# Multicast功能介绍

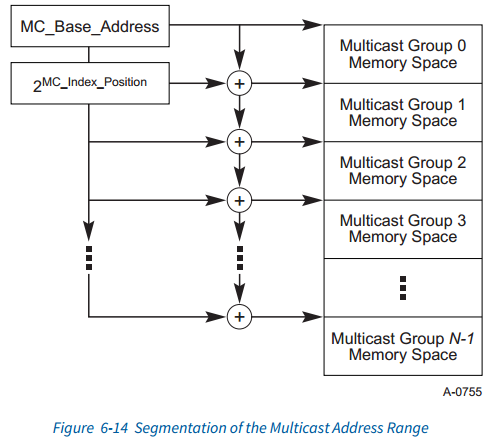
## 背景介绍

PCIe事务传输支持单播、组播和广播三种方式，其中以单播和广播应用最多。单播是点对点的传输，只有一个请求者和一个完成者，同一时刻主机只能与众多子设备中的一个设备进行通信；广播是点到面的传输，有一个请求者和多个完成者，请求者下游全局地址范围内的所有PCIe组件作为完成者，常见于消息事务的传播，比如RC广播消息，从RC发送消息给其所有下游组件；组播是介于单播和广播之间的传输方式，有一个请求者和多个完成者，请求者下游指定地址范围内的多个PCIe组件都可以作为完成者，同一时刻允许主机向多个设备或单个设备向多个主机发送数据，用于PCIe存储器写事务或地址路由的事务。

相较于单播，组播能够把数据发给一组设备，避免了发送重复的数据流，能够提升系统传输性能；相较于广播，组播具有更好的指向性和安全性。

## multicast业务

Multicast Capability结构定义了组播的地址范围，该地址范围被等分为大小相同的N个组播窗口，一个组播窗口被称为一个组播组（Multicast Group，MCG），如图所示。每个支持组播的Function都应实现该组播能力结构，以提供事务的路由方向、决定每个MCG的事务接收或透传。PCIe 5.0中MCG是6bit位宽的域，最多能够支持64个组播组。



### 组播TLP的处理

组播命中的触发需要满足所有以下三个条件

* MC\_Enable置1；
* TLP为memory write或地址路由，并且都是posted request；
* MC\_Base\_Address<=AddressTLP<=(MC\_Base\_Address+(2MC\_Index\_Position\*(MC\_Num\_Group + 1)))

在支持组播的Switch节点的所有function及多function上行端口中，通过软件把MC\_Enable、MC\_Base\_Address、MC\_Index\_Position及MC\_Num\_Group等相同的域配置为相同的值，这样交换节点入端口及其他组件可以使用其任意一个function的MC相关域的值。否则入端口就有可能取到不同的值而导致不确定的情况发生。

PCIe组播操作只支持P（Posted）请求，若NP（Non-Posted）请求的地址刚好在某个MCG地址范围内，这仍然不是一个组播操作，仍应按照单播的处理方式对齐进行处理。

一旦触发了组播操作，常规的地址路由规则不再适用，而是遵循以下规则：

从TLP地址中提取组播组号MCG，计算方式为MCG=((AddressTLP - MC\_Base\_Address) >> MC\_Index\_Position) & 3Fh，其中MC\_Base\_Address及MC\_Index\_Position可以采用任何一个function中该域的值。

接下来，组件根据MC\_Block\_All及MC\_Block\_Untranslated域比特向量的值来检查提取到的MCG：switch及RP采用入端口的MC\_Block\_All/MC\_Block\_Untranslated值检测入端口组播TLP，EP Function采用其function内的MC\_Block\_All/MC\_Block\_Untranslated值来检测待发的组播TLP。若MCG中存在MC\_Block\_All要阻塞的TLP，则对应TLP当作组播阻塞TLP处理；若MCG中存在MC\_Block\_Untranslated要阻塞的TLP，且TLP中的地址为未转换地址（根据头标AT域判断），则对应TLP同样当作组播阻塞TLP处理。

对于switch及RC，若TLP在该组件中没有被阻塞，且从TLP提取出的MCG在MC\_Receive寄存器中的对应位都有置1，那么该TLP会被前推到该组件除入端口外的所有端口；对于EP，该TLP会被所有function接收；若没有port前推或function接收该TLP，该TLP会被静默丢掉。

为了防止形成组播环路，即便RP（Root Port）或Switch Port的MC\_Receive设置符合转发条件，也绝对不能沿入端口推回该TLP。有一个例外，若RC收到了一笔单播请求且刚好落在了其组播窗口内，在这种情况下，当与该RP入口关联的MC\_Receive寄存器指定时，RC需要将落入组播窗口的TLP发送到它最初进入的同一Root端口。

组播命中暂停正常地址路由，包括switch中的默认上行路由。当发生组播命中时，TLP只会将设置了MC\_Receive位的出端口转发出去，MC\_Receive位与TLP中地址提取的MCG相关联。如果TLP中的地址没有使用正常地址解码解码到任何下游端口，则只有在上游端口的MC\_Receive寄存器指定的情况下，TLP才会被复制到上游端口。

若开启了ACS机制，ACS服务中唯一能用于组播操作的访问控制功能是ACS源地址验证。

组播事务的多个最终接收者组的播窗口大小是很有可能不同的。一个极端例子，某个组播窗口需要的窗口大小可能会覆盖整个设备的存储空间范围。反之，某个组播窗口大小可能仅需要覆盖FIFO寄存器内小段特定的偏移。为了应对这种情况，EP通过设置其组播能力寄存器内的组播窗口大小域MC\_Window\_Size\_Requested来表明该EP需要的窗口大小。

在可用地址空间允许的情况下，资源分配软件应尽可能满足所有组件对组播窗口大小的最低请求，并通过设置MC\_Index\_Position来设置组播窗口大小以满足最大请求。有些情况下，可将单个较大组播窗口的请求拆分为多个窗口较小的连续组播组，从而满足对较大窗口的需求。

### 组播排序

没有为处理组播tlp定义新的排序规则。所有组播TLP都是Posted请求，并遵循Posted请求排序规则。通过复制点，组播TLP按照相对于组件ingress stream中的其他TLP的正常排序规则进行排序。一旦复制到egress stream中，组播TLP遵循与stream中其他Posted请求相同的顺序。

### 组播Capability结构字段更新

组播能力结构中部分域支持随时更新，包括MC\_Enable, MC\_Num\_Group, MC\_Receive, MC\_Block\_All, MC\_Block\_Untranslated。在更新这些域的值的时候必须按顺序更新。采用相同TC的两笔TLP A和B先后从相同入端口到达，若A在某个组播域把值更新为X，则B相同域也应更新为X或者采用更新的值。

除可随时更新的域之外，其他域只有在组件内所有function的MC\_Enable都清零后才能更新，这样的组播域有MC\_Base\_Address, MC\_Index\_Position。

### 阻塞TLP的处理

若组播事务被MC\_Block\_All或MC\_Block\_Untranslated机制阻塞住了，该笔事务会被丢弃。阻塞住该笔事务的function就是该事务的完成者，完成者必须按照常规的错误处理方式记录并上报MC\_Blocked TLP错误。此外，完成者必须在其状态寄存器或第二状态寄存器合适的地方将Signaled Target Abord位置一。为了隔离错误并分析错误原因，强烈建议在具有组播能力的function内实现AER。

在RC或switch内，如果入端口接收TLP时发生错误，则由入端口上报错误。若EP function在准备发送TLP的过程中出现错误，则由EP function上报错误。

### MC\_Overlay机制

MC\_Overlay机制允许端点中不包含组播能力结构的单个BAR同时用于组播和单播TLP接收。软件可以通过在下游端口设置MC\_Overlay\_BAR来配置MC\_Overlay机制来影响这一点，以便多播地址范围或它的一部分被重新映射(覆盖)到端点的BAR接受的内存空间范围。在Switch的上游端口，该机制可用于将部分组播地址范围覆盖到与主机内存相关的内存空间范围上。

下游端口的MC\_Overlay机制适用于从该端口出去的TLP。上游端口的MC\_Overlay机制适用于退出该端口的TLP。端口的MC\_Overlay机制不适用于从该端口接收到的TLPs、以端口内内存空间为目标的TLPs、多功能上游端口内功能间对等路由的TLPs。

当开启了MC\_Overlay机制，overlay操作指定将组播TLP地址中位数大于或等于MC\_Overlay\_Size字段的位替换为MC\_Overlay\_BAR中相应的位。即：

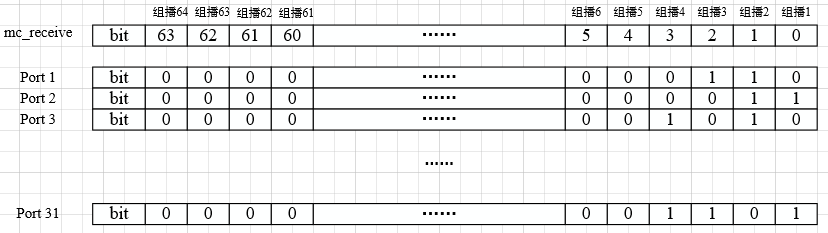
Egress\_TLP\_Addr=(MC\_Overlay\_Size<6)?Ingress\_TLP\_Addr :{ MC\_Overlay\_BAR[63:MC\_Overlay\_Size], Ingress\_TLP\_Addr[MC\_Overlay\_Size-1:0] }。

## multicast机制

在全部满足1.2章节中的三个条件时，则认为当前TLP是需要通过组播进行转发的，也就是组播命中。

组播组数是通过mc\_num\_group控制的，最多可以有64组组播；而且这个域段和mc\_index\_position、mc\_base\_address共同决定了组播地址的范围。

而下游port和组播组的映射关系是一个组播组会存在多个下游port，每个下游port配置的mc\_receive寄存器和上游端口接收到的TLP目标地址决定的mc\_group（MCG），即mc\_receive[mc\_group]共同确定该下游port是否属于该组播组。

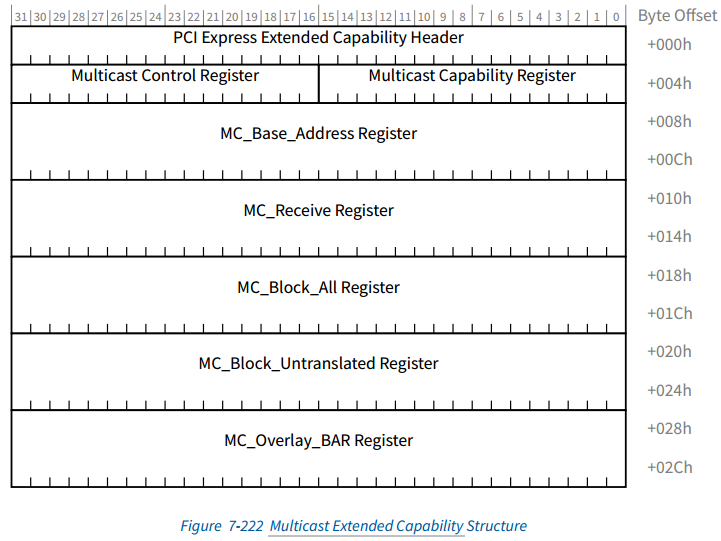


如上图所示，根据mc\_receive的每个bit对应一个组播组，从图中可以看到有4个组播。第一个组播组有port2和port31；第二个组播组有port1、port2和port3；第三个组播组有port1和port31；第四个组播组有port3和port31。

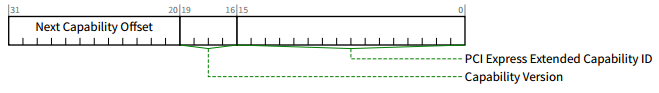
# Multicast Extended Capability

Multicast是一种可选的规范功能，由Multicast Extended Capability控制。Multicast Extended Capability适用于RC ports、RCRBs、Switch ports、EP functions和RCiEP。它不适用于PCI Express到PCI/PCI-x桥。

在switch或RC或包含multi-function component的情况下，存在需要此Capability结构的多个副本。每个支持多播端的EP、switch port或RC port都需要一个该副本。为了提高实现效率，组件中的每个Multicast Extended Capability中的某些字段必须被配置为相同的；如果这些字段没有配置为相同值，就会出现问题。必须配置相同值的字段和寄存器包括MC\_Enable、MC\_Num\_Group、MC\_Base\_Address和MC\_Index\_Position。EP的Multicast Extended Capability结构中的这些相同字段必须与EP上的Switch或RC或集成了RCiEP的Multicast Extended Capability结构中的配置相匹配。



## Multicast Extended Capability Header(Offset 00h)



### PCI Express Extended Capability ID（已实现）

This field is a PCI-SIG defined ID number that indicates the nature and format of the Extended Capability. PCI Express Extended Capability ID for the Multicast Extended Capability is 0012h.

该字段是PCI-SIG定义的标识号，表示扩展能力的性质和格式。PCI Express Extended Capability的组播扩展能力ID为0012h。

225已实现，具体在pcie5\_extcap模块中，895行中固定赋值为16’h0012，并传递给信号mcast00中，但是后续向其他模块传输过程中并没有再使用。

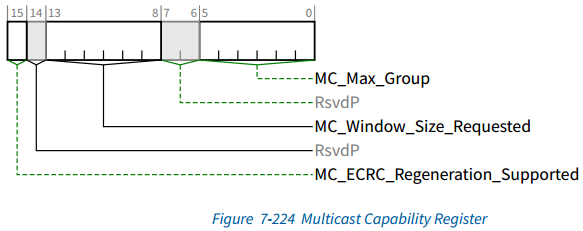
### Capability Version（已实现）

This field is a PCI-SIG defined version number that indicates the version of the Capability structure present.

该字段是PCI-SIG定义的版本号，指示当前Capability结构的版本。

225已实现，具体在pcie5\_extcap模块中，895行中固定赋值为4’h1，指示当前的版本为第1版本。该域段信号传输给信号mcast00之后没有再使用过。

## Multicast Capability Register(Offset 04h)



### MC\_Max\_Group（已实现）

Value indicates the maximum number of Multicast Groups that the component supports, encoded as M-1. A value of 00h indicates that one Multicast Group is supported.

该值表示组件支持的最大组播组数，编码为M-1。00h表示只支持一个组播组。

225已实现，具体在pcie5\_extcap模块220行k\_mc\_maxgroup信号，该信号表示最大的组播数，并直接通过配置信号k\_pciconf[`KPCICONF\_MCAST\_MAX\_GROUP+: 6]进行配置，当该信号的值为00h时，表示最大允许有一个组播组。

该信号仅在pcie5\_extcap模块用作判断mc\_receive、mc\_blockall、mc\_blockunt等寄存器的每bit是否可以被cfg\_data赋值的条件。该信号为6bit信号，即数值范围为0~64，而mc\_receive等寄存器在每bit赋值过程中都会判断该bit位是否小于mc\_max\_group，如果小于mc\_max\_group，则表示满足条件，没有超过最大的组播组数，并将mc\_receive的对应bit赋值；而如果大于mc\_max\_group，则不会给mc\_receive的该bit位赋值。

该域段信号k\_mc\_maxgroup在pcie5\_extcap模块中赋值给mcast04信号，mcast04的该域段在外部信号传递过程中没有被其他模块使用。

### MC\_Window\_Size\_Requested（已实现未使用）

In Endpoints, the log2 of the Multicast Window size requested. RsvdP in Switch and Root Ports.

在EP中，请求的多播窗口大小的log2。对于Switch和RC为预留值。

225已实现，具体在模块pcie5\_extcap模块898行，通过将配置信号k\_pciconf[`KPCICONF\_MCAST\_WINDOW\_SIZE+: 6]赋值，直接传递给mcast04信号，但是该部分域段在本模块和其他模块中都没有使用到。并且该域段只有EP port才会使用，表示在EP中请求的组播组数，而该域段对于switch和RC都是预留值。

### MC\_ECRC\_Regeneration\_Supported（已实现）

If Set, indicates that ECRC regeneration is supported.

This bit must not be Set unless the Function supports Advanced Error Reporting, and the ECRC Check Capable bit in the Advanced Error Capabilities and Control register is also Set. However, if ECRC regeneration is supported, its operation is not contingent upon the setting of the ECRC Check Enable bit in the Advanced Error Capabilities and Control register. This bit is applicable to Switch and Root Ports and is RsvdP in all other Functions.

如果该域段置1，表示支持ECRC重构。

除非功能支持高级错误报告，并且Advanced Error Capabilities和Control寄存器中的ECRC Check Capable位也被置1，否则不能设置此位。但是如果支持ECRC重构，则其操作不取决于Advanced Error Capabilities和Control寄存器中ECRC Check Capable位的设置。该位适用于Switch和RC端口，在所有其他功能中为预留值。

225已实现，在模块pcie5\_extcap模块中，该部分域段直接通过配置信号k\_pciconf[`KPCICONF\_MCAST\_ECRC\_REGEN\_SUP]进行配置，拼接赋值给信号mcast04，但是该部分域段没有再传输给其他信号。而是在hsio\_top\_pipe模块中例化hsio\_decoder\_up和hsio\_decoder\_dn模块时直接将配置信号k\_pciconf赋给ecrc\_regen\_en信号，表示是否支持ecrc重构。

TLP中是否有ECRC取决于TD字段，如果原TD为1，表示原TLP中包含有ECRC部分，如果原TD为0，则表示原TLP中没有ECRC。在decoder模块中ecrc\_regen\_en会传递给multicast模块，并与接收到的TLP的TD字段做逻辑与判断后，其结果作为新的TD替换掉原来TLP中的TD。

下图为mc\_overlay与ECRC重构规则的对应关系。225的代码实现便是根据下图的对应关系，开启了MC\_Overlay机制后才会涉及到ECRC的剥离操作，此时TLP是否有ECRC以及是否支持ECRC重构将决定TLP的TD字段以及ECRC的重构计算。

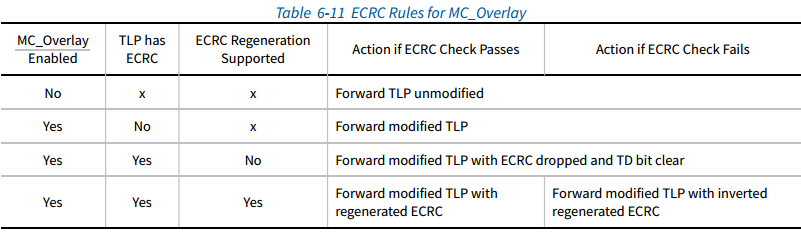


表1 225代码实现情况

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| MC\_Overlay  Enabled | TLP has  ERCR | ECRC regen  supported | TD modified | ECRC  removed | Address  modified |
| No | x | x | No | No | No |
| Yes | No | x | No | No | Yes |
| Yes | Yes | No | Yes | Yes | Yes |
| Yes | Yes | Yes | No | Yes | Yes |

根据上表，共有四种情况：

1. 当mc\_overaly机制没有打开时，不管TLP中是否有ECRC以及ECRC是否支持重构，TLP不会有任何操作和改动，直接将其转发出去。
2. 当mc\_overlay机制打开时，如果TLP中没有ECRC，那么不管是否支持ECRC重构，TD都不会发生变化而且也不存在ECRC被剔除，而此时地址会发生改变，将原TLP地址替换为mc\_overlay\_bar的地址。
3. 当mc\_overlay机制打开时，如果TLP中有ECRC，但是不支持ECRC重构，那么TD就会发生变化（TD由1变为0），并且在multicast模块中会将ECRC剔除掉（PCIe5.0协议中的第199页的ECRC rule提到有些receiver接收到的TLP中没有ECRC，这些是正常处理，不是一个错误）。此时地址会发生改变。
4. 当mc\_overlay机制打开时，如果TLP中有ECRC并且支持ECRC重构，此时TD保持为1不变，但是ECRC同样将会在multicast模块中被剔除掉，并在txtl模块中根据最新TD字段使pcie5\_ecrc模块重新计算ECRC，并在txtl模块中将重新计算好的ECRC插入到TLP中。

在txtl模块中，addecrc\_n信号控制是否附加ECRC，而该信号addecrc\_n拉高是由TLP的TD字段和tl\_tx\_err[3]决定的，而在decoder模块中tl\_tx\_err[3]是由multicast发送的TLP的TD字段决定的，因此可以看到addecrc\_n该信号完全是由TD字段决定的。

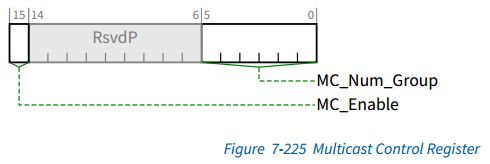
另外PCIe5.0中有以下描述：

如果TLP中的地址会被修改，此时便存在一个问题，当TLP开启了ECRC，TLP地址被修改了但ECRC没改，那么ECRC校验会出错。当然，switch和RC中的端口可以选择是否支持ECRC重构。若支持重构，强烈建立尽全力缩减原TLP的ECRC校验到地址修改后TLP ECRC重构之间这段时间。因为中间这段时间TLP处于不受保护的状态，存在数据完整性的漏洞。

对于开启了MC\_Overlay机制但不支持ECRC重构的端口，可以从穿过端口的组播TLP中剥离ECRC，这时接收端EP可以开启ECRC校验。这种情况下，EP将用于检测非组播TLP的ECRC，而不用检测MC\_Overlay机制修改过后的组播TLP的ECRC。

如果支持组播ECRC重构，并且再TLP修改之前检测到ECRC错误，这种情况下可以反转重构的ECRC来确保不会因为重构的ECRC而导致此次ECRC错误漏报。

## Multicast Control Register(Offset 06h)



### MC\_Num\_Group（已实现）

Value indicates the number of Multicast Groups configured for use, encoded as N-1. The default value of 00 0000b indicates that one Multicast Group is configured for use. Behavior is undefined if value exceeds MC\_Max\_Group. This parameter indirectly defines the upper limit of the Multicast address range. This field is ignored if MC\_Enable is Clear. Default value is 00 0000b.

该域段值表明配置的组播组个数，编码为N-1。默认值00 0000b表示配置了一个组播组。如果该值超过MC\_Max\_Group，则为未定义行为。该参数间接定义组播地址范围的上限值。如果MC\_Enable清零，则忽略该字段。并且默认值为00 0000b。

225已实现，在pcie5\_extcap模块中的888行，在判断cfg\_addr=12’h184时，该域段通过cfg\_data[16+: 6]进行赋值配置，cfg\_data信号为decoder\_up模块中的APB配置接收到的数据信息。并且该域段在pcie5\_extcap模块中和MC\_Index\_Position会根据协议的公式MC\_Base\_Address <= AddressTLP <= (MC\_Base\_Address +(2MC\_Index\_Position\*(MC\_Num\_Group + 1)))用于计算组播地址范围。

该域段信号在pcie5\_extcap模块中赋值给mcast04信号，mcast04的该域段在外部信号传递过程中没有被其他模块使用。

### MC\_Enable（已实现）

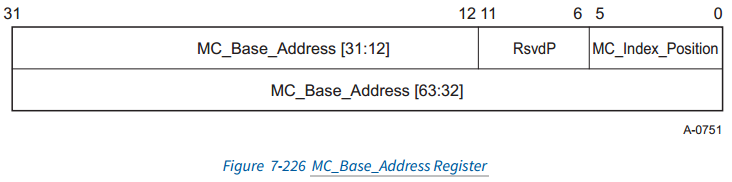
When Set, the Multicast mechanism is enabled for the component. Default value is 0b.

当该域段置1时，表示组件启用组播机制。默认值为0b。

225已实现，在pcie5\_extcap模块中的891行，在判断cfg\_addr=12’h184时，该域段信号通过cfg\_data[31]进行赋值配置。该信号会在pcie5\_extcap模块中用作判断组播是否命中的条件：也就是组播命中需要保证mc\_enable置1。

该域段信号在pcie5\_extcap模块中赋值给mcast04信号后，mcast04的该域段在外部信号传递过程中没有被其他模块使用。

## MC\_Base\_Address Register(Offset 08h)



### MC\_Index\_Position（已实现）

The location of the LSB of the Multicast Group number within the address. Behavior is undefined if this value is less than 12 and MC\_Enable is Set. Default is 0.

组播组号所在的LSB在地址内的位置。如果该值小于12且MC\_Enable置1，则行为未定义。该域段默认为0。

225已实现，在pcie5\_extcap模块的920行，通过判断是否cfg\_addr=12’h188，将cfg\_data[5:0]赋值给mc\_index\_pos\_r。该信号mc\_index\_pos\_r在pcie5\_extcap模块中用于计算组播的组数MCG（multicast group）和最大的组播地址mc\_addmax，最大的组播地址和mc\_base\_address将用于判断组播是否命中的条件。

### MC\_Base\_Address（已实现）

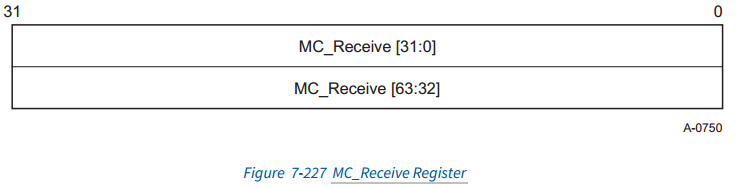
The base address of the Multicast address range. The behavior is undefined if MC\_Enable is Set and bits in this field corresponding to address bits that contain the Multicast Group number or address bits less than MC\_Index\_Position are non-zero. Default is 0.

组播地址范围的基地址。如果设置了MC\_Enable，并且该字段中包含组播组号的地址位或小于MC\_Index\_Position的地址位对应的位不为零，则该行为未定义。默认值为0。

225已实现，具体在pcie5\_extcap模块的925行，通过判断cfg\_addr=12’h188或者cfg\_addr=12’h18c，将cfg\_data的高20bit或cfg\_data全部32bit赋值给mc\_addr\_r。该信号mc\_addr\_r为组播范围的基地址，用作判断是否组播命中的条件。

信号mc\_addr\_r在pcie5\_extcap模块将mc\_addr\_r[31:12]和mc\_addr\_r[63:32]分别赋值给mcast08[31:12]和mcast10两个信号。这个两个信号的mc\_base\_address域段没有在外部其他模块中进行使用。

## MC\_Receive Register(Offset 10h)



### MC\_Receive（已实现）

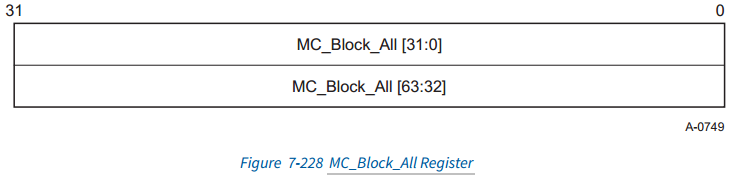
For each bit that’s Set, this Function gets a copy of any Multicast TLPs for the associated Multicast Group. Bits above MC\_Num\_Group are ignored by hardware. Default value of each bit is 0b.

对于设置的每一个bit，该功能为关联的组播组获取任何组播tlp的副本。而高于MC\_Num\_Group的位将被硬件忽略。每个bit的默认值为0b。

225已实现，具体在pcie5\_extcap模块的961行，通过判断是否cfg\_addr=12’h190以及cfg\_addr=12’h194，将全部的64bit数据cfg\_data赋值给mc\_receiver\_r，这部分信号分别将高32bit和低32bit的值分给mcast14和mcast10两个信号。

这两个信号拼接之后作为mcast\_cap信号的一部分输出到hsio\_top\_pipe模块，在hsio\_top\_pipe模块中会将32个端口配置的mc\_receive拼接为mc\_receive[64\*32-1:0]，然后将该mc\_receive[64\*32-1:0]传递给decoder模块，decoder\_up模块内部例化了31个multicast模块，然后在decoder\_up内部将mc\_receive[64\*32-1:0]按照之前的拼接顺序分别传递给各个与port对应的multicast模块中，multicast模块中通过mc\_group（MCG）选择判断mc\_receive[mc\_group]是否拉高。如果mc\_receive[mc\_group]拉高，则可以确定该port的TLP可以作为组播的TLP发送出去。

## MC\_Block\_All Register(Offset 18h)



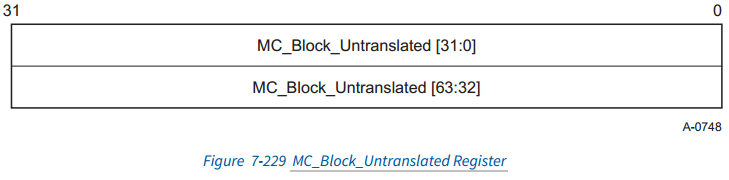
### MC\_Block\_All（已实现）

For each bit that is Set, this Function is blocked from sending TLPs to the associated Multicast Group. Bits above MC\_Num\_Group are ignored by hardware. Default value of each bit is 0b.

每设置一个bit，就会阻止该功能向关联的组播组发送tlp。高于MC\_Num\_Group的位将被硬件忽略。每个位的默认值为0b。

225已实现，具体在pcie5\_extcap模块中974行，通过判断是否cfg\_addr=12’h198以及cfg\_addr=12’h19C，将全部的64bit数据cfg\_data赋值给mc\_blockall\_r，这部分信号分别将高32bit和低32bit的值分给mcast1c和mcast18两个信号。1145行中该域段结合MCG用于判断mc\_block是否置1的条件，即当mc\_blockall\_r[mc\_group]=1时，mc\_block拉高。并且mc\_block标志信号会和mc\_hit、mc\_group信号拼接为mc\_status\_out后传递出去，最后提供给decoder使用，也就是decoder模块的tl\_rx\_mchit0[6]信号。在decoder模块中，当检测到mc\_tlp\_blocked信号，那么就会丢弃该TLP报文。

## MC\_Block\_Untranslated Register(Offset 20h)



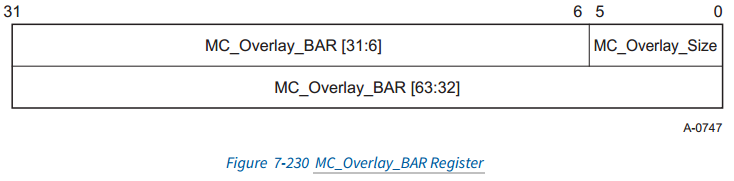
### MC\_Block\_Untranslated（已实现）

For each bit that is Set, this Function is blocked from sending TLPs containing Untranslated Addresses to the associated MCG. Bits above MC\_Num\_Group are ignored by hardware. Default value of each bit is 0b.

每设置一个比特，该功能将被阻止发送包含未转换地址的tlp到相关的MCG。高于MC\_Num\_Group的位将被硬件忽略。每个位的默认值为0b。

225已实现，具体在pcie5\_extcap模块的990行，通过判断是否cfg\_addr=12’h1A0以及cfg\_addr=12’h1A4，将全部的64bit数据cfg\_data赋值给mc\_blockunt\_r，这部分信号分别将高32bit和低32bit的值分给mcast24和mcast20两个信号。1145行中该域段mc\_blockunt\_r[mc\_group]以及tlp的Attr=00h（表示未转换的地址）以及是否配置ATS用于判断mc\_block是否置1的条件，前面已经描述了mc\_block的信号传递，这里不再赘述。

## MC\_Overlay\_BAR Register(Offset 28h)



The MC\_Overlay\_BAR Register is required in Switch and Root Complex Ports that support the Multicast Extended Capability and not implemented in Endpoints. Software must interpret the Device/Port Type field in the PCI Express Capabilities Register to determine if the MC\_Overlay\_BAR Register is present in a Function.

MC\_Overlay\_BAR寄存器需要在Switch和RC端口中支持Multicast Extended Capability，而不是在EP中实现。软件必须说明PCI Express能力寄存器中的设备/端口类型字段，以确定MC\_Overlay\_BAR寄存器是否存在于Function中。

The MC\_Overlay\_BAR specifies the base address of a window in unicast space onto which Multicast TLPs going out an Egress Port are overlaid by a process of address replacement. This allows a single BAR in an Endpoint attached to the Switch or Root Port to be used for both unicast and Multicast traffic. At a Switch Upstream Port, it allows the Multicast address range, or a portion of it, to be overlayed onto host memory.

MC\_Overlay\_BAR指定单播空间中一个窗口的基址，通过地址替换过程将从出口端口发出的多播tlp覆盖到该窗口上。这样将允许连接到Switch或Root端口的EP中的单个BAR用于单播和多播传输。在Switch的上行端口，Overlay允许多播地址范围或部分地址范围被覆盖到主机内存。

### MC\_Overlay\_Size（已实现）

If 6 or greater, specifies the size in bytes of the overlay aperture as a power of 2. If less than 6, disables the overlay mechanism. Default value is 00 0000b.

如果该域段的值大于等于6，那么指定覆盖范围的大小为2的次幂。如果该域段值小于6，就不会启用Overlay机制。默认值为00 0000b。

225已实现，在pcie5\_extcap模块中的1026行，通过判断是否cfg\_addr=12’h1A8，将6bit的cfg\_data信号全部赋值给mc\_overlay\_size\_r信号，并在拼接后赋值给mcast28[5:0]信号，这个信号和其他mcast信号拼接之后作为mcast\_cap信号的一部分输出到hsio\_top\_pipe模块，在hsio\_top\_pipe模块中对该部分信号处理方式与mc\_receive一样，需要将32个端口配置的mc\_overlay\_size拼接为mc\_overlay\_size[6\*32-1:0]，然后在decoder内部将mc\_overlay\_size [6\*32-1:0]按照之前的拼接顺序分别传递给各个与port对应的multicast模块中，在multicast模块内判断mc\_overlay\_size是否大于等于6，如果大于等于6，则需要将TLP中的地址对应替换为mc\_overlay\_bar，即为tlp\_addr[63:mc\_overlay\_size]替换为mc\_overlay\_bar [63:mc\_overlay\_size]，地址其他部分bit保持不变。

### MC\_Overlay\_BAR（已实现）

Specifies the base address of the window onto which MC TLPs passing through this Function will be overlaid. Default value is 0.

指定通过此功能传递的MC tlp将被覆盖到的窗口的基址。默认值为0。

225已实现，在pcie5\_extcap模块中的1026行，通过判断是否cfg\_addr=12’h1A8以及cfg\_addr=12’h1AC，将58bit的cfg\_data信号全部赋值给mc\_overlay\_bar\_r信号，并将mc\_overlay\_bar\_r[63:32]和mc\_overlay\_bar\_r[31:12]分别赋值给mcast2c和mcast28信号，这两个信号和其他mcast信号拼接之后作为mcast\_cap信号的一部分输出到hsio\_top\_pipe模块，在hsio\_top\_pipe模块中对该部分信号解析出mc\_overlay\_bar后，提供给hsio\_decoder\_xx模块使用。

这部分信号的操作和mc\_overlay\_size操作一样，在hsio\_top\_pipe模块中将32个端口配置的mc\_overlay\_bar拼接为mc\_overlay\_bar[58\*32-1:0]，然后分别给各个port对应的multicast使用，具体使用方法为前面mc\_overlay\_size中的描述，这里不在赘述。

在225的代码实现中，在mc\_overlay机制打开情况下，才会存在TLP报文中的地址字段替换，此时会存在四种情况，也就是32bit到32bit、32bit到64bit、64bit到32bit以及64bit到64bit。此时如何判断替换的地址为32bit或者64bit是根据mc\_overlay\_bar地址的高32bit，因为替换mc\_overlay\_bar地址过程中会打拍寄存，而在TLP报文拼接修改过程中会判断打拍寄存的mc\_overlay\_bar地址的高32bit是否为0。如果mc\_overlay\_bar地址的高32bit为0，那么就认为该mc\_overlay\_bar地址为32bit，并且只截取mc\_overlay\_bar[31:0]作为TLP的替换地址；如果mc\_overlay\_bar地址的高32bit不为0，那么就认为该mc\_overlay\_bar地址为64bit；此时便判断出替换后的地址是32bit还是64bit。

协议中在mc\_overlay机制打开时，只会关心mc\_overlay\_bar地址从哪bit开始替换，并不关心是32bit还是64bit，真正关心这个问题的是225的multicast模块在发送TLP时候，需要将原TLP的地址字段替换为mc\_overlay\_bar的地址，这个时候会判断替换的地址是32bit还是64bit。

并且一旦开启了MC\_Overlay机制，overlay操作会把组播TLP地址中比特数大于等于MC\_Overlay\_Size的部分替换为MC\_Overlay\_BAR中的对应比特。即：

Egress\_TLP\_Addr=(MC\_Overlay\_Size<6)?Ingress\_TLP\_Addr :{ MC\_Overlay\_BAR[63:MC\_Overlay\_Size], Ingress\_TLP\_Addr[MC\_Overlay\_Size-1:0] }。