



Protocol Analysis Methodologies of Deadlock Scenarios

Isaac Livny
LSI Corporation



Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express® architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe® link level
 - ✓ Example: Credit leakage

Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe link level
 - ✓ Example: Credit leakage

General

- Deadlock issues unavoidable for bus architectures
- Occur when multiple agents mutually depend on response from each other or common resource
 - ✓ Access to a resource cannot complete until other agent has completed its access to the same resource
- Conventional PCI back-off schemes
 - ✓ Disconnect
 - ✓ Retry
- PCI Express replaces retry and disconnect termination schemes with split transaction protocols.
 - ✓ Less immune to deadlock situations due to their complexity.

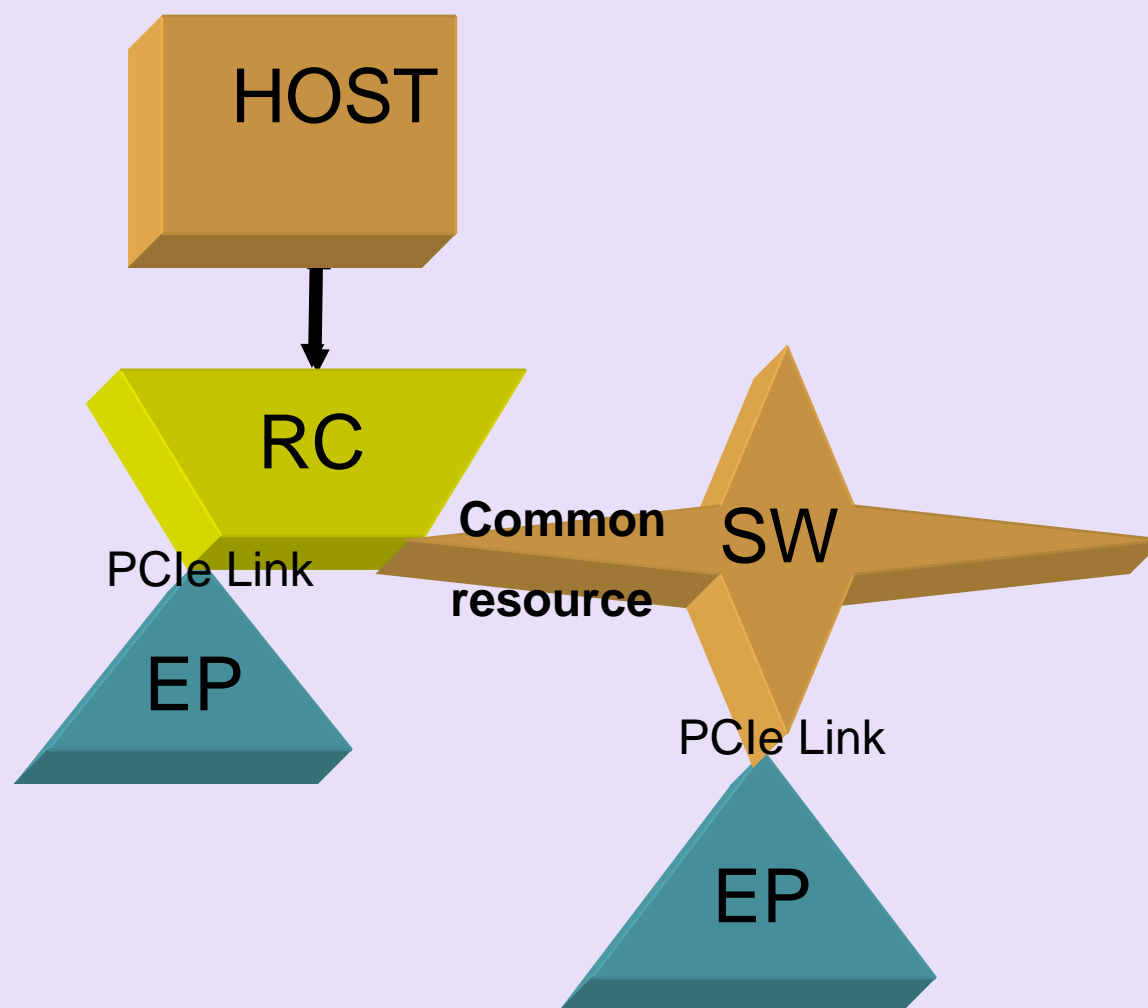
Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe link level
 - ✓ Example: Credit leakage

Deadlock scenarios

- 4 types of deadlock scenarios
 - characterized by where they can occur in the PCI Express fabric
 - ✓ Within an endpoint
 - ✓ Associated with a particular PCI Express link
 - ✓ Associated with a resource common to a number of PCI express devices or within a PCI express switch
 - ✓ At the Root Complex level, within the chipset or at the system level

PCI Express Fabric



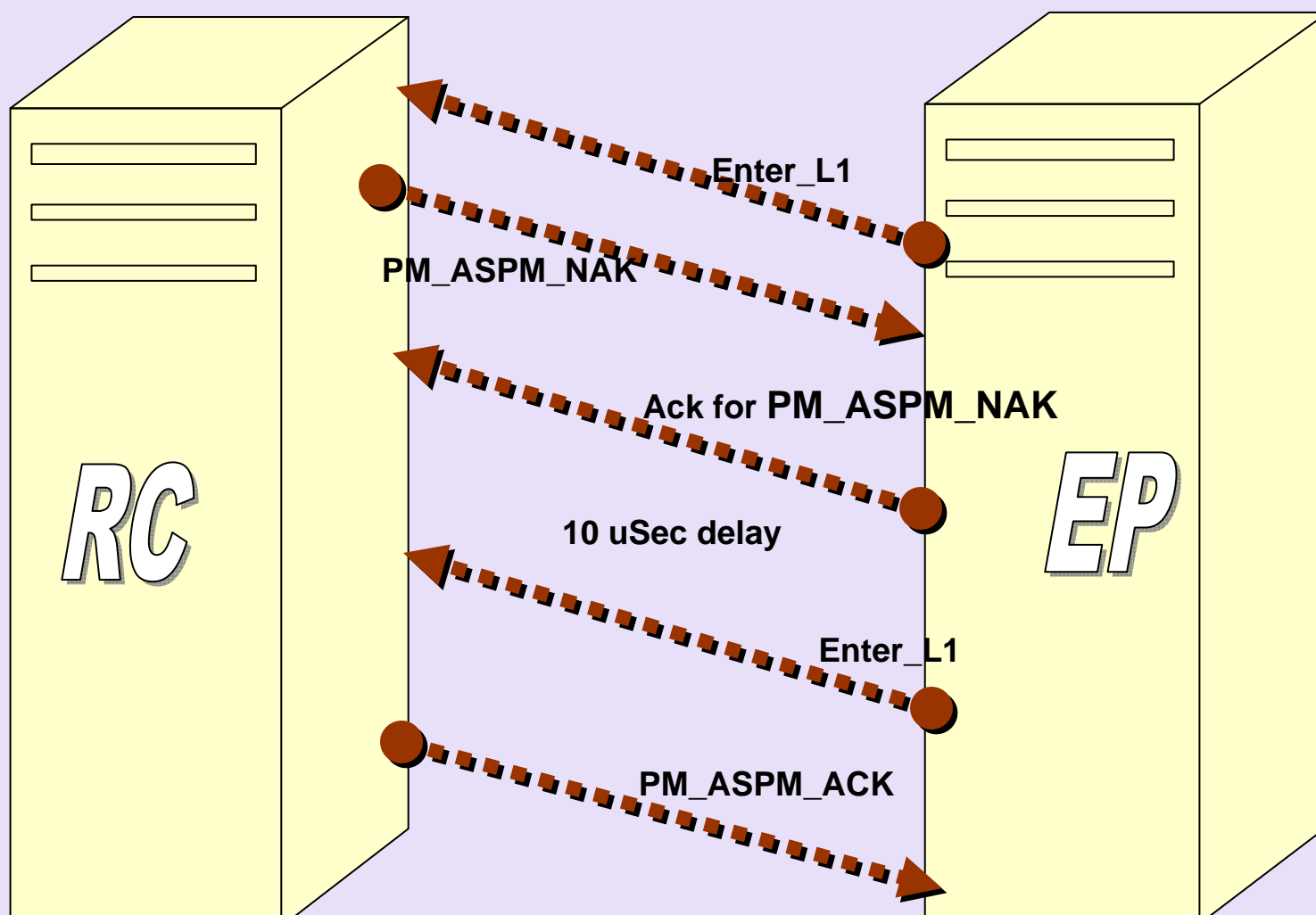
Deadlocks and avoidance

- PCIe protocols affected
 - ✓ ASPM L1 entry and exit protocols
 - ✓ PME fence mechanism
 - ✓ Error reporting schemes
- Deadlock avoidance schemes
 - ✓ At the device level
 - ✓ Root Complex level
 - ✓ System level
 - ✓ Possible specification enhancements

Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe link level
 - ✓ Example: Credit leakage

ASPM L1 Re-entry delay



Deadlock example at common resource level

Packet	R←	2.5	TLP	Cpl	CplID	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Register Data
140713	x1		1138		10:01010	000:00:0	1	004:00:0	SC	0	4	0x00	0x10110042
Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Time Delta						
140714	x1				1138	0x14AD	536.000 ns						
Packet	R←	2.5	DLLP	UpdateFC-NP	VC ID	HdrFC	16-bit CRC	16	Time Delta				
140716	x1				0	105	6C	8.568 μs					
Packet	R←	2.5	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	CRC 16	Idle				
140719	x1				0	68	81	0x6285	0.000 ns				
Packet	R←	2.5	DLLP	UpdateFC-NP	VC ID	HdrFC	DataFC	CRC 16	Time Delta				
140720	x1				0	105	146	0x726C	2.704 μs				
Packet	R→	2.5	DLLP	UpdateFC-NP	VC ID	HdrFC	DataFC	CRC 16	Idle				
140723	x1				0	68	0	0x2D8B	96.000 ns				
Packet	R→	2.5	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	CRC 16	Time Delta				
140724	x1				0	75	136	0xFFFF1	13.768 μs				
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140729	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140730	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140731	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140732	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140733	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140734	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140735	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Time Delta					
140736	x1						0xEB05	24.000 ns					
Packet	R→	2.5	TLP	Msg	Msg	Msg Routing	RequesterID	Tag	Message Code	LCRC	Time Delta		
140737	x1		1191		01:10100	Local	000:00:0	0	PME_Active_State_Nak	0x0B870B0C	8.000 ns		
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140738	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140739	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140740	x1						0xEB05	0.000 ns					
Packet	R←	2.5	DLLP	PM_Active_State_Request_L1			CRC 16	Idle					
140741	x1						0xEB05	0.000 ns					

ASPM L1 re-entry delay deadlock

Packet 140736	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Time Delta 24.000 ns		
Packet 140737	R→	2.5 x1	TLP	Msg	Msg 01:10100	Msg Routing Local	RequesterID 000:00:0	Tag 0	Message Code PME_Active_State_Nak	
Packet 140738	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140739	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140740	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140741	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140742	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Time Delta 56.000 ns		
Packet 140744	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140745	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140746	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140747	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140748	R→	2.5 x1	DLLP	ACK	AckNak_Seq_Num 1191	CRC 16 0x3C7A	Idle 0.000 ns			
Packet 140749	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 912.000 ns		
Packet 140751	R→	2.5 x1	DLLP	UpdateFC-P	VC ID 0	HdrFC 69	DataFC 81	CRC 16 0x8EEB	Time Delta 13.232 µs	
Packet 140756	R→	2.5 x1	DLLP	UpdateFC-P	VC ID 0	HdrFC 75	DataFC 136	CRC 16 0xFFFF1	Idle 96.000 ns	
Packet 140757	R→	2.5 x1	DLLP	UpdateFC-NP	VC ID 0	HdrFC 68	DataFC 0	CRC 16 0x2D8B	Time Delta 1.264 µs	
Packet 140759	R→	2.5 x1	DLLP	UpdateFC-P	VC ID 0	HdrFC 69	DataFC 81	CRC 16 0x8EEB	Idle 0.000 ns	
Packet 140760	R→	2.5 x1	DLLP	UpdateFC-NP	VC ID 0	HdrFC 105	DataFC 146	CRC 16 0x726C	Time Delta 16.600 µs	
Packet 140767	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		
Packet 140768	R→	2.5 x1	DLLP	PM_Active_State_Request_L1			CRC 16 0xEB05	Idle 0.000 ns		

L1 Re-entry delay deadlock

- ASPM L1 enabled at endpoint, disabled at root port
- Traffic recess – EP requests L1, Root port NAKs
- EP Acks PM_Active_State_Ack
- EP required ASPM L1 re-entry delay > 10 usec
- If re-entry is too early – RC interprets as old request
- RC resources congestion blocks other traffic
- Deadlock situation arises.
- Deadlock can also happen with 10 usec delay met
- RC should flash L1 requests not to consume resources

Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe link level
 - ✓ Example: Credit leakage

Deadlock example at RC/Chipset level

- Correct D3 cold transition protocol
 - ✓ System software puts device to D3
 - ✓ Endpoint initiates L1 transition

Packet	R→	2.5	TLP	Cmd	CmdWr0	RequesterID	Tag	DeviceID	Register	1st BE	Register Data
1484067	R→	x1	2814	Cmd	10:00100	000:00:0	0	002:00:0	0x044	0011	0x00000003
Packet	R→	2.5	TLP	Cmd	Cmd	RequesterID	Tag	CompleterID	Status	Byte Cnt	Lwr Addr
1484069	R→	x1	1805	Cmd	00:01010	000:00:0	0	002:00:0	SC	4	0x00
Packet	R→	2.5	DLLP	PM_Enter_L1						CRC 16	
1484071	R→	x1	DLLP	PM_Enter_L1						0x65AD	
Packet	R→	2.5	DLLP	PM_Enter_L1						CRC 16	
1484072	R→	x1	DLLP	PM_Enter_L1						0x65AD	
Packet	R→	2.5	DLLP	PM_Enter_L1						CRC 16	
1484073	R→	x1	DLLP	PM_Enter_L1						0x65AD	

Correct D3 cold transition protocol

Packet	R→	2.5	DLLP	PM_Request_Ack	CRC 16	Time Delta
1484089		x1			0x930C	16.000 ns
Packet number (Post-Trigger)						
Packet	R→	2.5	DLLP	PM_Enter_L1	CRC 16	Idle
1484090		x1			0x65AD	0.000 ns
Packet	R→	2.5	DLLP	PM_Enter_L1	CRC 16	Idle
1484091		x1			0x65AD	0.000 ns
Packet	R→	2.5	DLLP	PM_Enter_L1	CRC 16	Time Delta
1484092		x1			0x65AD	16.000 ns
Packet	R→	2.5	DLLP	PM_Request_Ack	CRC 16	Time Delta
1484093		x1			0x930C	16.000 ns
Packet	R→	2.5	DLLP	PM_Enter_L1	CRC 16	Time Delta
1484094		x1			0x65AD	16.000 ns
Packet	R→	2.5	DLLP	PM_Request_Ack	CRC 16	Time Delta
1484095		x1			0x930C	16.000 ns
Packet	R→	2.5	DLLP	PM_Enter_L1	CRC 16	Idle
1484096		x1			0x65AD	32.000 ns
Packet	R→	2.5	EIOS	COM EIOS Symbols	Time Delta	
1484097		x1		K28.5 K28.3 K28.3 K28.3	16.000 ns	
Packet	R→	2.5	DLLP	PM_Request_Ack	CRC 16	Idle
1484098		x1			0x930C	0.000 ns
Packet	R→	2.5	DLLP	PM_Request_Ack	CRC 16	Idle
1484099		x1			0x930C	32.000 ns
Packet	R→	2.5	DLLP	PM_Request_Ack	CRC 16	Idle
1484100		x1			0x930C	0.000 ns
Packet	R→	2.5	DLLP	PM_Request_Ack	CRC 16	Idle
1484101		x1			0x930C	648.000 ns
Packet	R→	2.5	EIOS	COM EIOS Symbols	Time Delta	
1484104		x1		K28.5 K28.3 K28.3 K28.3	1.320 μs	
Packet	R→	2.5	TS1	COM Link Lane N_FTS Training Control Data Rate TS1 Symbols	Idle	
1484109		x1		K28.5 0 0 24 0 0 0 0 0 2.5 GT/s D10.2 ...	0.000 ns	
Packet	R→	2.5	TS1	COM Link Lane N_FTS Training Control Data Rate TS1 Symbols	Idle	
1484110		x1		K28.5 0 0 24 0 0 0 0 0 2.5 GT/s D10.2 ...	0.000 ns	

Correct D3 cold transition protocol

- D3 cold transition protocol - continued
 - ✓ L1 exit, Link transitions back to L0
 - ✓ System software/RC initiate PME fence mechanism
 - ✓ Endpoint Acks PME Shutdown message
 - ✓ Transition L2/L3 ready

PME Fence mechanism

Packet	R→	2.5	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Idle				
1484949		x1		K28.5	0	0	24	0 0 0 0	2.5 GT/s	D05.2 ...	0.000 ns				
Packet	R→	2.5	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Idle				
1484950		x1		K28.5	0	0	24	0 0 0 0	2.5 GT/s	D05.2 ...	0.000 ns				
Packet	R→	2.5	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Idle				
1484951		x1		K28.5	0	0	24	0 0 0 0	2.5 GT/s	D05.2 ...	0.000 ns				
Packet	R→	2.5	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Idle				
1484952		x1		K28.5	0	0	24	0 0 0 0	2.5 GT/s	D05.2 ...	180.000 ns				
Packet	R→	2.5	TLP	Cfg	CfgRd0	RequesterID	Tag	DeviceID	Register	1st BE	LCRC	Time Delta			
1484953		x1	2815	00:00100	000:00:0	1	002:00:0	0x044	0011	0xB1E45B7D	292.000 ns				
Packet	R←	2.5	TLP	Cpl	CplID	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Register Data	LCRC	
1484955		x1	1806	10:01010	000:00:0	1	002:00:0	SC	0	4	0x00	0x00000003	0x341B74A8		
Packet	R→	2.5	TLP	Cfg	CfgRd0	RequesterID	Tag	DeviceID	Register	1st BE	LCRC	Time Delta			
1484957		x1	2816	00:00100	000:00:0	0	002:00:0	0x004	0011	0x541B3A60	324.000 ns				
Packet	R←	2.5	TLP	Cpl	CplID	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Status	Command	LCRC
1484959		x1	1807	10:01010	000:00:0	0	002:00:0	SC	0	4	0x00	0x0010	0x0106	0x868B509	
Packet	R→	2.5	TLP	Cfg	CfgWr0	RequesterID	Tag	DeviceID	Register	1st BE	Status	Command	LCRC	Time Delta	
1484961		x1	2817	10:00100	000:00:0	1	002:00:0	0x004	0011	0x0000	0x0100	0xE9889B5A	344.000 ns		
Packet	R←	2.5	TLP	Cpl	Cpl	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	LCRC	Time Delta	
1484963		x1	1808	00:01010	000:00:0	1	002:00:0	SC	0	4	0x00	0x400DF844	5.389 sec		
Packet	R→	2.5	TLP	Msg	Msg	Msg Routing	RequesterID	Tag	Message Code	LCRC	Time Delta				
1484965		x1	2818	01:10011	Broadcast	000:00:0	0	PME_Turn_Off	0xDF294396	248.000 ns					
Packet	R←	2.5	TLP	Msg	Msg	Msg Routing	RequesterID	Tag	Message Code	LCRC	Idle				
1484967		x1	1809	01:10101	Gathered	002:00:0	0	PME_TO_Ack	0xD14D2D5F	0.000 ns					
Packet	R←	2.5	TLP	Msg	Msg	Msg Routing	RequesterID	Tag	Message Code	LCRC	Idle				
1484968		x1	1810	01:10101	Gathered	002:00:0	0	PME_TO_Ack	0x5516B70C	456.000 ns					
Packet	R←	2.5	DLLP	PM_Enter_L23	CRC 16	Idle									
1484970		x1		0x1055	0.000 ns										
Packet	R←	2.5	DLLP	PM_Enter_L23	CRC 16	Idle									
1484971		x1		0x1055	0.000 ns										

■ L1 exit, Link back to L0

■ PME mechanism

■ L2/L3 ready trans

■ L1 exit, Link back to L0

■ PME Fence mechanism

■ L2/L3 ready transition

Going to D3 cold

- L23 entry message ack'ed
- Link goes to electrical idle

Packet 1484985	R←	2.5 x1	DLLP	PM_Enter_L23	CRC 16 0x1055	Time Delta 0.000 ns
Packet 1484986	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 32.000 ns
Packet 1484987	R←	2.5 x1	DLLP	PM_Enter_L23	CRC 16 0x1055	Idle 32.000 ns
Packet 1484988	R←	2.5 x1	EIOS	COM K28.5	EIOS Symbols K28.3 K28.3 K28.3	Time Delta 0.000 ns
Packet 1484989	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 0.000 ns
Packet 1484991	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 88.000 ns
Packet 1484992	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 0.000 ns
Packet 1484993	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 616.000 ns
Packet 1484996	R→	2.5 x1	EIOS	COM K28.5	EIOS Symbols K28.3 K28.3 K28.3	

Power sequencing deadlock

- Power on
- Wait for the BIOS to kick in
- Toggle main supply/RESET
- Windows starts
- Disable the device in device manager, or go to Hibernate
- Machine hangs

Power sequencing hang-up

Packet	R→	2.5	TLP	Cfg	CfgWr0	RequesterID	Tag	DeviceID	Register	1st BE	Register Data
4320		x1	2773		10:00100	000:00:0	1	003:00:0	0x044	1111	0x00008003

Packet	R←	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Idle
4321		x1			2773	0x3F37	80.000 ns

Packet	R←	2.5	TLP	Cpl	Cpl	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lv
4322		x1	3107		00:01010	000:00:0	1	003:00:0	SC	0	4	

Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Time Delta
4323		x1			3107	0x8519	216.000 ns

Packet	R←	2.5	DLLP	PM_Enter_L23		CRC 16	Idle
4324		x1				0x1055	0.000 ns

Packet	R←	2.5	DLLP	PM_Enter_L23		CRC 16	Idle
4325		x1				0x1055	0.000 ns

Packet	R←	2.5	DLLP	PM_Enter_L23		CRC 16	Idle
4326		x1				0x1055	0.000 ns

Packet	R←	2.5	DLLP	PM_Enter_L23		CRC 16	Idle
4327		x1				0x1055	0.000 ns

Packet	R←	2.5	DLLP	PM_Enter_L23		CRC 16	Time Delta
4328		x1				0x1055	8.000 ns

Packet	R→	2.5	DLLP	PM_Request_Ack		CRC 16	Time Delta
4329		x1				0x930C	24.000 ns

Power sequencing hang-up

Packet 4338	R←	2.5 x1	DLLP	PM_Enter_L23	CRC 16 0x1055	Time Delta 8.000 ns
Packet 4339	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 24.000 ns
Packet 4340	R←	2.5 x1	DLLP	PM_Enter_L23	CRC 16 0x1055	Time Delta 8.000 ns
Packet 4341	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 0.000 ns
Packet 4342	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 24.000 ns
Packet 4343	R←	2.5 x1	EIOS	COM K28.5	EIOS Symbols K28.3 K28.3 K28.3	Time Delta 8.000 ns
Packet 4344	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 24.000 ns
Packet 4345	R←	2.5 x1	Link Event Link Down	Time Delta 8.000 ns		
Packet 4346	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 0.000 ns
Packet 4347	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 16.000 ns
Packet 4348	R←	2.5 x1	EIOS	COM K28.5	EIOS Symbols K28.3 K28.3 K28.3	Time Delta 824.000 ns

PME Fence mechanism deadlock summary

- Device transitions to L23 ready instead of L1

Packet	R→	2.5	TLP	Cfg	CfgWr0	RequesterID	Tag	DeviceID	Register	1st BE	Register Data	
4320		x1	2773		10:00100	000:00:0	1	003:00:0	0x044	1111	0x00008003	
Packet	R→	2.5	DLLP	ACK	AdkNak_Seq_Num	CRC 16	Idle					
4321		x1			2773	0x3F37	80.000 ns					
Packet	R→	2.5	TLP	Cpl	Cpl	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr
4322		x1	3107		00:01010	000:00:0	1	003:00:0	SC	0	4	0
Packet	R→	2.5	DLLP	ACK	AdkNak_Seq_Num	CRC 16	Time Delta					
4323		x1			3107	0x6519	216.000 ns					
Packet	R→	2.5	DLLP	PM_Enter_L23			CRC 16	Idle				
4324		x1					0x1055	0.000 ns				
Packet	R→	2.5	DLLP	PM_Enter_L23			CRC 16	Idle				
4325		x1					0x1055	0.000 ns				
Packet	R→	2.5	DLLP	PM_Enter_L23			CRC 16	Idle				
4326		x1					0x1055	0.000 ns				
Packet	R→	2.5	DLLP	PM_Enter_L23			CRC 16	Idle				
4327		x1					0x1055	0.000 ns				
Packet	R→	2.5	DLLP	PM_Enter_L23			CRC 16	Time Delta				
4328		x1					0x1055	8.000 ns				
Packet	R→	2.5	DLLP	PM_Request_Ack			CRC 16	Time Delta				
4329		x1					0x930C	24.000 ns				

PME Fence mechanism deadlock summary

- Initial boot-power-sequence leaves device at D3-cold ready state

Packet 4338	R←	2.5 x1	DLLP	PM_Enter_L23	CRC 16 0x1055	Time Delta 8.000 ns
Packet 4339	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 24.000 ns
Packet 4340	R←	2.5 x1	DLLP	PM_Enter_L23	CRC 16 0x1055	Time Delta 8.000 ns
Packet 4341	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 0.000 ns
Packet 4342	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 24.000 ns
Packet 4343	R←	2.5 x1	EIOS	COM K28.5	EIOS Symbols K28.3 K28.3 K28.3	Time Delta 8.000 ns
Packet 4344	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 24.000 ns
Packet 4345	R←	2.5 x1	Link Event Link Down	Time Delta 8.000 ns		
Packet 4346	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 0.000 ns
Packet 4347	R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Idle 16.000 ns
Packet 4348	R→	2.5 x1	EIOS	COM K28.5	EIOS Symbols K28.3 K28.3 K28.3	Time Delta 824.000 ns

PME fence hangup analysis

- Host tries to communicate while device in D3 cold
- RC blocks transactions since it has transitioned the link to L23
- Machine hangs on no completions coming back from the device
- Power down state machine should not transition to L23 before receiving PME_Turn_Off message and fully executing the PME fence mechanism

Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe link level
 - ✓ Example: Credit leakage

Deadlock at the device level: Handling of illegitimate NAK

■ Compliant NAK

Packet	R→	2.5	TLP	Mem	MRd(32)	TC	TD	EP	Attributes	Length	RequesterID	Tag	Address	1st BE	Last BE	LCRC	Time Delta	
93203		x1	3835		00:00000	0	0	0	01	1	000:00:0	1	FEAFF180	1111	0000	0xBD332CCA	272.000 ns	
Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Idle											
93204		x1			3835	0x1139	56.000 ns											
Packet	R→	2.5	TLP	Cpl	CplID	TC	TD	EP	Attributes	Length	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Data
93205		x1	510		10:01010	0	0	0	01	1	000:00:0	1	002:00:0	SC	0	4	0x00	11040000
Packet	R→	2.5	TLP	Cpl	CplID	TC	TD	EP	Attributes	Length	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Data
93207		x1	510		10:01010	0	0	0	01	1	000:00:0	1	002:00:0	SC	0	4	0x00	11040000
Packet	R→	2.5	TLP	Cpl	CplID	TC	TD	EP	Attributes	Length	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Data
93210		x1	510		10:01010	0	0	0	01	1	000:00:0	1	002:00:0	SC	0	4	0x00	11040000
Packet	R→	2.5	DLLP	NAK	AckNak_Seq_Num	CRC 16	Time Delta											
93211		x1			509	0xCAC4	280.000 ns											
Packet	R→	2.5	TLP	Cpl	CplID	TC	TD	EP	Attributes	Length	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Data
93214		x1	510		10:01010	0	0	0	01	1	000:00:0	1	002:00:0	SC	0	4	0x00	11040000
Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Idle											
93215		x1			510	0xC28F	480.000 ns											
Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Idle											
93216		x1			510	0xC28F	1.312 µs											

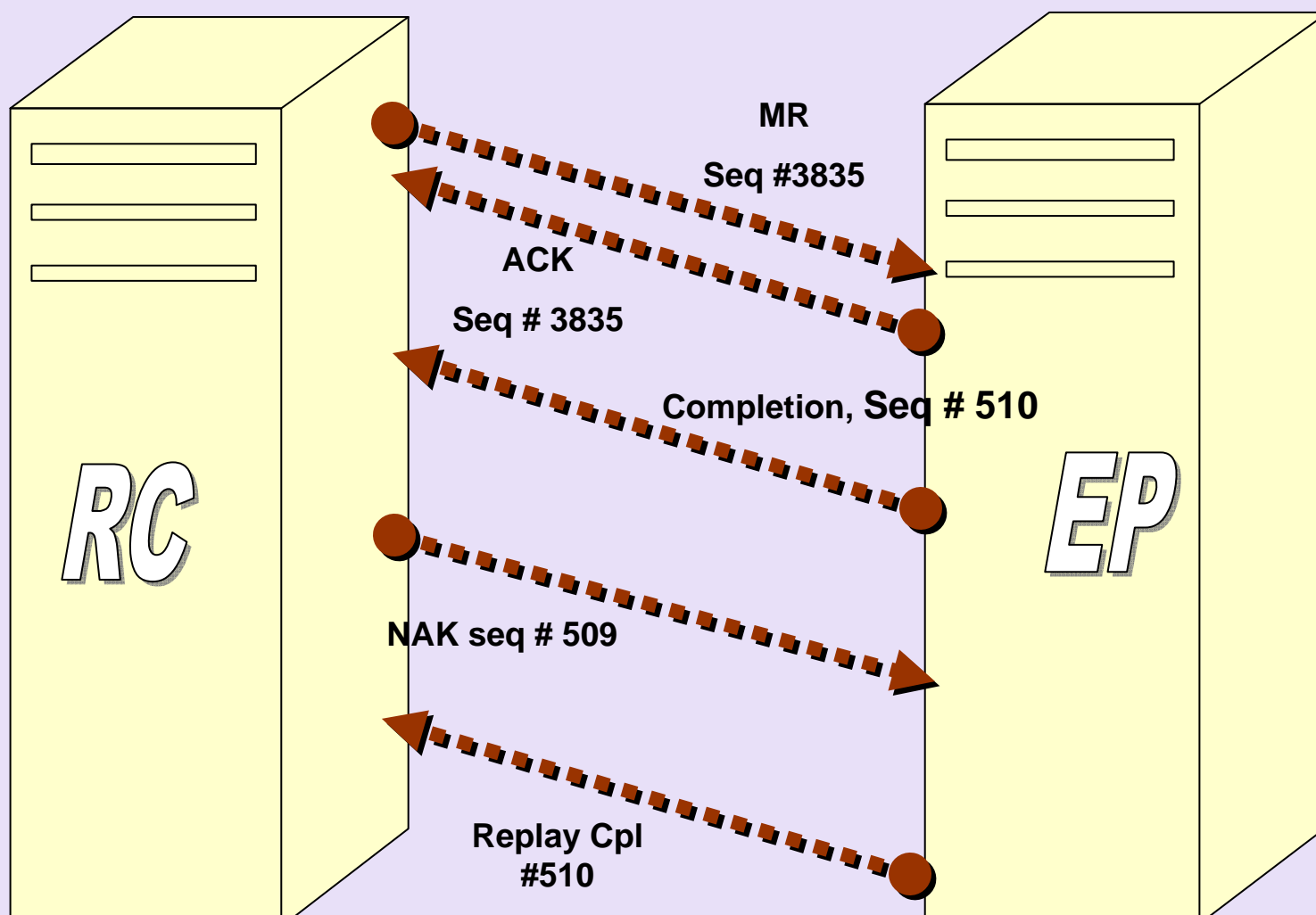
✓ TLP # 510 Nak'd by NAK

✓ TLP #510 is replayed

✓ TLP # 510 Nak'd by NAK 509

✓ TLP #510 is replayed

Compliant NAK



Compliant NAK

Packet	R→	2.5	TLP	Mem	MWr(32)	RequesterID	Tag	Address	1st BE	Last BE	Data	LCRC	Time Delta
3213		x1	868		10:00000	002:00:0	0	06037880	1111	1111	32 dwords	0xC872881E	952.000 ns

Packet	R→	2.5	DLLP	ACK	AdtNak_Seq_Num	CRC 16	Time Delta
3214		x1			868	0x38A9	376.000 ns

✓ TLP # 867, 868 Nak'd by NAK 866

Packet	R→	2.5	TLP	Mem	MWr(32)	RequesterID	Tag	Address	1st BE	Last BE	Data	LCRC	Idle
3215		x1	867		10:00000	002:00:0	0	06037900	1111	1111	32 dwords	0x59E9112C	720.000 ns

Packet	R→	2.5	TLP	Mem	MWr(32)	RequesterID	Tag	Address	1st BE	Last BE	Data	LCRC	Time Delta
3216		x1	868		10:00000	002:00:0	0	06037980	1111	1111	32 dwords	0x6F85E849	1.128 µs

Packet	R→	2.5	DLLP	NAK	AdtNak_Seq_Num	CRC 16	Time Delta
3217		x1			868	0xDDCE	168.000 ns

✓ TLP #859, 868 are replayed

Packet	R→	2.5	TLP	Mem	MWr(32)	RequesterID	Tag	Address	1st BE	Last BE	Data	LCRC	Idle
3218		x1	869		10:00000	002:00:0	0	06037A00	1111	1111	32 dwords	0x46BE0908	0.000 ns

Packet	R→	2.5	TLP	Mem	MWr(32)	RequesterID	Tag	Address	1st BE	Last BE	Data	LCRC	Idle
3219		x1	867		10:00000	002:00:0	0	06037900	1111	1111	32 dwords	0x59E9112C	0.000 ns

Packet	R→	2.5	TLP	Mem	MWr(32)	RequesterID	Tag	Address	1st BE	Last BE	Data	LCRC	Idle
3220		x1	868		10:00000	002:00:0	0	06037980	1111	1111	32 dwords	0x6F85E849	24.000 ns

Incompliant NAK

Packet	R→	2.5	TLP	Mem	MRd(32)	RequesterID	Tag	Address	1st BE	Last BE	LCRC	Time Delta
75124		x1	576		00:00000	000:00:0	1	FEAFF180	1111	0000	0x0BD503A0	272.000 ns

■ Undelivered TLP # 3

Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Idle
75125		x1			576	0xC855	56.000 ns

■ Nak'd by NAK 3852

Packet	R→	2.5	TLP	Cpl	CpID	RequesterID	Tag	CompleterID	Status	BCM	Byte Cnt	Lwr Addr	Data	LCRC	Time Delta
75126		x1	3852		10:01010	000:00:0	1	002:00:0	SC	0	4	0x00	1 dword	0x99A353D5	472.000 ns

Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Idle
75127		x1			3852	0xC912	224.000 ns

Packet	R→	2.5	TLP	Mem	MW(32)	RequesterID	Tag	Address	1st BE	Last BE	Data	LCRC	Time Delta
75128		x1	577		10:00000	000:00:0	0	FEAFF180	1111	0000	1 dword	0x1AAAC0DA	288.000 ns

Endpoint concludes that completion is for TLP sent and NAK's that T

Packet	R→	2.5	DLLP	ACK	AckNak_Seq_Num	CRC 16	Time Delta
75129		x1			577	0x874E	2.337 ms

RC hangs for 345 ms

Packet	R→	2.5	DLLP	NAK	AckNak_Seq_Num	CRC 16	Time Delta
75130		x1			3852	0x2275	2.764 ms

Packet	R→	2.5	DLLP	NAK	AckNak_Seq_Num	CRC 16	Time Delta
75131		x1			577	0x8C29	345.602 ms

Packet	R→	2.5	TLP	Mem	MRd(32)	RequesterID	Tag	Address	1st BE	Last BE	LCRC	Time Delta
75132		x1	578		00:00000	000:00:0	0	FEAFF180	1111	0000	0xE5B3030B	256.000 ns

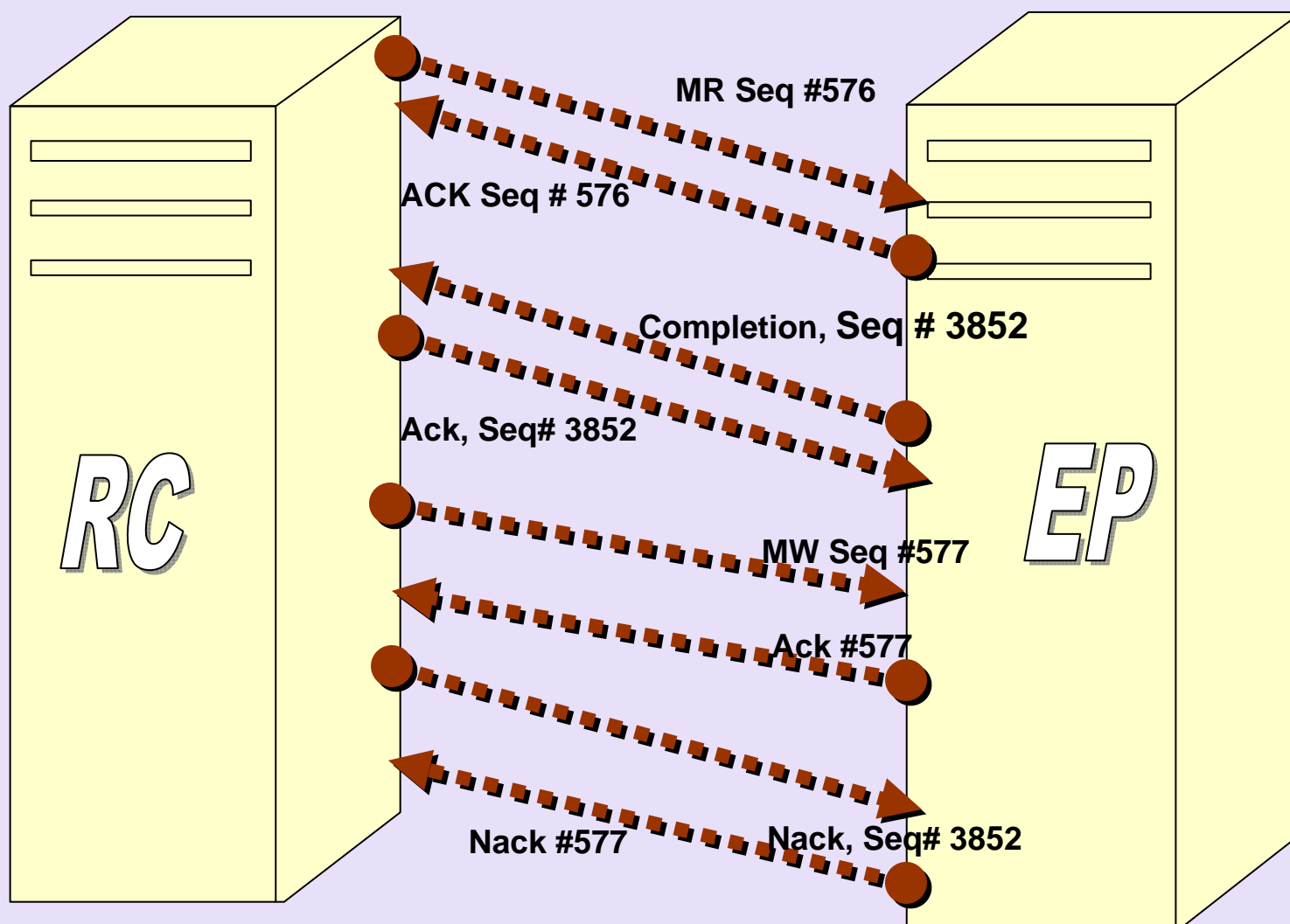
■ Undelivered TLP # 3853

■ Nak'd by NAK 3852

Endpoint concludes that NAK for completion is for TLP it never sent and NAK's that TLP

RC hangs for 345 msec

Incompliant NAK deadlock



Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe link level
 - ✓ Example: Credit leakage



ASPM L1 and Clock power management



- L1 exit latency does not account for CLKREQ enabled
- PLL shuts down in preparation for loss of REFCLK#
- System timeout due to long L1 recovery
- Use backup slow clock to replace bit clock when CLKREQ# de-asserted

ASPM L1 and Clock power management

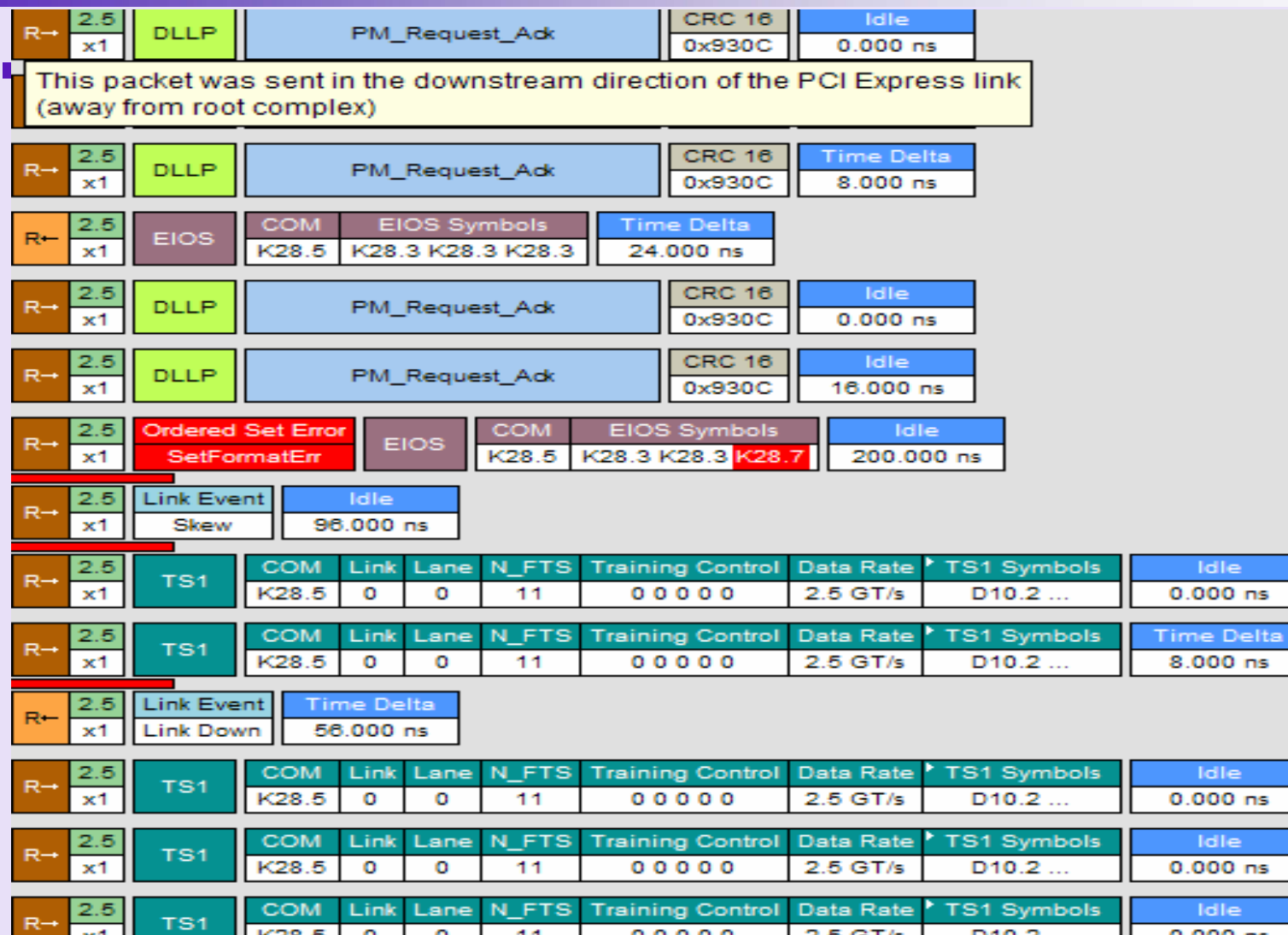
R→	2.5 x1	TS2	COM K28.5	Link 0	Lane 0	N_FTS 11	Training Control 0 0 0 0	Data Rate 2.5 GT/s	TS2 Symbols D05.2 ...	Time Delta 32.000 ns
Packet number (Post-Trigger)										
R→	2.5 x1	TS2	COM K28.5	Link 0	Lane 0	N_FTS 15	Training Control 0 0 0 0	Data Rate 2.5 GT/s	TS2 Symbols D05.2 ...	Time Delta 32.000 ns
R→	2.5 x1	TS2	COM K28.5	Link 0	Lane 0	N_FTS 11	Training Control 0 0 0 0	Data Rate 2.5 GT/s	TS2 Symbols D05.2 ...	Idle 0.000 ns
R→	2.5 x1	TS2	COM K28.5	Link 0	Lane 0	N_FTS 11	Training Control 0 0 0 0	Data Rate 2.5 GT/s	TS2 Symbols D05.2 ...	Idle 0.000 ns
R→	2.5 x1	TS2	COM K28.5	Link 0	Lane 0	N_FTS 11	Training Control 0 0 0 0	Data Rate 2.5 GT/s	TS2 Symbols D05.2 ...	Time Delta 376.000 ns

R→	2.5 x1	TLP 3208	Cpl	CpID 10:01010	RequesterID 000:00:0	Tag 14	CompleterID 001:00:0	Status SC	BCM 0	Byte Cnt 4	Lwr Addr 0x00	Register Data 0x00000000
----	-----------	-------------	-----	------------------	-------------------------	-----------	-------------------------	--------------	----------	---------------	------------------	-----------------------------

R→	2.5 x1	DLLP	ACK	AckNak_Seq_Num 3208	CRC 16 0x840C	Time Delta 2.656 μs
----	-----------	------	-----	------------------------	------------------	------------------------

R→	2.5 x1	DLLP	PM_Enter_L1	CRC 16 0x65AD	Idle 0.000 ns
R→	2.5 x1	DLLP	PM_Enter_L1	CRC 16 0x65AD	Idle 0.000 ns
R→	2.5 x1	DLLP	PM_Enter_L1	CRC 16 0x65AD	Idle 0.000 ns
R→	2.5 x1	DLLP	PM_Enter_L1	CRC 16 0x65AD	Idle 0.000 ns
R→	2.5 x1	DLLP	PM_Enter_L1	CRC 16 0x65AD	Time Delta 24.000 ns
R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 8.000 ns
R→	2.5 x1	DLLP	PM_Enter_L1	CRC 16 0x65AD	Time Delta 24.000 ns
R→	2.5 x1	DLLP	PM_Request_Ack	CRC 16 0x930C	Time Delta 8.000 ns

ASPM L1 and Clock power management



ASPM L1 and Clock power management

R→	2.5 x1	TS1	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS1 Symbols	Idle	
			K28.5	PAD	PAD	11	0 0 0 0 0	2.5 GT/s	D10.2 ...	0.000 ns	
R→	2.5 x1	TS1	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS1 Symbols	Idle	
			K28.5	PAD	PAD	11	0 0 0 0 0	2.5 GT/s	D10.2 ...	0.000 ns	
R→	2.5 x1	TS1	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS1 Symbols	Idle	
			K28.5	PAD	PAD	11	0 0 0 0 0	2.5 GT/s	D10.2 ...	0.000 ns	
R→	2.5 x1	TS1	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS1 Symbols	Idle	
			K28.5	PAD	PAD	11	0 0 0 0 0	2.5 GT/s	D10.2 ...	0.000 ns	
R→	2.5 x1	TS1	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS1 Symbols	Idle	
			K28.5	PAD	PAD	11	0 0 0 0 0	2.5 GT/s	D10.2 ...	0.000 ns	
R→	2.5 x1	TS1	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS1 Symbols	Idle	
			K28.5	PAD	PAD	11	0 0 0 0 0	2.5 GT/s	D10.2 ...	0.000 ns	
R→	2.5 x1	Ordered Set Error AlignErr, LenErr	TS1	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS1 Symbols	Idle
			K28.5	PAD	PAD	11	0 0 0 0 0	2.5 GT/s	D10.2 ...	0.000 ns	
R→	2.5 x1	PATN	PATN Symbols						Idle		
			K28.5	K28.5	K28.5	D21.5	K28.5	D10.2	0.000 ns		
R→	2.5 x1	PATN	PATN Symbols						Idle		
			K28.5	K28.5	K28.5	D21.5	K28.5	D10.2	0.000 ns		
R→	2.5 x1	PATN	PATN Symbols						Idle		
			K28.5	D21.5	K28.5	D10.2			0.000 ns		
R→	2.5 x1	PATN	PATN Symbols						Idle		
			K28.5	D21.5	K28.5	D10.2			0.000 ns		
R→	2.5 x1	PATN	PATN Symbols						Idle		
			K28.5	D21.5	K28.5	D10.2			0.000 ns		
R→	2.5 x1	PATN	PATN Symbols						Idle		
			K28.5	D21.5	K28.5	D10.2			0.000 ns		
R→	2.5 x1	PATN	PATN Symbols						Idle		
			K28.5	D21.5	K28.5	D10.2			0.000 ns		

Agenda

- Deadlock issues in bus architectures
- Deadlocks in PCI Express architectures
- Deadlocks at common resource level
 - ✓ Example: ASPM L1 re-entry delay
- Deadlock at RC/chipset level
 - ✓ Example: PME fence mechanism hang-up
- Deadlock at device level – EP or RC
 - ✓ Example 1: Incompliant NAK
 - ✓ Example 2: CLKREQ# readiness
- Deadlock at PCIe link level
 - ✓ Example: Credit leakage

Deadlock associated with particular PCIe link

- Credit starvation deadlocks
 - ✓ Malformed TLP going upstream
 - ✓ Downstream device accounts for credit
 - ✓ Upstream device does not account for credit
 - ✓ Additional malformed TLPs lead to credit starvation

Malformed TLP deadlock

R→	2.5 x1	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Time Delta
			K28.5	1	0	31	0 0 0 0	2.5 GT/s	D05.2 ...	16.000 ns
R→	2.5 x1	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Time Delta
			K28.5	1	0	15	0 0 0 0	2.5 GT/s	D05.2 ...	48.000 ns
R→	2.5 x1	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Time Delta
			K28.5	1	0	31	0 0 0 0	2.5 GT/s	D05.2 ...	16.000 ns
R→	2.5 x1	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Time Delta
			K28.5	1	0	15	0 0 0 0	2.5 GT/s	D05.2 ...	48.000 ns
R→	2.5 x1	TS2	COM	Link	Lane	N_FTS	Training Control	Data Rate	TS2 Symbols	Idle
			K28.5	1	0	31	0 0 0 0	2.5 GT/s	D05.2 ...	96.000 ns
R→	2.5 x1	SKIP	COM	SKIP Symbols			Idle			
			K28.5	K28.0	K28.0	K28.0	0.000 ns			
R→	2.5 x1	DLLP	InitFC1-P	VC ID	HdrFC	DataFC	CRC 16	Idle		
				0	4	16	0xFBB9	0.000 ns		
R→	2.5 x1	DLLP	InitFC1-NP	VC ID	HdrFC	DataFC	CRC 16	Idle		
				0	8	4	0x9809	0.000 ns		
R→	2.5 x1	DLLP	InitFC1-Cpl	VC ID	HdrFC	DataFC	CRC 16	Idle		
				0	0	0	0xD892	0.000 ns		
R→	2.5 x1	DLLP	InitFC1-P	VC ID	HdrFC	DataFC	CRC 16	Time Delta		
				0	4	16	0xFBB9	32.000 ns		
R→	2.5 x1	DLLP	InitFC1-P	VC ID	HdrFC	DataFC	CRC 16	Time Delta		
				0	5	18	0x55E0	0.000 ns		
R→	2.5 x1	DLLP	InitFC1-NP	VC ID	HdrFC	DataFC	CRC 16	Time Delta		
				0	8	4	0x9809	32.000 ns		
R→	2.5 x1	DLLP	InitFC1-NP	VC ID	HdrFC	DataFC	CRC 16	Time Delta		
				0	5	18	0xBE87	0.000 ns		
R→	2.5 x1	DLLP	InitFC1-Cpl	VC ID	HdrFC	DataFC	CRC 16	Time Delta		
				0	0	0	0xD892	32.000 ns		
R→	2.5 x1	DLLP	InitFC1-Cpl	VC ID	HdrFC	DataFC	CRC 16	Time Delta		
				0	0	0	0xD892	0.000 ns		

- Downstream FC Init for posted transactions is 16, or $16 \times 16 = 256$ bytes

Malformed TLP deadlock

TLP Length 16

- RC advertises buffer space for a request for $4 \times 16 = 64$ posted data bytes
- EP sends memory write request for $16 \times 4 = 64$ bytes
- RC consumes 64 bytes and advertises release of buffer space
- EP sends additional memory write request
- RC receives a malformed TLP
- RC does not consume, does not update
- Result is credit leakage

Malformed TLP deadlock

TLP Length 16

- Memory Write TLPs require free resource of $16 \times 4 = 64$ bytes = 4 credits
- TLP 2705 is sent, but only TLP 2704 is ack'ed
- 4 credits are consumed and FC updated
- EP lags 4 credits behind RC
- Each malformed TLP adds additional lag of 4 credits

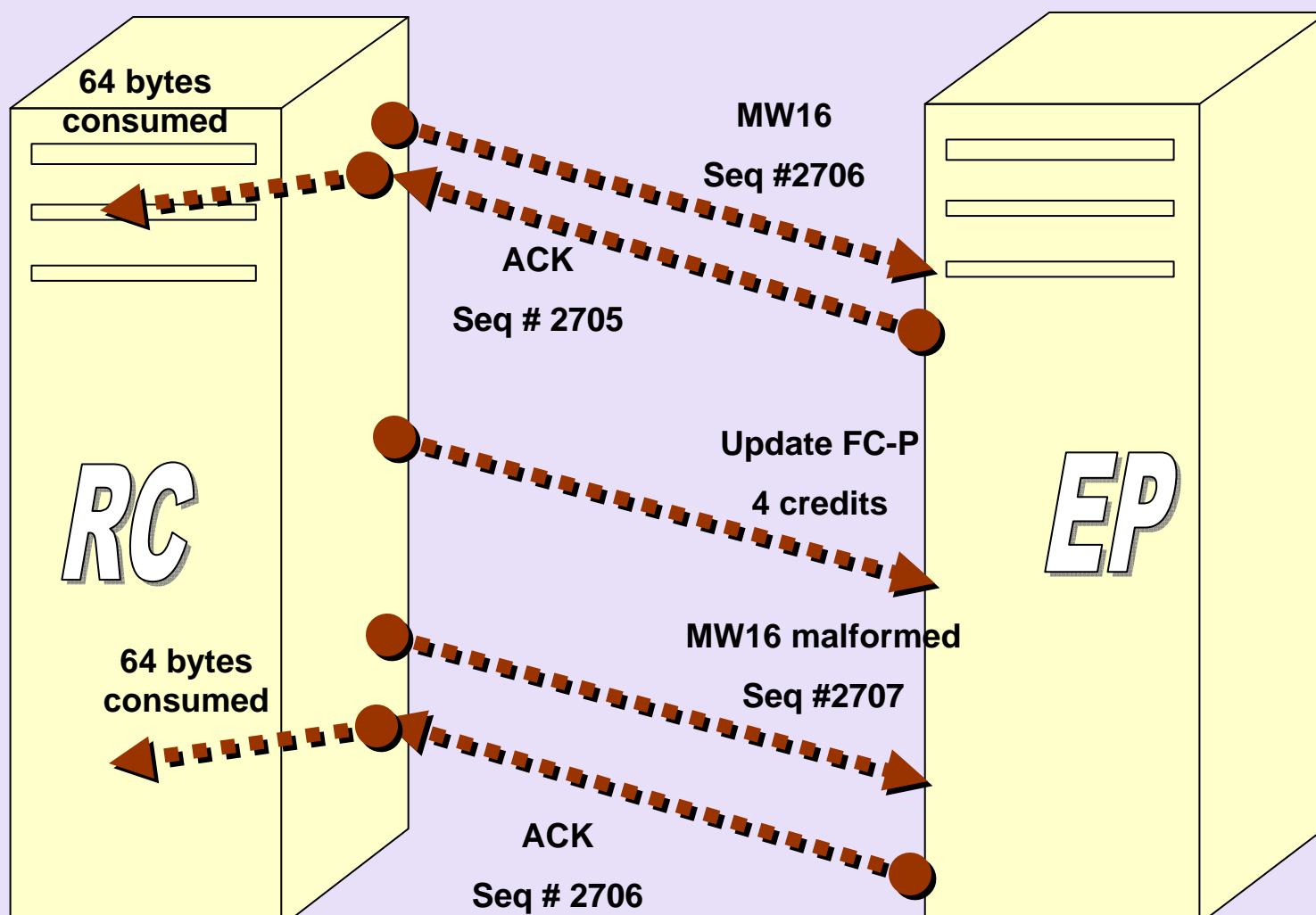
Malformed TLP deadlock

TLP length 16

R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Time Delta				
			0	255	1276	232.000 ns				
R←	TLP	Mem	MWrr(32)	Length	RequesterID	Tag	Address	Data	Time Delta	
	2705		10:00000	16	002:00:0	0	15FB2AC0	16 dwords	8.000 ns	
R→	DLLP	ACK	AckNak_Seq_Num	Idle						
			2704	0.000 ns						
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Idle				
			0	0	1276	0.000 ns				
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Idle				
			0	0	1278	0.000 ns				
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Time Delta				
			0	0	1280	232.000 ns				
R←	TLP	Mem	MWrr(32)	Length	RequesterID	Tag	Address	Data	Time Delta	
	2706		10:00000	16	002:00:0	0	15FB2B00	16 dwords	60.000 ns	
R→	DLLP	ACK	AckNak_Seq_Num	Idle						
			2705	0.000 ns						
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Idle				
			0	1	1280	0.000 ns				
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Idle				
			0	1	1282	0.000 ns				
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Time Delta				
			0	1	1284	180.000 ns				
R←	Packet Error TLPPayloadErr	TLP	Mem	MWrr(32)	Length	RequesterID	Tag	Address	Data	Time Delta
		2707		10:00000	16	002:00:0	0	15FB2B40	15 dwords	68.000 ns
R→	DLLP	ACK	AckNak_Seq_Num	Idle						
			2706	0.000 ns						
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Idle				
			0	2	1284	0.000 ns				
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Idle				
			0	2	1285	0.000 ns				
R→	DLLP	UpdateFC-P	VC ID	HdrFC	DataFC	Time Delta				
			0	2	1288	188.000 ns				
R←	Packet Error TLPPayloadErr	TLP	Mem	MWrr(32)	Length	RequesterID	Tag	Address	Data	Time Delta
		2708		10:00000	16	002:00:0	0	15FB2800	12 dwords	36.000 ns
R→	DLLP	ACK	AckNak_Seq_Num	Idle						
			2707	0.000 ns						

Malformed TLP deadlock

TLP length 16



Malformed TLP deadlock

TLP Length 16

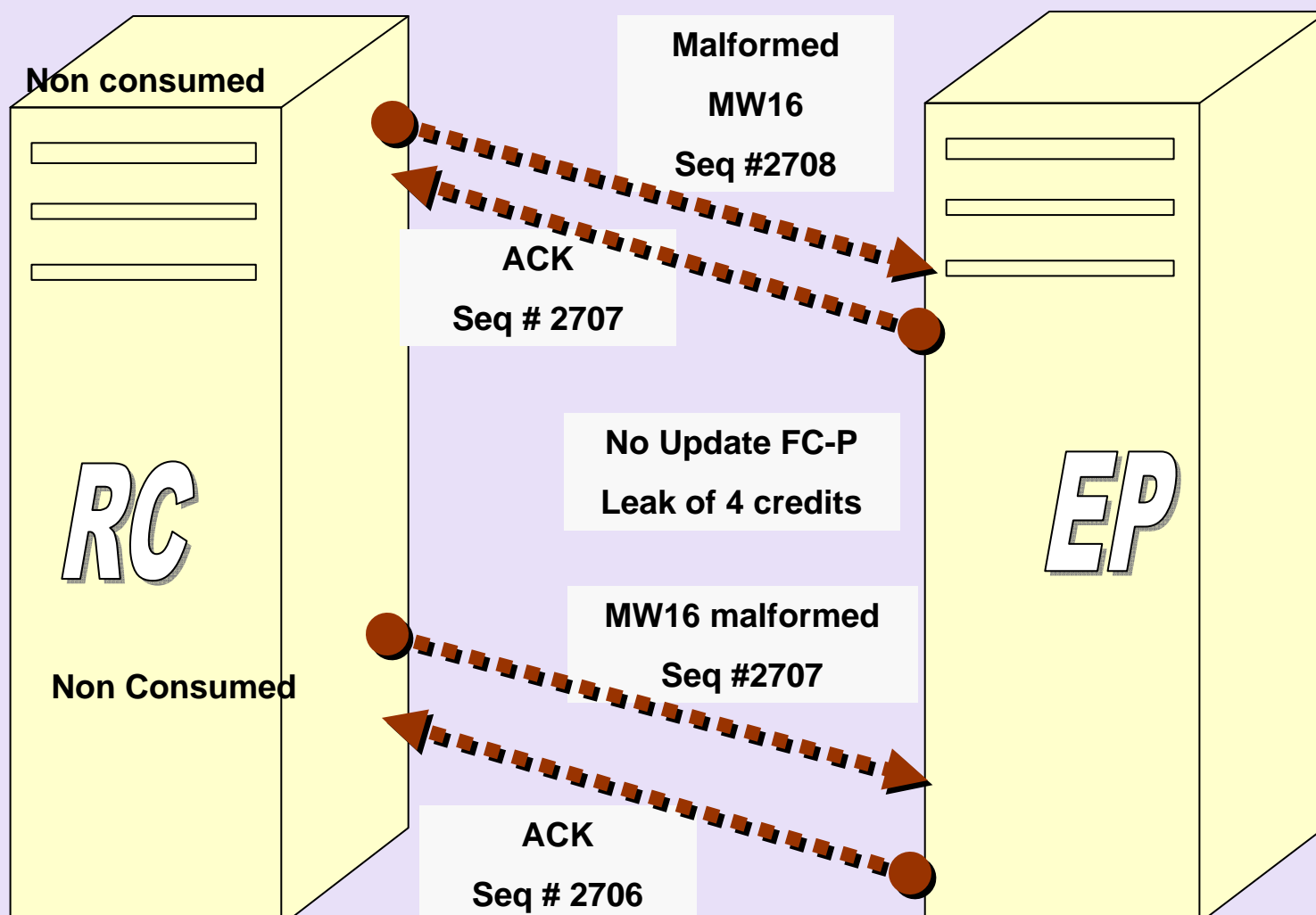
- 3 malformed TLPs create a lag of 12 credits
- Leads to EP credit starvation and blocks further traffic
- Deadlock scenario resulting from credit starvation

Malformed TLP deadlock TLP length 16

R←	Packet Error TLPPayloadErr	TLP 2708	Mem	MW(32)	Length 16	RequesterID 002:00:0	Tag 0	Address 15FB2800	Data 12 dwords	Time Delta 36.000 ns
R→	DLLP	ACK	AckNak_Seq_Num 2707	Idle						
R→	DLLP	UpdateFC-NP	VC ID 0	HdrFC 242	DataFC 4					Time Delta 236.000 ns
R←	Packet Error TLPPayloadErr	TLP 2709	Mem	MW(32)	Length 16	RequesterID 002:00:0	Tag 0	Address 15FB2800	Data 12 dwords	Time Delta 40.000 ns
R→	DLLP	ACK	AckNak_Seq_Num 2708	Idle						
R→	DLLP	ACK	AckNak_Seq_Num 2709	Time Delta						3.560 µs
R←	DLLP	UpdateFC-P	VC ID 0	HdrFC 44	DataFC 1849					Idle 0.000 ns
R←	DLLP	UpdateFC-NP	VC ID 0	HdrFC 176	DataFC 95					Time Delta 23.960 µs
R→	DLLP	UpdateFC-P	VC ID 0	HdrFC 2	DataFC 1288					Idle 964.000 ns
R→	DLLP	UpdateFC-NP	VC ID 0	HdrFC 242	DataFC 4					Time Delta 7.788 µs
R←	DLLP	UpdateFC-P	VC ID 0	HdrFC 44	DataFC 1849					Idle 0.000 ns
R←	DLLP	UpdateFC-NP	VC ID 0	HdrFC 176	DataFC 95					Time Delta 20.296 µs
R→	DLLP	UpdateFC-P	VC ID 0	HdrFC 2	DataFC 1288					Idle 972.000 ns
R→	DLLP	UpdateFC-NP	VC ID 0	HdrFC 242	DataFC 4					Time Delta 11.460 µs
R←	DLLP	UpdateFC-P	VC ID 0	HdrFC 44	DataFC 1849					Idle 0.000 ns
R←	DLLP	UpdateFC-NP	VC ID 0	HdrFC 176	DataFC 95					Time Delta 16.660 µs
R→	DLLP	UpdateFC-P	VC ID 0	HdrFC 2	DataFC 1288					Idle 976.000 ns
R→	DLLP	UpdateFC-NP	VC ID 0	HdrFC 242	DataFC 4					Time Delta 15.100 µs

Malformed TLP deadlock

TLP length 16



Malformed TLP deadlock

TLP length 32

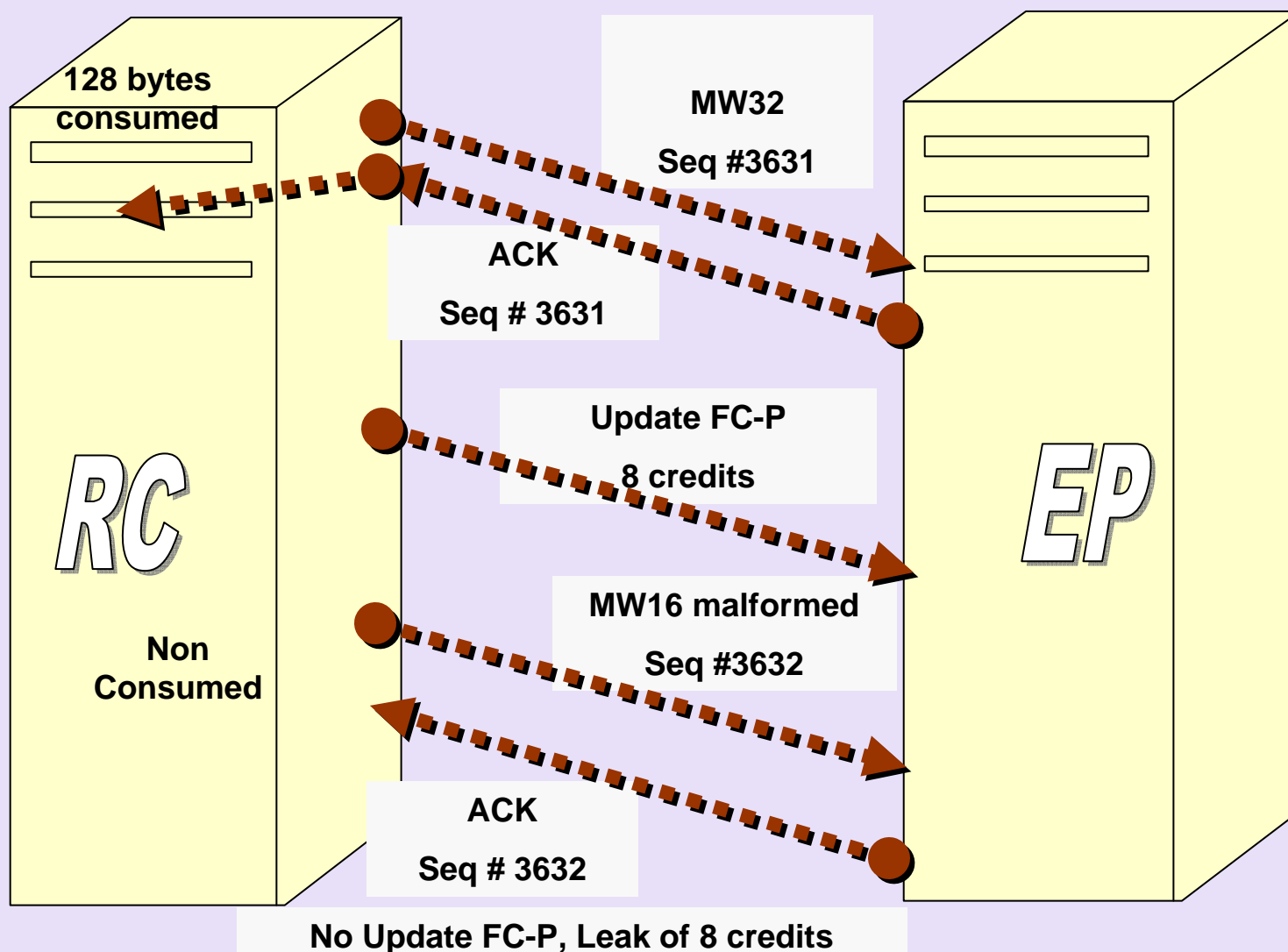
- TLP of length 32 requires $32 * 4 / 16 = 8$ credits
- No Updates
- Credit starvation will occur after only 2 malformed TLPs

Malformed TLP deadlock TLP length 32

R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Idle	Time Stamp				
				170	7	0.000 ns	0147 . 191 591 984 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Time Delta	Time Stamp				
				170	9	48.000 ns	0147 . 191 592 016 s				
R→	2.5 x1	DLLP	UpdateFC-P DLLP: Updating Flow Control Credits for Virtual Channel								
				170	11	224.000 ns	0147 . 191 592 064 s				
R←	2.5 x1	TLP	Mem	MWrr(32)	Attributes	Length	RequesterID	Tag	Address	Data	Time Delta
		3631		10:00000	00	32	002:00:0	0	285DC580	32 dwords	672.000 ns
R→	2.5 x1	DLLP	ACK	AckNak_Seq_Num		Idle	Time Stamp				
				3631		0.000 ns	0147 . 191 592 960 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Idle	Time Stamp				
				171	11	0.000 ns	0147 . 191 592 992 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Idle	Time Stamp				
				171	13	0.000 ns	0147 . 191 593 024 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Idle	Time Stamp				
				171	15	68.000 ns	0147 . 191 593 056 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Idle	Time Stamp				
				171	17	0.000 ns	0147 . 191 593 156 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Time Delta	Time Stamp				
				171	19	260.000 ns	0147 . 191 593 188 s				
R←	2.5 x1	Packet Error TLPPayloadErr	TLP	Mem	MWrr(32)	Attributes	Length	RequesterID	Tag	Address	Data
			3632		10:00000	00	32	002:00:0	0	285DC600	15 dwords
R→	2.5 x1	DLLP	ACK	AckNak_Seq_Num		Time Delta	Time Stamp				
				3632		9.912 µs	0147 . 191 593 820 s				
R→	2.5 x1	DLLP	UpdateFC-NP	HdrFC	DataFC	Time Delta	Time Stamp				
				116	4	18.088 µs	0147 . 191 603 732 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Time Delta	Time Stamp				
				171	19	11.052 µs	0147 . 191 621 820 s				
R→	2.5 x1	DLLP	UpdateFC-NP	HdrFC	DataFC	Time Delta	Time Stamp				
				116	4	18.084 µs	0147 . 191 632 872 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Time Delta	Time Stamp				
				171	19	11.024 µs	0147 . 191 650 956 s				
R→	2.5 x1	DLLP	UpdateFC-NP	HdrFC	DataFC	Time Delta	Time Stamp				
				116	4	18.068 µs	0147 . 191 661 980 s				
R→	2.5 x1	DLLP	UpdateFC-P	HdrFC	DataFC	Time Delta	Time Stamp				
				171	19	11.060 µs	0147 . 191 680 048 s				

Malformed TLP deadlock

TLP length 32



Malformed TLP deadlock

- Error reporting
 - ✓ Malformed TLP Error reporting message required for downstream TLP, meaningless for upstream TLP
- Specification enhancement recommendation
 - ✓ Downstream FC Init or link retrain should be mandatory following malformed TLP, or other scenarios leading to credit starvation

Thank you for attending the
PCI-SIG Developers Conference 2008

For more information please go to
www.pcisig.com