Dynamic skew for MLD

Mark Gustlin - Cisco

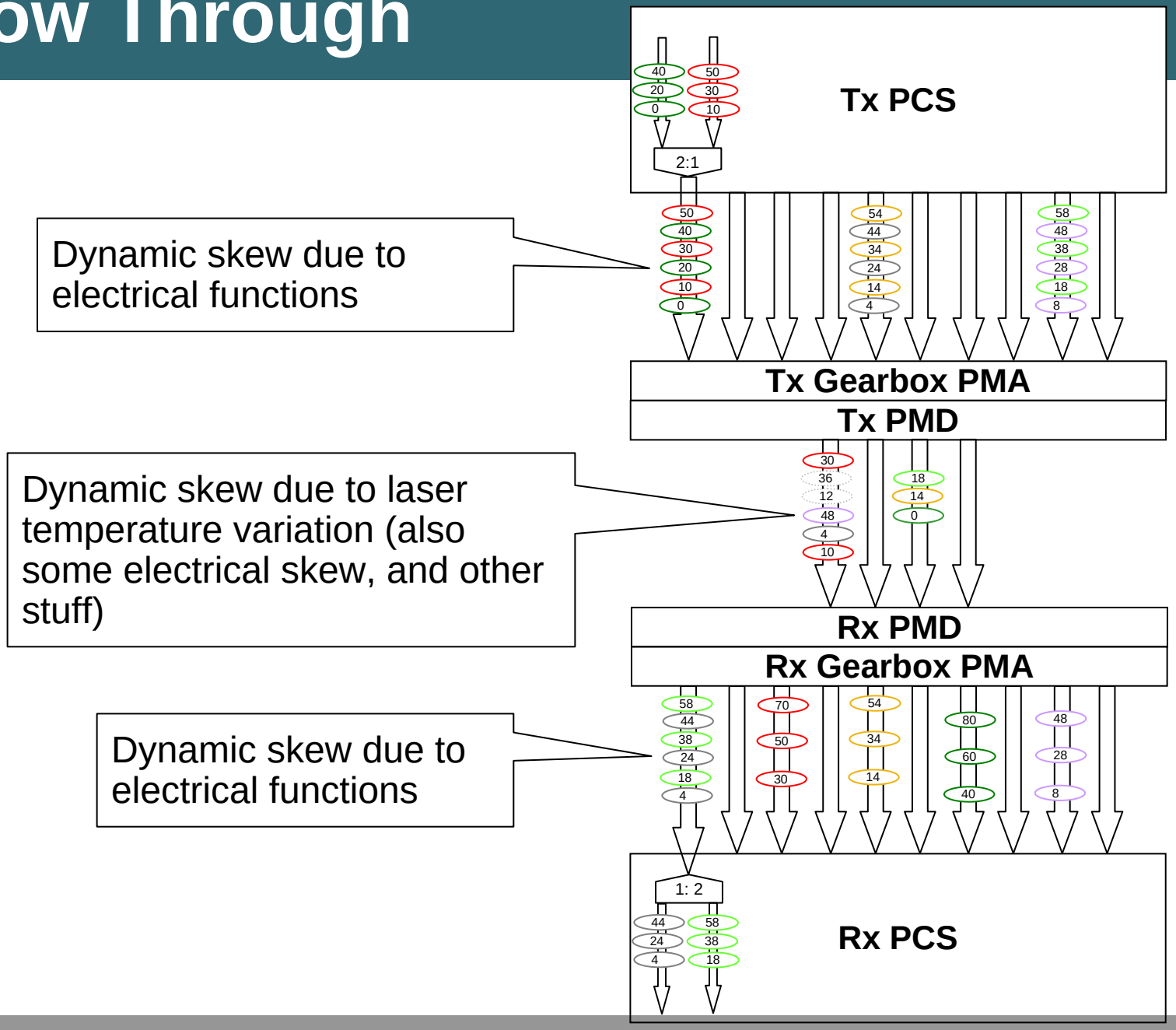
Pete Anslow – Nortel

IEEE P802.3ba, Munich, May 2008

Dynamic Skew Handling

- What happens in MLD if bit skew varies over time?
  - Two main reasons it will
    - Electrical effects
    - Laser temperature variation which changes the wavelengths and hence the relative lane propagation times.
- For the PMD variants with a gearbox, changes in skew must not change the route of a virtual lane through the link
- For PMDs without gearboxes, we need to total up the dynamic skew so that the rx PCS can compensate for it
- First question is how much the skew can vary over time?

Bit Flow Through

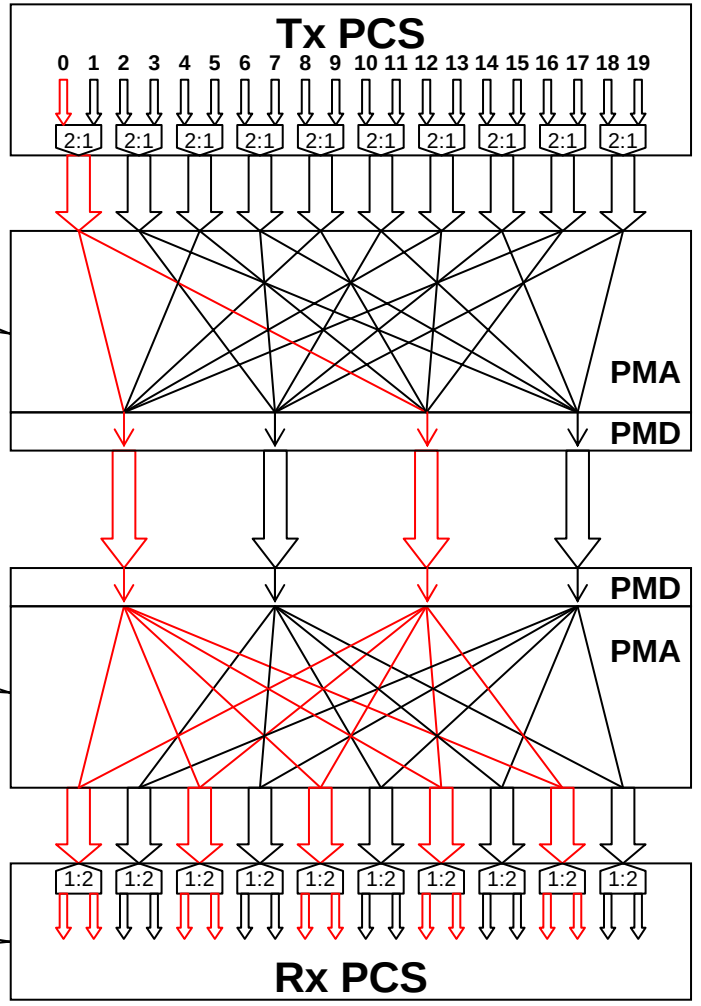


Possible Paths Through the Link

Skew on input electrical interface determines which optical lane VL 0 passes through

Skew on optical interface and previous electrical interface determines which output electrical lane VL 0 passes through

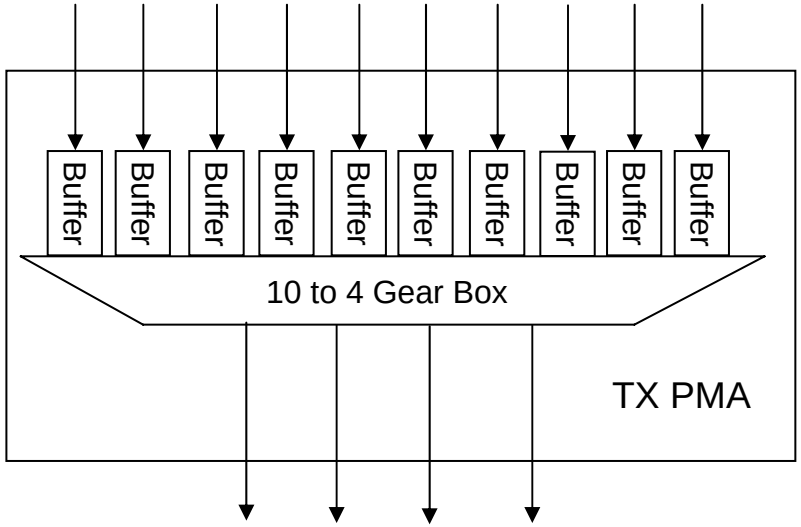
VL 0 can appear on any of the red outputs depending on the skew of the electrical and optical interfaces

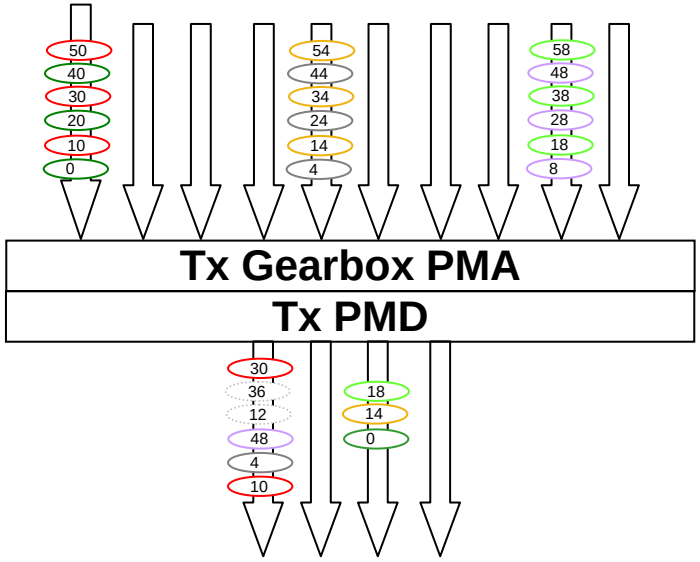


Note: These possible paths are based on a 10:4 and 4:10 function based on round-robin distribution. Other arrangements which give different paths are possible.

TX PMA Design

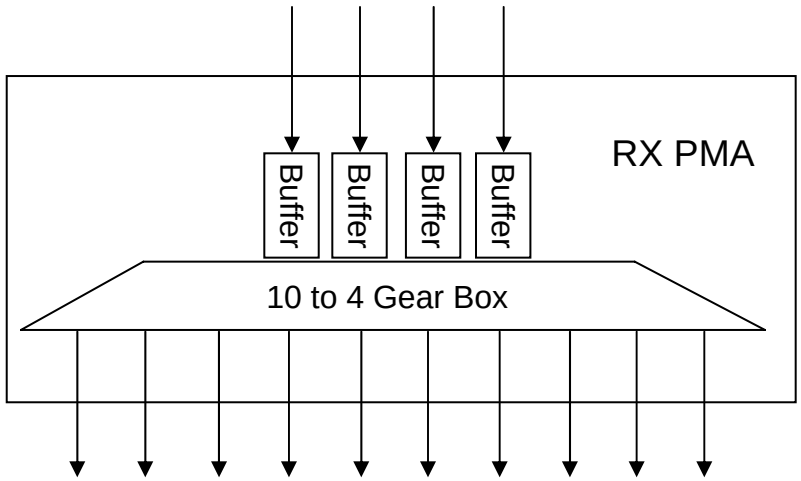
- For the case where the PMA is a gear box, 10 bits to 4
- Dynamic skew on electrical I/F will be very small, most skew is constant (PCB traces, ASIC skew etc.)
- All PMD lanes driven with the same clock, should need at most a few bits of retiming buffer per lane
- SFI5.2 relative wander is 1.5UI, assume we would need a similar amount for electrical functions

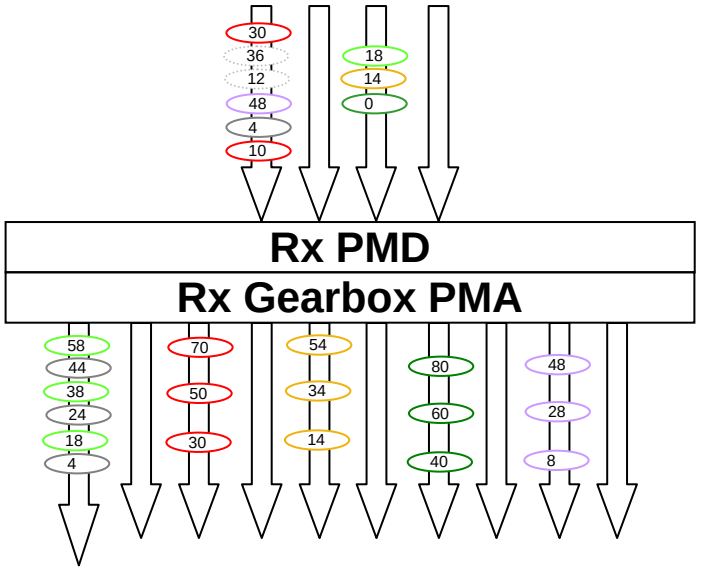




RX PMA Design

- We can have much more variable skew for this case
- You would need retiming buffers at around 2x the worst case delay variation expected to handle this case

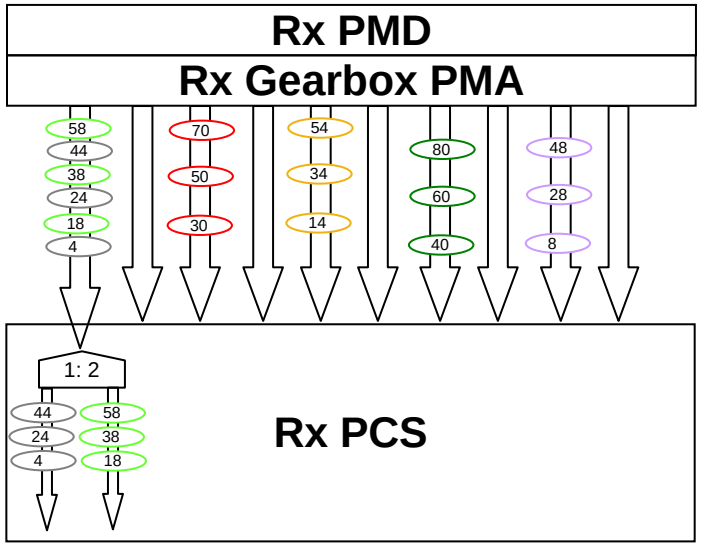




RX PCS Design

- Dynamic skew on electrical I/F will be very small, most skew is constant (PCB traces, ASIC skew etc.)
- All lanes driven with the same clock, might need at most a few bits of retiming buffer per lane?

  Could be included in the de-skew logic that must exist in the RX PCS



Dynamic Skew Handling

PMD	'Standard'	Skew change due to lambdas
100GE 10km	LAN WDM 800 GHz	53 ps (1.4 UI)
100GE 40km	LAN WDM 800 GHz	213 ps (5.5 UI)
100GE 10km	CWDM (1271 nm)	766 ps (19.7 UI)

Dynamic Skew – Electrical Functions

- Electrical functions will require some dynamic skew handling
- SFI-5.2 specifies 1.5UI of relative wander

  Relative Wander: Components of wander that are uncorrelated between any two in band signals

- Poll of three vendors: ~ 1UI - 1.5UI

  Should we round up to 2UI? How do you handle 0.5 anyhow?

- If we use an FPGA with external serdes does that increase this?

  One view is no, still about 1UI

- This would all be per hop
- Anyone else have a view?

Total Dynamic skew budget – 100GE SM

- First electrical interface: 2UI
- Optical interface: 2 + (6 or 20) = 8 or 22 UI total
- Second electrical interface: 2UI
  - Just added into the overall skew budget and handled by the rx MLD function
- Skew for speculative 80 km interface
  - Would have to be at 1550 nm because of fiber loss
  - Assume 800 GHz spacing in C band
  - Assume dispersion compensator does not remove skew between channels
  - Maximum optical skew variation = 244 UI (4 lanes at 25.8 GBd)

Total Dynamic skew budget – 100/40GE MM

- What about non muxed PMDs?
  - Is it 2UI of total dynamic skew, or 3 * this? Or something else?
  - Add this total to the static skew, and it is handled in the rx PCS
  - And what about the temperature variation on the VCSELs?
- For Parallel Fiber Multi-mode:
  - First electrical I/F: 1 UI
  - Optical interface: 1 UI + stress + DMD + λ change
  - Second electrical I/F: 2 UI
  - Stress + DMD + λ change = $4.4^{1} + 0.3^{1} + 2.1^{2}$ = 6.8 ps/m
  - 680 psec for 100m, or ~7.0 UI
  - So total dynamic skew = 1 + 1 + 7.0 + 2 = 11 UI
  - [For 300m interface would be 1 + 1 + 7x3 + 2 = 25 UI]

1 – kolesar_01_0508

2 – 840 to 860 nm, λ0 = 1318nm, S0 = 0.102 ps/(nm^2.km)

Total Dynamic skew budget – 40GE SM

- 40GE 10km Single Mode
  - If optics is serial dynamic skew only on electrical interfaces
  - If optics is 4 x 10G could be LAN WDM or more likely CWDM
  - CWDM case gives worst dynamic skew
- 4 x 10G CWDM case:
  - First electrical I/F: 1UI
  - Optical interface: 1UI + λ change
  - second electrical I/F: 2UI
  - λ change = 7.9[1] UI
  - So total dynamic skew = 1 + 1 + 7.9 + 2 = 11.9UI

1 – Fibre characteristics spreadsheet version 3 with lane 0 at 1271 nm
http://www.ieee802.org/3/ba/public/tools/Fibre_characteristics_V_3_0.xls

Summary – required dynamic skew tolerance

PMD	'Standard'	Tx PMA (Elec.)	Optical	Rx PCS (Elec.)
100GE 40 km	LAN WDM	194ps (2UI)	310ps (8UI)	194ps (2UI)
100GE 10 km	LAN WDM	194ps (2UI)	155ps (4UI)	194ps (2UI)
100GE 10 km	CWDM	194ps (2UI)	853ps (22UI)	194ps (2UI)
100GE 100 m	// fibers	-	-	1.07ns (11UI)
[100GE 300 m]	[// fibers]	-	-	[2.42ns (25UI)]
[100GE 80km]	[1550nm]	[194ps (2UI)]	[9.53ns (246UI)]	[194ps (2UI)]
40GE 10 km	CWDM	-	-	1.16ns (12UI)
40GE 100 m	// fibers	-	-	1.07ns (11UI)
[40GE 300 m]	[// fibers]	-	-	[2.42ns (25UI)]
Overall Recommendation?		2UI (per physical lane)	30UI (per physical lane)	30UI (15UI per virtual lane for 100GbE)

More on Dynamic Skew

- Ok, I have the dynamic skew numbers, now what?
- For designs with a PMA gearbox (m != n), the gearbox has a wander buffer per input lane

  Size is 2x the max dynamic skew for that corresponding path

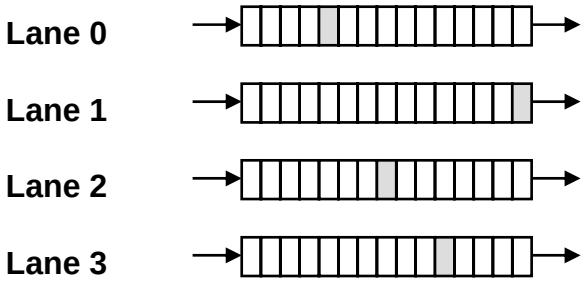
  Start reading out of the wander buffers when they are half full
- For designs without a PMA gearbox (m=n), the maximum skew already includes the dynamic skew numbers, your receive PCS input FIFOs/buffers need to be able to track the dynamic skew

  Note: An increase in maximum skew capability does not impact latency, only buffer depth. An increase in dynamic skew capability does increase latency because you must wait to start reading out data from the receive FIFOs until there is enough data in the least filled FIFO to allow for the maximum dynamic skew variation that we expect for a worst case interface.

Wander buffer (m != n)

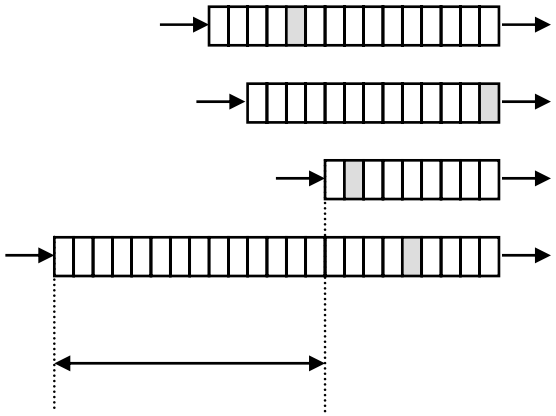
Example – 4 x 25G receiver gearbox

When link is established



All buffers half full

Some time later



Dynamic skew

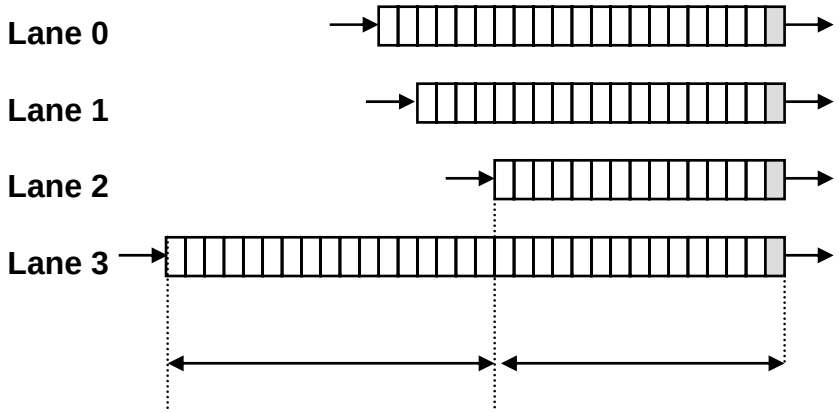
Takes no account of the lane markers

= first bit of the lane markers

De-skew buffer

Example – 4 x 10G receiver PCS

When link is established

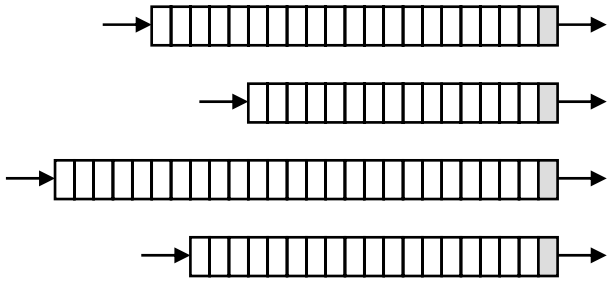


Initial skew

All buffers must contain at least as much as the max dynamic skew

Lane markers read out of each lane at the same time

Some time later



Dynamic skew has changed the fill level of the buffers

The maximum skew is the largest difference in the fill level of the buffers at any time