.0.0	1.0.0	2.0.0				IDLE	1.00	1.00	1.00	0	0
1	0.1.0	1.1.0	2.1.0	0.0.0	1.0.0	2.0.0	IDLE	1.00	1.00	1.00	0	0
2	0.2.0	1.2.0	2.2.0	0.1.0	1.0.0/1.1.0	2.0.0/2.1.0	0.0.0	1.33	1.33	1.33	1	0
3	0.3.0	1.3.0	2.3.0	0.1.0/0.2.0	1.1.0/1.2.0	2.0.0/2.1.0/2.2.0	1.0.0	1.67	1.67	1.67	2	0
4	0.4.0	1.4.0	2.4.0	0.1.0/0.2.0/0.3.0	1.1.0/1.2.0/1.3.0	2.1.0/2.2.0/2.3.0	2.0.0	2.00	2.00	2.00	0	0
5	0.5.0	1.5.0	2.5.0	0.2.0/0.3.0/0.4.0	1.1.0/1.2.0/1.3.0/1.4.0	2.1.0/2.2.0/2.3.0/2.4.0	0.1.0	1.33	2.33	2.33	0	1
6				0.2.0/0.3.0/0.4.0/0.5.0	1.2.0/1.3.0/1.4.0/1.5.0	2.1.0/2.2.0/2.3.0/2.4.0/2.5.0	1.1.0	1.67	1.67	2.67	0	2
7				0.2.0/0.3.0/0.4.0/0.5.0	1.2.0/1.3.0/1.4.0/1.5.0	2.2.0/2.3.0/2.4.0/2.5.0	2.1.0	2.00	2.00	2.00	0	0
8				0.3.0/0.4.0/0.5.0	1.2.0/1.3.0/1.4.0/1.5.0	2.2.0/2.3.0/2.4.0/2.5.0	0.2.0	1.33	2.33	2.33	0	1
9				0.3.0/0.4.0/0.5.0	1.3.0/1.4.0/1.5.0	2.2.0/2.3.0/2.4.0/2.5.0	1.2.0	1.67	1.67	2.67	0	2
A				0.3.0/0.4.0/0.5.0	1.3.0/1.4.0/1.5.0	2.3.0/2.4.0/2.5.0	2.2.0	2.00	2.00	2.00	0	0
B				0.4.0/0.5.0	1.3.0/1.4.0/1.5.0	2.3.0/2.4.0/2.5.0	0.3.0	1.33	2.33	2.33	0	1
C				0.4.0/0.5.0	1.4.0/1.5.0	2.3.0/2.4.0/2.5.0	1.3.0	1.67	1.67	2.67	0	2
D				0.4.0/0.5.0	1.4.0/1.5.0	2.4.0/2.5.0	2.3.0	2.00	2.00	2.00	0	0
E				0.5.0	1.4.0/1.5.0	2.4.0/2.5.0	0.4.0	1.33	2.33	2.33	0	1
F				0.5.0	1.5.0	2.4.0/2.5.0	1.4.0	1.67	1.67	2.67	0	2
G				0.5.0	1.5.0	2.5.0	2.4.0	2.00	2.00	2.00	0	0
H					1.5.0	2.5.0	0.5.0	1.33	2.33	2.33	0	1
I						2.5.0	1.5.0	1.67	1.67	2.67	0	2
J							2.5.0	2.00	2.00	2.00	0	0
K							IDLE	2.00	2.00	2.00	0	0
L							IDLE	2.00	2.00	2.00	0	0
M							IDLE	2.00	2.00	2.00	0	0
N							IDLE	2.00	2.00	2.00	0	0
P							IDLE	2.00	2.00	2.00	0	0
R							IDLE	2.00	2.00	2.00	0	0
S							IDLE	2.00	2.00	2.00	0	0
T							IDLE	2.00	2.00	2.00	0	0
U							IDLE	2.00	2.00	2.00	0	0
V							IDLE	2.00	2.00	2.00	0	0
W							IDLE	2.00	2.00	2.00	0	0
X							IDLE	2.00	2.00	2.00	0	0
Y							IDLE	2.00	2.00	2.00	0	0
Z							IDLE	2.00	2.00	2.00	0	0