

\*\*\*\*\*

ATM Forum Document Number: ATM\_Forum/96-1762

\*\*\*\*\*

**Title:** MIMO Latency: Revised Proposed Text

\*\*\*\*\*

**Abstract:** Replacement text for Appendix A of the ATM forum performance testing specification is presented.

\*\*\*\*\*

**Source:**

Raj Jain, Gojko Babic

The Ohio State University

Department of CIS Columbus, OH 43210-1277

Phone: 614-292-3989, Fax: 614-292-2911, Email: Jain@ACM.Org

The presentation of this contribution at the ATM Forum is sponsored by NASA.

\*\*\*\*\*

**Date:** December 1996

\*\*\*\*\*

**Distribution:** ATM Forum Technical Working Group Members (AF-TEST, AF-TM)

\*\*\*\*\*

**Notice:**

This contribution has been prepared to assist the ATM Forum. It is offered to the Forum as a basis for discussion and is not a binding proposal on the part of any of the contributing organizations. The statements are subject to change in form and content after further study. Specifically, the contributors reserve the right to add to, amend or modify the statements contained herein.

\*\*\*\*\*

In the October meeting, we presented a revised definition for MIMO latency. This definition allows for discontinuous frames on input. This contribution presents the text corresponding to that presentation.

Motion: Adopt the text of this contribution as a replacement for Appendix A of the baseline document.

# Appendix A: MIMO Latency

## A.1. Definition

MIMO latency (Message-In Message-Out) is a general definition of the latency that applies to an ATM switch or a group of ATM switches and it is defined as follows:

$$\text{MIMO latency} = \min \{ \text{LILO latency}, \text{FILO latency} - \text{NFOT} \}$$

where:

- LILO latency = Time between the last-bit entry and the last-bit exit
- FILO latency = Time between the first-bit entry and the last-bit exit
- NFOT = Nominal Frame Output Time = FIT  $\times$  Input Rate/Output Rate
- FIT = Frame Input Time = Time between the first-bit entry and the last-bit entry

Note that for contiguous frames on input:

$$\text{Frame Input Time} = \text{Frame Size} / \text{Input rate}$$

and then it follows:

$$\begin{aligned} \text{Normalized Frame Output Time} &= \text{Frame Size} / \text{Input Rate} \times \text{Input Rate} / \text{Output Rate} \\ &= \text{Frame Size} / \text{Output rate} \end{aligned}$$

The following is an equivalent definition for MIMO Latency:

$$\text{MIMO latency} = \begin{cases} \text{LILO latency} & \text{if input rate} \leq \text{output rate} \\ \text{FILO latency} - \text{NFOT} & \text{if input rate} \geq \text{output rate} \end{cases}$$

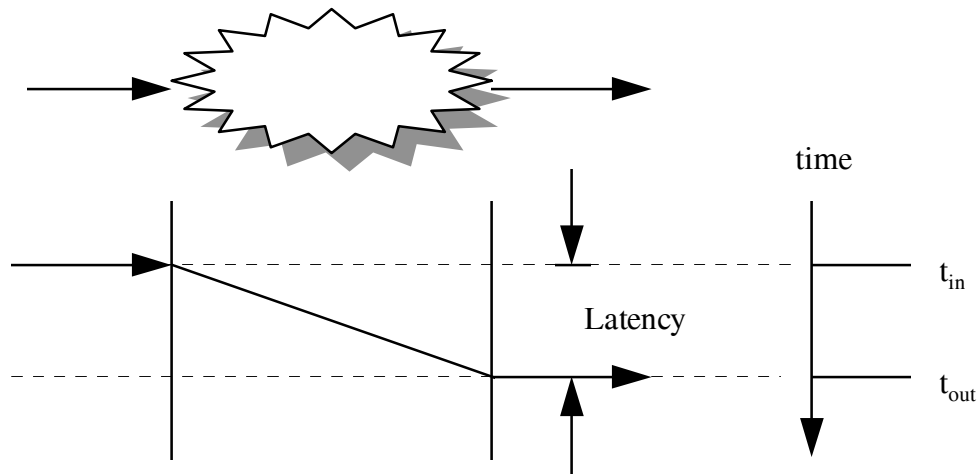
Note that for input rate = output rate:

$$\text{LILO} = \text{FILO} - \text{NFOT}$$

## A.2. Introduction

In the rest of the Appendix we justify the MIMO latency definition. In this section, we start with a single bit case and a simple contiguous frame case. Then we systematically consider contiguous frame cases and then discontinuous frame cases in the ATM environment in Section A.3 and Section A.4, respectively.

For a single bit case (see Figure A.1), the latency is generally defined as the time between the instant the bit enters the system to the instant the bit exits from the system.



**Figure A.1:** Latency for a single bit

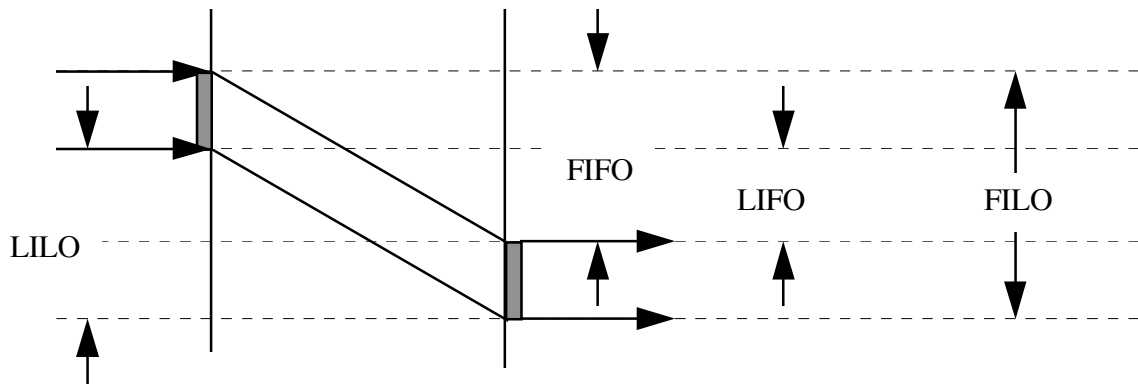
For multi-bit frames, there are several possible definitions. Consider the case of contiguous frames, i.e. all bits of the frames are sent (on input) and delivered (on output) contiguously without any gap between bits. In this case, latency can be defined in one of the following four ways:

1. FIFO latency: Time between the first-bit entry and the first-bit exit
2. LILO latency: Time between the last-bit entry and the last-bit exit
3. FILO latency: Time between the first-bit entry and the last-bit exit
4. LIFO latency: Time between the last-bit entry and the first-bit exit

If the input link and the output link are of the same speed and frames are contiguous (see Figure A.2), FIFO and LILO latencies are identical. In this case FILO and LIFO latencies can be computed from FIFO (or LILO) latency given the frame length and input rate or output rate:

$$\begin{aligned} \text{FILO} &= \text{FIFO} + \text{Frame Size}/\text{Input rate} = \text{FIFO} + \text{Frame Size}/\text{Output rate} \\ \text{LIFO} &= \text{FIFO} - \text{Frame Size}/\text{Input rate} = \text{FIFO} - \text{Frame Size}/\text{Output rate} \end{aligned}$$

It is clear that FIFO (or LILO) is a preferred metric in this case since it is independent of the frame length while FILO and LIFO would be different for each frame size. That is one of reasons why we shall not further consider FILO and LIFO.



**Figure A.2:** Latency for multi-bit frames, input rate = output rate

Unfortunately, none of the above four metrics apply to an ATM network (or switch) latency since:

- the input and output link may be of different speeds, and
- the frames are not always sent in (on input) or delivered out (on output) contiguously, i.e., there may be idle times between cells of a frame either on input and/or output.

In the following, we consider first contiguous frames and then discontinuous frames in an ATM network. We compare FIFO, LILO and MIMO metrics and show that MIMO is the correct metric in all cases while other metrics apply to some cases but give incorrect results in others.

### **A.3. Contiguous Frames**

In this section we consider cases where frames on input as well as on output are contiguous, i.e without any gaps between their cells. Depending upon the relative magnitude of input and output rates and the delay through the switch, there are six possible cases. These cases and the applicability of the three metrics are shown in Table A.1.

**Table A.1:** Applicability of Various Latency Definitions For Contiguous Frames

No.	Case	FIFO	LILO	MIMO
1aC	Input rate = Output rate, zero-delay switch	√	√	√
1bC	Input rate = Output rate, nonzero-delay switch	√	√	√
2aC	Input rate < Output rate, zero-delay switch	×	√	√
2bC	Input rate < Output rate, nonzero-	×	√	√

	delay switch			
3aC	Input rate > Output rate, zero-delay switch	√	×	√
3bC	Input rate > Output rate, nonzero-delay switch	√	×	√

√ ⇒ The matrix gives a valid result.

× ⇒ The matrix gives an invalid result.

As indicated above, we consider a zero-delay switch and a nonzero-delay switch. The cases with a zero-delay switch are especially useful to verify the validity of a latency definition, because the switch delay is known in advance (equal to zero).

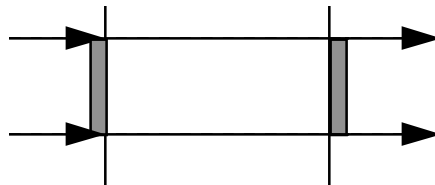
It should be noted that in all cases for contiguous frames on input and on output, the following relation always holds:

FIFO = FILO – Frame size/Output rate,

and this relation will be used in this section.

### Case 1aC: Contiguous Frames, Input rate = Output rate, Zero-Delay Switch

Figure A.1aC shows the flow in this case.



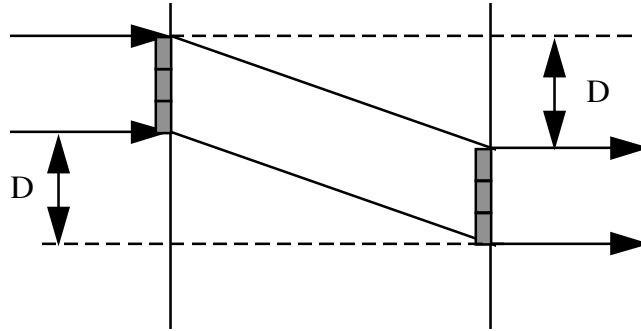
**Figure A.1aC:** Contiguous frames, Input rate = Output rate, Zero-delay switch

In this case, the bits appear on the output as soon as they enter on the input. Here we have:

- FIFO = 0, correct
- LILO = 0, correct
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{Frame Size}/\text{Output rate} \} = \min \{ 0, \text{FIFO} \} = 0$ , correct

### Case 1bC: Contiguous Frames, Input rate = Output rate, Nonzero-Delay Switch

Figure A.1bC shows the flow in this case.



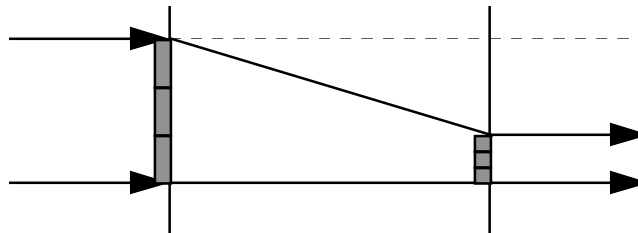
**Figure A.1bC:** Contiguous frames, Input rate = Output rate, Nonzero-delay switch

In this case, the switch latency  $D$  is determined by a delay of the first bit (or the last bit). Here we have:

- FIFO =  $D$ , correct
- LILO =  $D$ , correct
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{Frame Size}/\text{Output Rate} \} = \min \{ D, \text{FIFO} \} = D$ , correct

**Case 2aC: Contiguous Frames, Input rate < Output rate, Zero-Delay Switch**

Figure A.2aC shows the flow in this case.



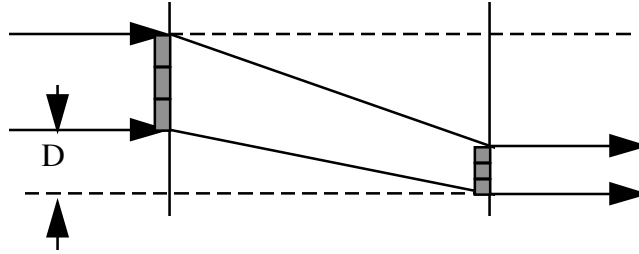
**Figure A.2aC:** Contiguous frames, Input rate < Output rate, Zero-delay switch

In this case, a contiguous frame on the output is possible only if the transmission of incoming bits is scheduled such that there will not be any buffer underflow until the last bit. Here we have:

- FIFO > 0, incorrect; Note that FIFO may change with changing output rate (while not changing the switch latency). So, FIFO does not correctly represent the switch latency.
- LILO = 0, correct
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{Frame Size}/\text{Output Rate} \} = \min \{ 0, \text{FIFO} \} = 0$ , correct

**Case 2bC: Contiguous Frames, Input rate < Output rate, Nonzero-Delay Switch**

Figure A.2bC shows the flow in this case.



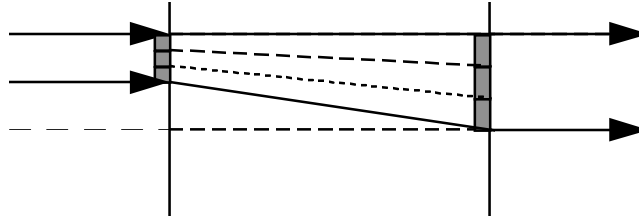
**Figure A.2bC:** Contiguous frames, Input rate < Output rate, Nonzero-delay switch

In this case, the switch latency  $D$  is determined by a delay of the last bit. Here we have:

- FIFO >  $D$ , incorrect; As in Case 2aC, FIFO may change with changing output rate (without changing the switch latency). So, FIFO does not correctly represent the switch latency.
- LILO =  $D$ , correct
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{Frame Size}/\text{Output Rate} \} = \min \{ D, \text{FIFO} \} = D$ , correct

**Case 3aC: Contiguous Frames, Input rate > Output rate, Zero-Delay Switch**

Figure A.3aC shows the flow in this case.



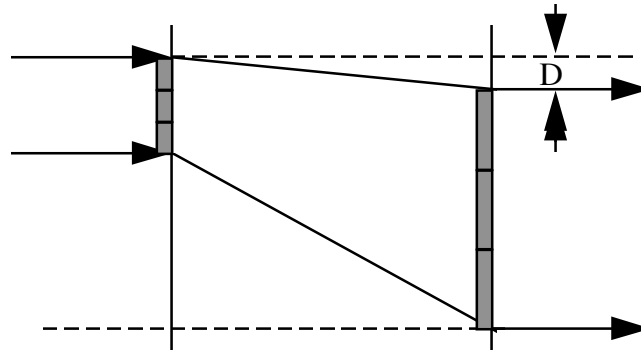
**Figure A.3aC:** Contiguous frames, Input rate > Output rate, Zero-delay switch

In this case, only the first bit on the input appears immediately on the output, and other bits have to be buffered, because the input rate is larger (more bits are input) than the output rate (fewer bits are output). Here we have:

- FIFO = 0, correct
- LILO > 0, incorrect; Note that LILO may change with changing the output rate and not changing the switch otherwise. So, LILO does not correctly represent the switch latency.
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{Frame Size}/\text{Output Rate} \} = \min \{ \text{LILO}, \text{FIFO} \} = 0$ , correct

**Case 3bC: Contiguous Frames, Input rate > Output rate, Nonzero-Delay Switch**

Figure A.3bC shows the flow in this case.



**Figure A.3bC:** Contiguous frames, Input rate > Output rate, nonzero-delay switch

In this case, the switch latency  $D$  is determined by a delay of the first bit. Here we have:

- FIFO =  $D$ , correct
- LILO >  $D$ , incorrect; As in Case 3aC, LILO may change with changing the output rate and not changing the switch otherwise. So, LILO does not correctly represent the switch latency.
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{Frame Size}/\text{Output Rate} \} = \min \{ \text{LILO}, \text{FIFO} \} = D$ , correct

#### **A.4. Discontiguous Frames**

In this section we consider cases where frames on input as well as on output are discontiguous, i.e. there are gaps between cells of frames. Depending upon the number of gaps on input and output, we have three possibilities:

- The number of gaps on output is same as that on input. This is the case of *no change in gaps*.
- The number of gaps on output is more than that on input. This is the case of *expansion of gaps*.
- The number of gaps on output is less than that on input. This is the case of *compression of gaps*.

It should be noted that cases with contiguous frames on input and/or output are special cases of discontiguous frames with no gaps.

The nine cases and the applicability of the three metrics (FIFO, LILO and MIMO) to those cases are shown in Table A.2. Each case includes a case with a nonzero delay switch and (if possible) a case with a zero-delay switch.

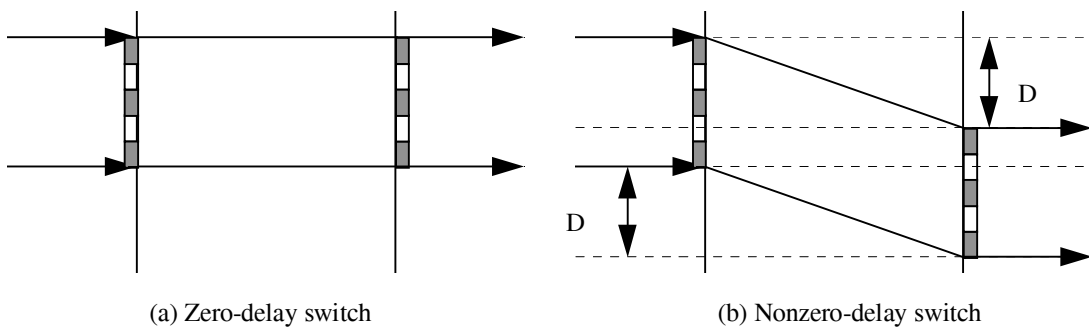
**Table A.2:** Applicability of Various Latency Definitions For Discontiguous Frames

No.	Case	FIFO	LILO	MIMO
1aD	Input rate = Output rate, no change in gaps	√	√	√
1bD	Input rate = Output rate, expansion of gaps	×	√	√
1cD	Input rate = Output rate, compression of gaps	×	√	√
2aD	Input rate < Output rate, no change in gaps	×	√	√
2bD	Input rate < Output rate, expansion of gaps	×	√	√
2cD	Input rate < Output rate, compression of gaps	×	√	√
3aD	Input rate > Output rate, no change in gaps	√	×	√
3bD	Input rate > Output rate, expansion of gaps	×	×	√
3cD	Input rate > Output rate, compression of gaps	×	×	√

√ ⇒ The matrix gives a valid result.  
 × ⇒ The matrix gives an invalid result.

**Case 1aD: Discontiguous Frames, Input rate = Output rate, No Changes in Gaps**

Figure A.1aD shows the flow for a zero-delay switch and a nonzero-delay switch.



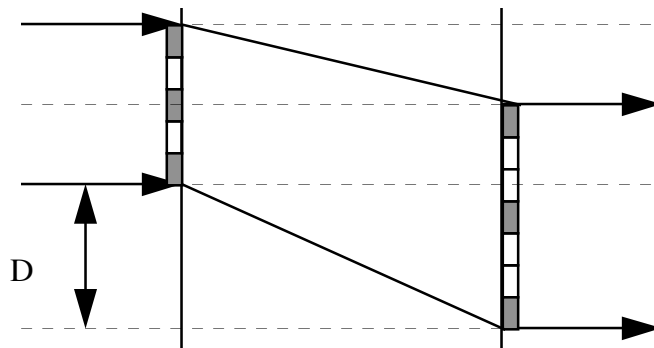
**Figure A.1aD:** Discontiguous frames, Input rate = Output rate, No change in gaps

This case is similar to cases 1aC and 1bC. The switch latency is determined by a delay of the first bit (or the last bit). Here we have:

- FIFO = D, correct
- LILO = D, correct
- Input rate = Output rate
- ⇒ MIMO = min {LILO, FILO – FIT} = {D, D} = D, correct

**Case 1bD: Discontiguous Frames, Input Rate = Output Rate, Expansion of Gaps**

Figure A.1bD shows the flow for a nonzero-delay switch, while a zero-delay switch with expansion of gaps is an impossible scenario.



**Figure A.1bD:** Discontiguous frames, Input rate = Output rate, Expansion of gaps

In this case, the switch latency D is given by:

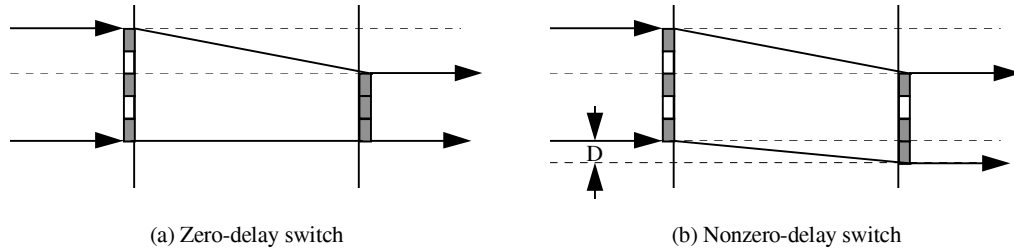
$$D = \text{first bit delay} + \text{time of additional gaps on output}$$

Here we have:

- FIFO < D, incorrect; FIFO is incorrect because it does not reflect expansion of gaps. Note that for a nonzero-delay switch, FIFO may be zero (the case of zero delay for the first bit)
- LILO = D, correct
- Input rate = Output rate
- ⇒ MIMO = min {LILO, FILO – FIT} = min {D, D} = D, correct

**Case 1cD: Discontiguous Frames, Input Rate = Output Rate, Compression of Gaps**

Figure A.1cD shows the flow for a zero-delay and a nonzero-delay switch with compression of gaps.



**Figure A.1cD:** Discontiguous frames, Input rate = Output rate, Compression of gaps  
 In this case, the switch latency  $D$  is given by:

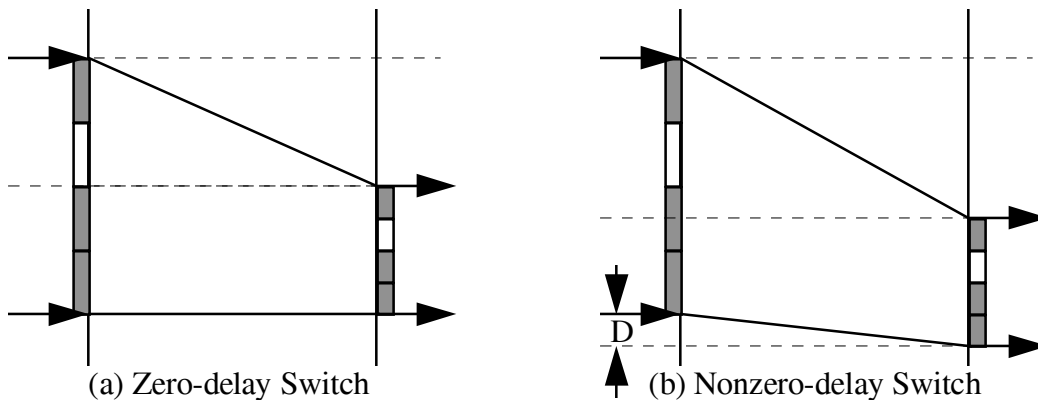
$$D = \text{Last bit delay} = \text{First bit delay} - \text{Time of additional gaps on input}$$

Here we have:

- FIFO >  $D$ , incorrect; FIFO is incorrect because it does not reflect compression of gaps.
- LILO =  $D$ , correct
- Input rate = Output rate
- $\Rightarrow$  MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{FIT} \} = \min \{ D, D \} = D$ , correct

**Case 2aD: Discontiguous Frames, Input Rate < Output Rate, No change in Gaps**

Figure A.2aD shows the flow for a zero-delay switch and a nonzero-delay switch.



**Figure A.2aD:** Discontiguous frames, Input rate < Output rate, No change in gaps  
 This case is similar to cases 2aC and 2bC. The switch latency  $D$  is determined by a delay of the last bit. Here we have:

- FIFO >  $D$ , incorrect; FIFO may change with changing the output rate and not changing the switch otherwise. So, FIFO does not correctly represent the switch latency.
- LILO =  $D$ , correct

- Input rate < Output rate
- $\Rightarrow$  FILO – FIT  $\times$  Input rate / Output rate > D
- $\Rightarrow$  MIMO = min {LILO, FILO – FIT  $\times$  Input rate / Output rate} = D, correct

**Case 2bD: Discontiguous Frames, Input Rate < Output Rate, Expansion of Gaps**

Figure A.2bD shows the flow for a zero-delay switch and a nonzero-delay switch.

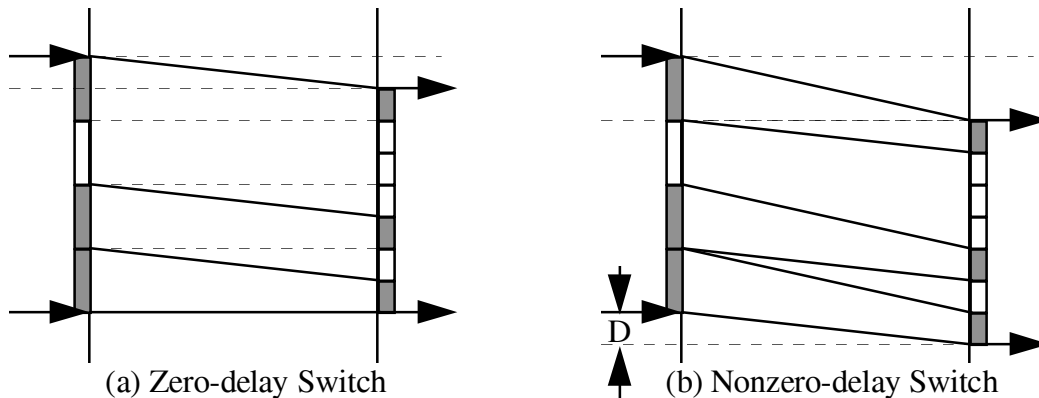


Figure A.2bD: Discontiguous frames, Input rate < Output rate, Expansion of gaps

In this case, the switch latency D is determined by a delay of the last bit. Here we have:

- FIFO is incorrect because:
  - a. FIFO may be affected by changing the output rate and not changing the switch (latency) otherwise.
  - b. FIFO may change by changing the number of gaps on the output while the switch (latency) is unchanged. It should be noted that for this case, with the given input rate and the given number of gaps on input, it is possible to produce cases with the appropriate output rate and the appropriate number of gaps on output such that FIFO > D or FIFO < D or even FIFO = D, all without changing the switch (latency).

- LILO = D, correct

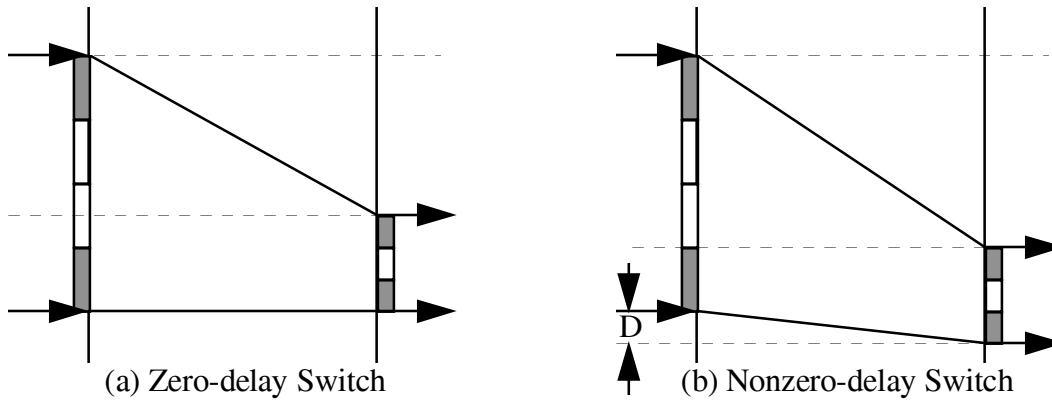
- Input rate < Output rate

$\Rightarrow$  FILO – FIT  $\times$  Input rate / Output rate > D

$\Rightarrow$  MIMO = min {LILO, FILO – FIT  $\times$  Input rate / Output rate} = D, correct

**Case 2cD: Discontiguous Frames, Input Rate < Output Rate, Compression of Gaps**

Figure A.2cD shows the flow for a zero-delay switch and a nonzero-delay switch.



**Figure A.2cD:** Discontiguous frames, Input rate < Output rate, Compression of gaps

In his case, the switch latency  $D$  is determined by the last bit delay. Here we have:

- FIFO >  $D$  incorrect; FIFO may be affected by changing the output rate or/and with changing the number of gaps on the output while the switch (latency) is unchanged. So, FIFO does not correctly represent the switch latency.

- LILO =  $D$ , correct

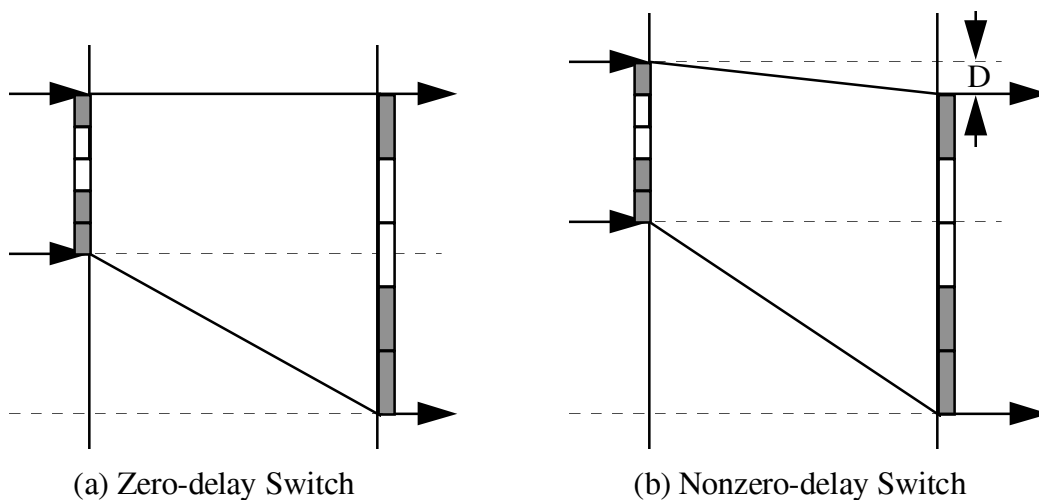
- Input rate < Output rate

$\Rightarrow$  FILO – FIT  $\times$  Input rate/Output rate >  $D$

$\Rightarrow$  MIMO = min {LILO, FILO – FIT  $\times$  Input rate/Output rate} =  $D$ , correct

**Case 3aD: Discontiguous Frames, Input Rate > Output Rate, No Change in Gaps**

Figure A.3aD shows the flow for a zero-delay switch and a nonzero-delay switch.



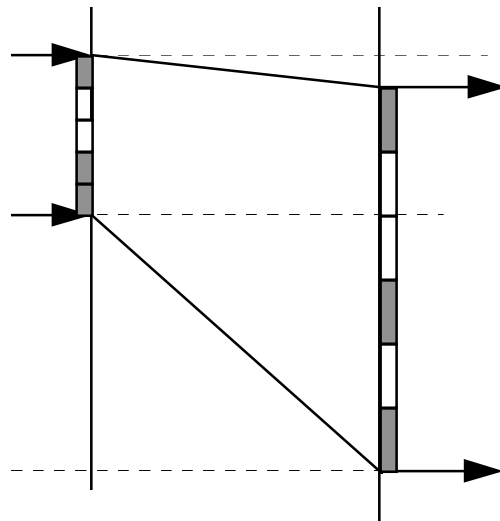
**Figure A.3aD:** Discontiguous frames, Input rate > Output rate, No Change in Gaps

This case is similar to cases 3aC and 3bC. The switch latency  $D$  is determined by a delay of the first bit. Here we have:

- FIFO =  $D$ , correct
- LILO >  $D$ , incorrect; Note that LILO may change with changing the output rate and not changing the switch otherwise. So, LILO does not correctly represent the switch latency.
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{FIT} \times \text{Input rate} / \text{Output rate} \} = \min \{ \text{LILO}, D \} = D$ , correct

### Case 3bD: Discontiguous Frames, Input Rate > Output Rate, Expansion of Gaps

Figure A.3bD shows the flow for a nonzero-delay switch, while a zero-delay switch with expansion of gaps is an impossible scenario.



**Figure A.3bD:** Discontiguous frames, Input rate > Output rate, Expansion of gaps

In this case, the switch latency  $D$  is given by:

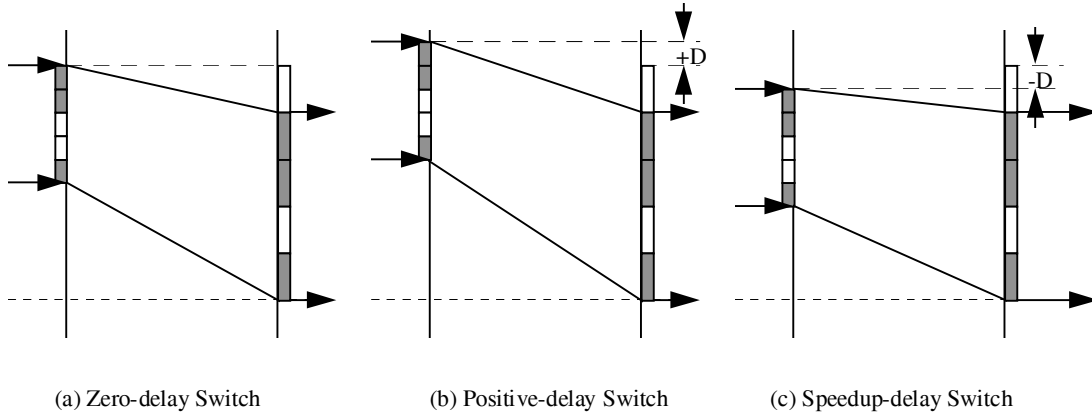
$D = \text{first bit delay} + \text{time of additional gaps on output}$

Here we have:

- FIFO <  $D$ , incorrect; FIFO is incorrect because it does not reflect expansion of gaps. Note for a nonzero-delay switch, FIFO may be even zero (the case of a zero delay for the first bit)
- LILO >  $D$ , incorrect; Here a similar argument applies as in Case 3aD for LILO incorrectly being influenced by the output rate, but with the observation that LILO correctly accounts for a time of additional gaps.
- MIMO =  $\min \{ \text{LILO}, \text{FILO} - \text{FIT} \times \text{Input rate} / \text{Output rate} \} = \min \{ \text{LILO}, D \} = D$ , correct

**Case 3cD: Discontiguous Frames, Input Rate > Output Rate, Compression of Gaps**

Figure A.3cD shows the flow for a zero-delay switch, the positive-delay switch and the speed-up switch.



**Figure A.3cD:** Discontiguous frames, Input rate > Output rate, Compression of gaps  
In this case the switch latency D is given by:

$$D = \text{first bit delay} - \text{time of additional gaps on output}$$

Three cases can be distinguished:

a. the case of a zero-delay switch, where:

$$\text{first bit delay} = \text{time of additional gaps on output}$$

b. the case of a positive-delay switch, where:

$$\text{first bit delay} > \text{time of additional gaps on output}$$

c. the case of a speedup-delay switch (a negative-delay switch), where:

$$\text{first bit delay} < \text{time of additional gaps on output}$$

- FIFO > D, incorrect; FIFO is incorrect because it does not reflect compression of gaps. Note that, here FIFO may be zero (the case of zero delay for the first bit) while the switch latency is negative

- LILO > D, incorrect; Here a similar argument applies as in Case 3aD for LILO incorrectly being influenced by the output rate, but with the observation that LILO correctly accounts for a time of missing gaps.

- MIMO = min {LILO, FILO - FIT × Input rate/Output rate} = min {LILO, D} = D, correct

In summary, MIMO latency is the only metric that applies to all cases.