



15<sup>th</sup> ANNUAL WORKSHOP 2019

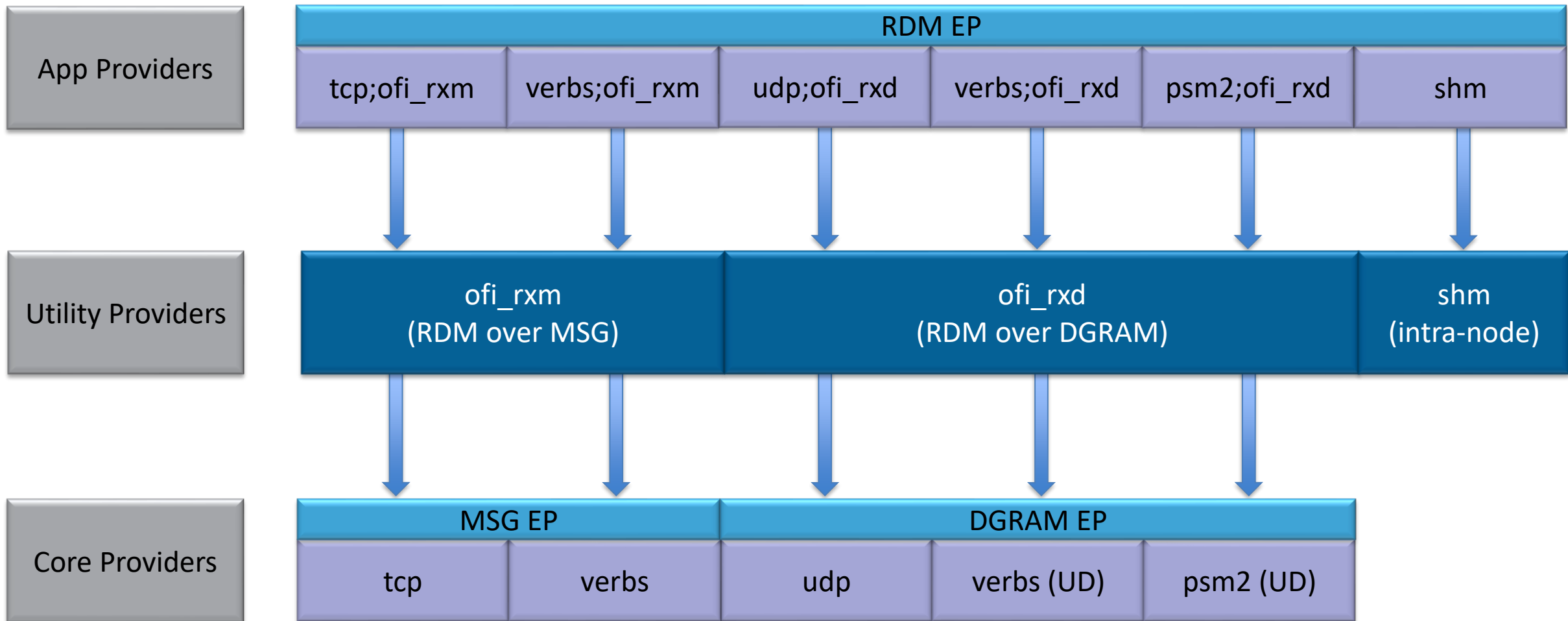
# AN OVERVIEW OF OFI UTILITY PROVIDERS: IMPLEMENTATION AND USES

Alexia Ingerson

Intel® Corporation

March 2019

# STRUCTURE OVERVIEW

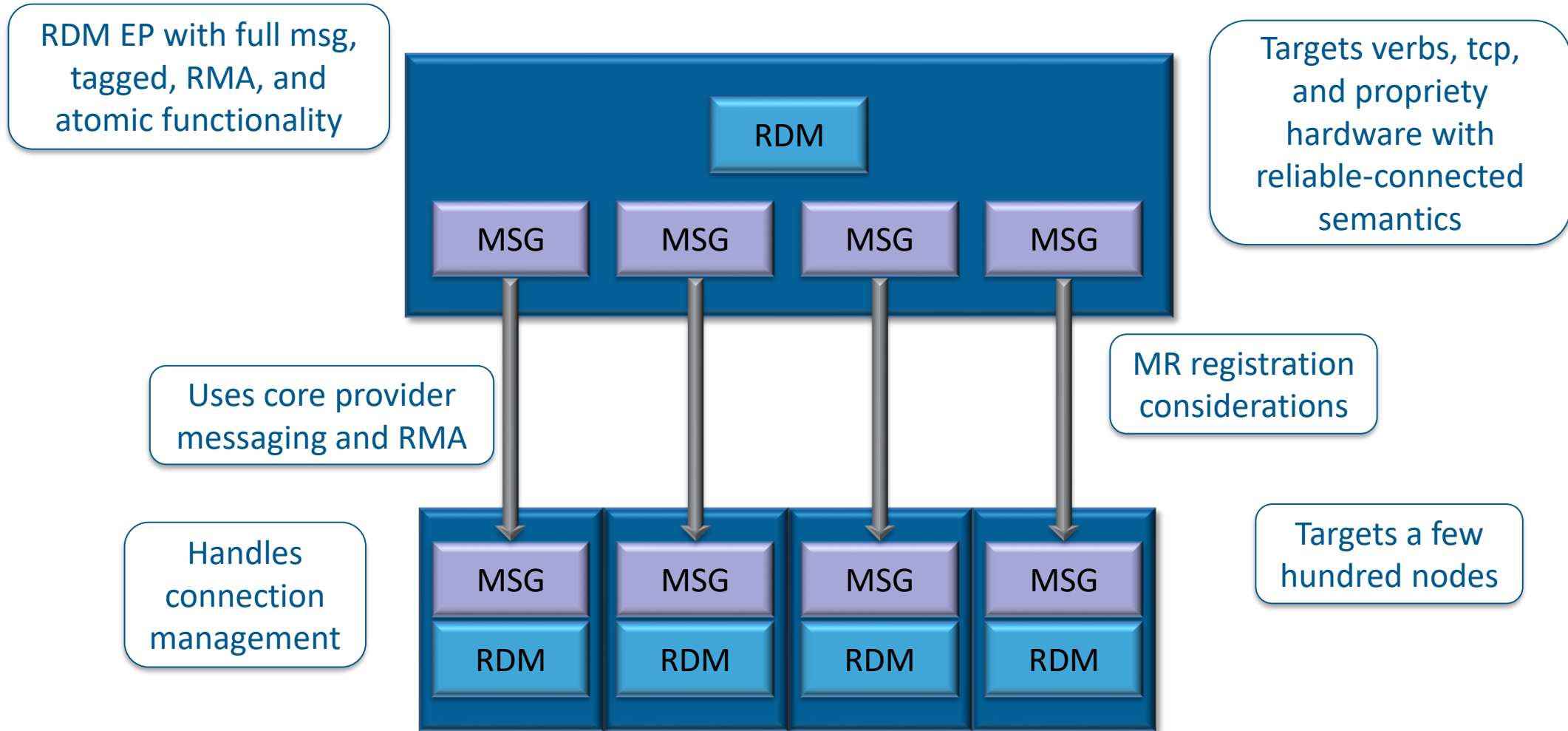






# RXM PROVIDER

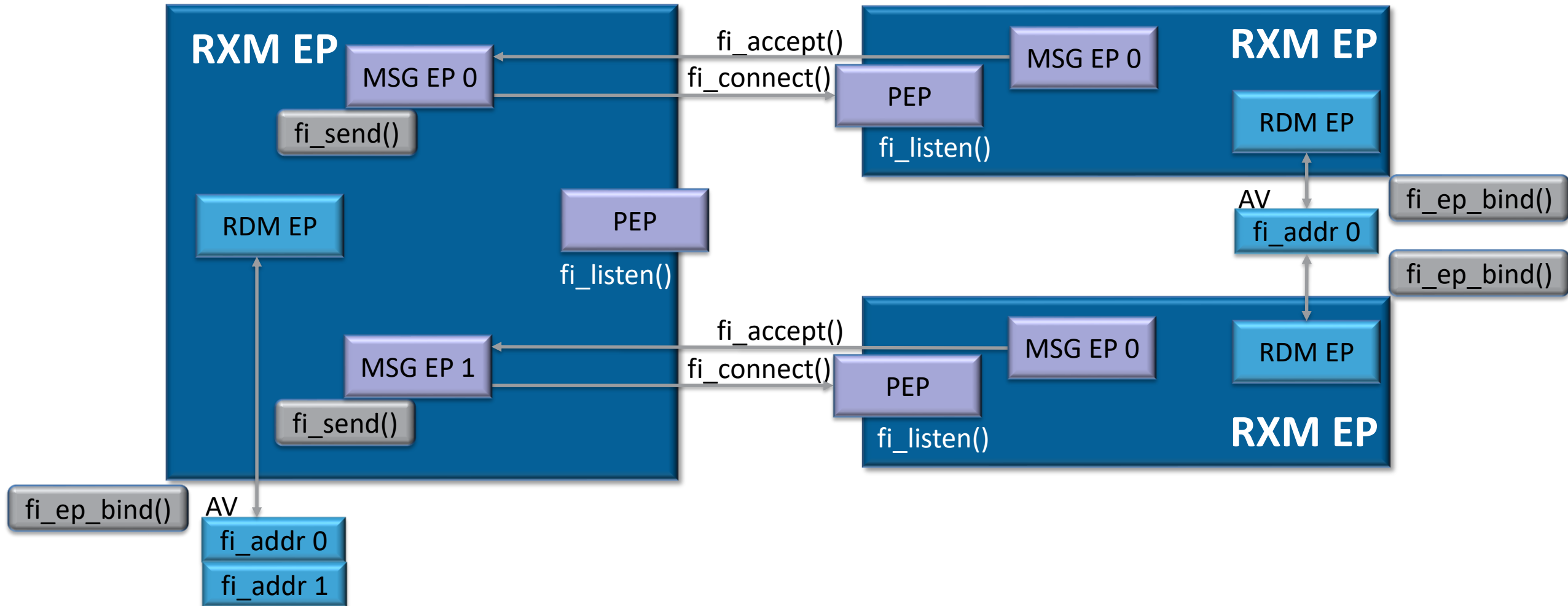
# RXM PROVIDER





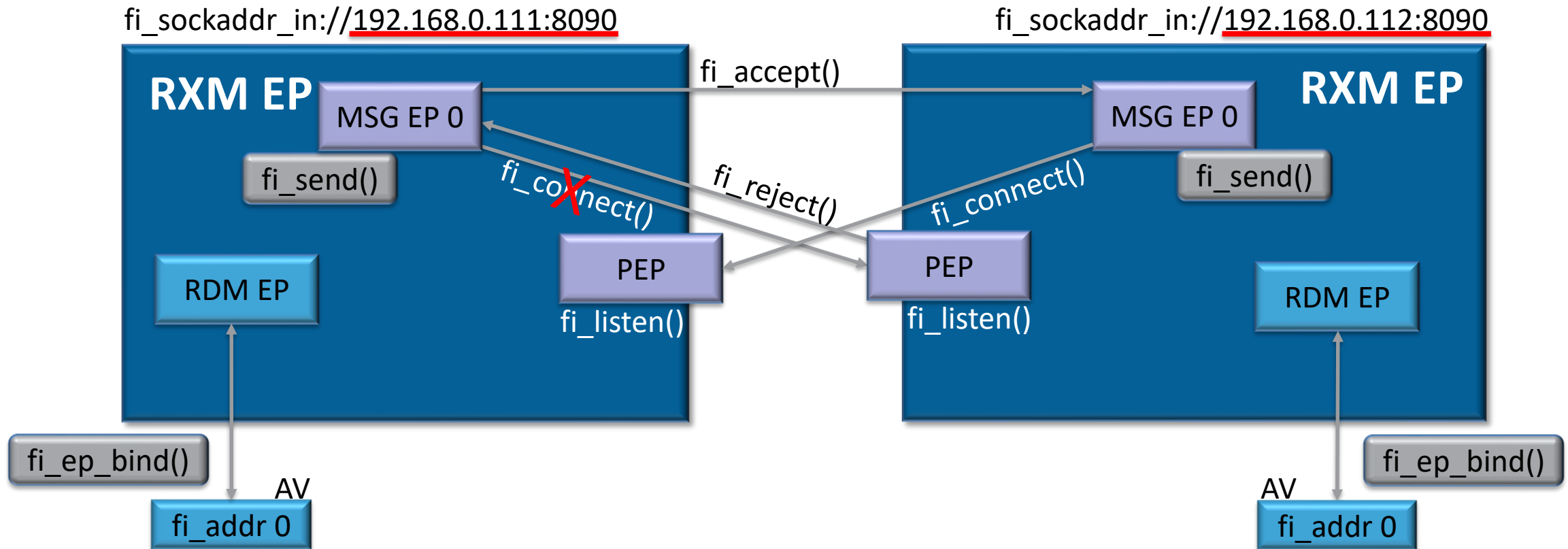
# RXM CONNECTION MANAGEMENT

## On-demand connections:



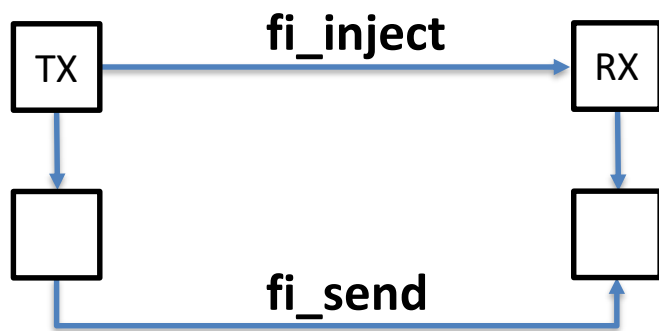
# RXM CONNECTION MANAGEMENT

## Simultaneous connections:



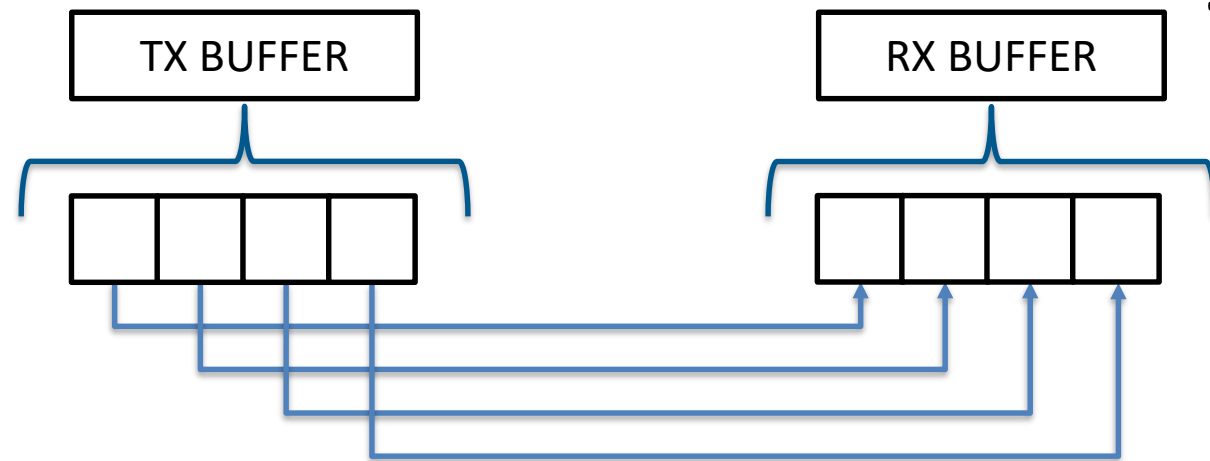
# RXM PROTOCOLS

## Eager

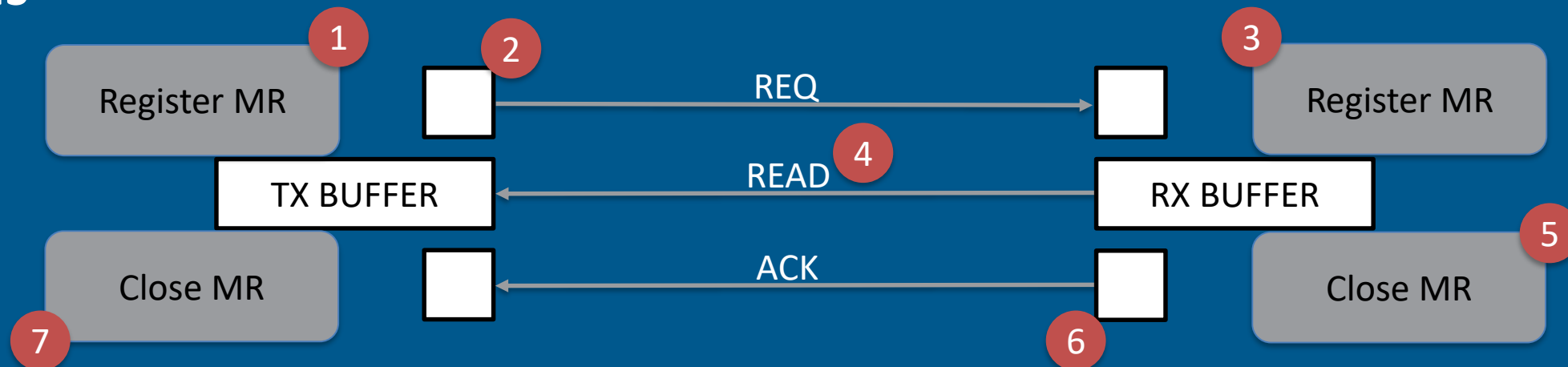


Pre-registered  
tx/rx MRs

## SAR



## Rendezvous



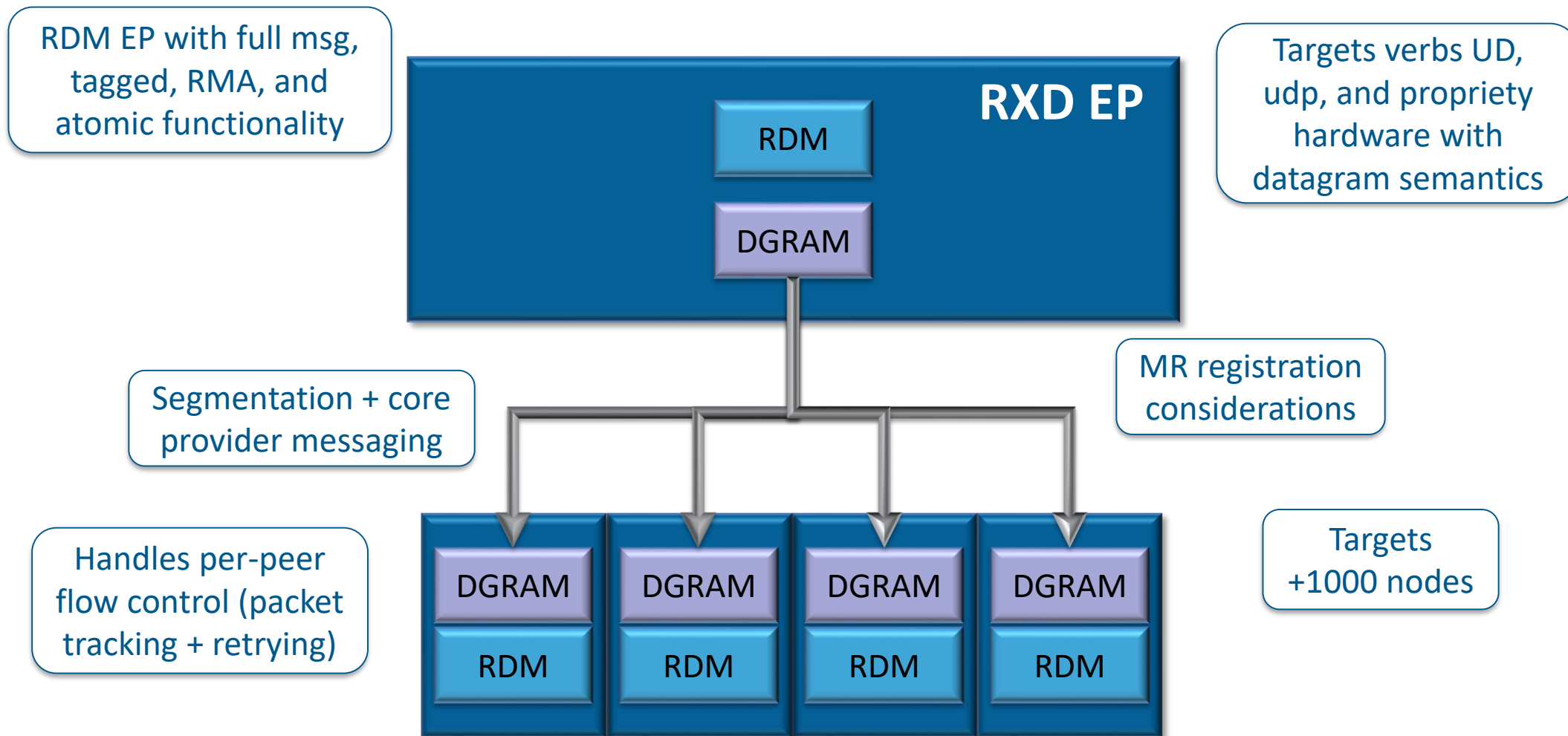




# RXD PROVIDER

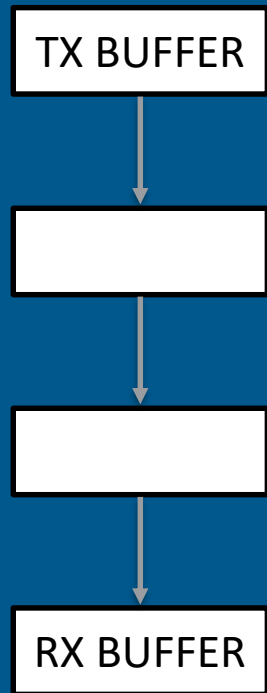


# RXD PROVIDER



# RXD PROTOCOLS

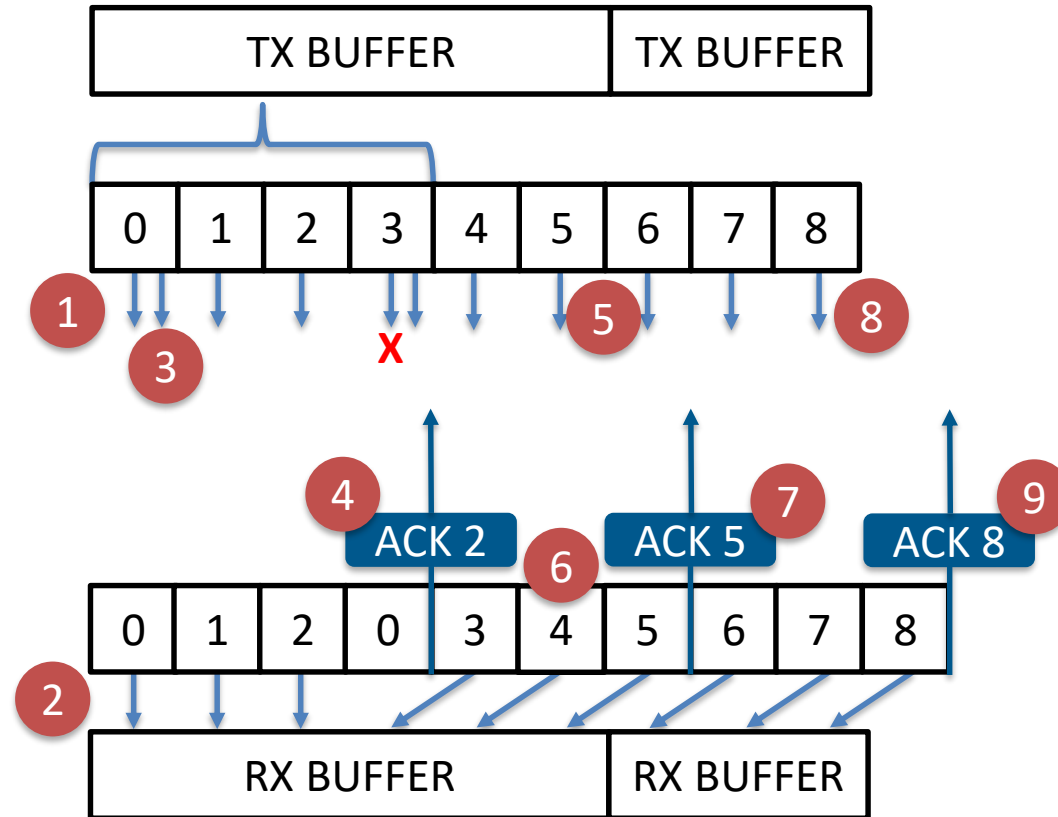
## Eager



Pre-registered  
tx/rx MRs

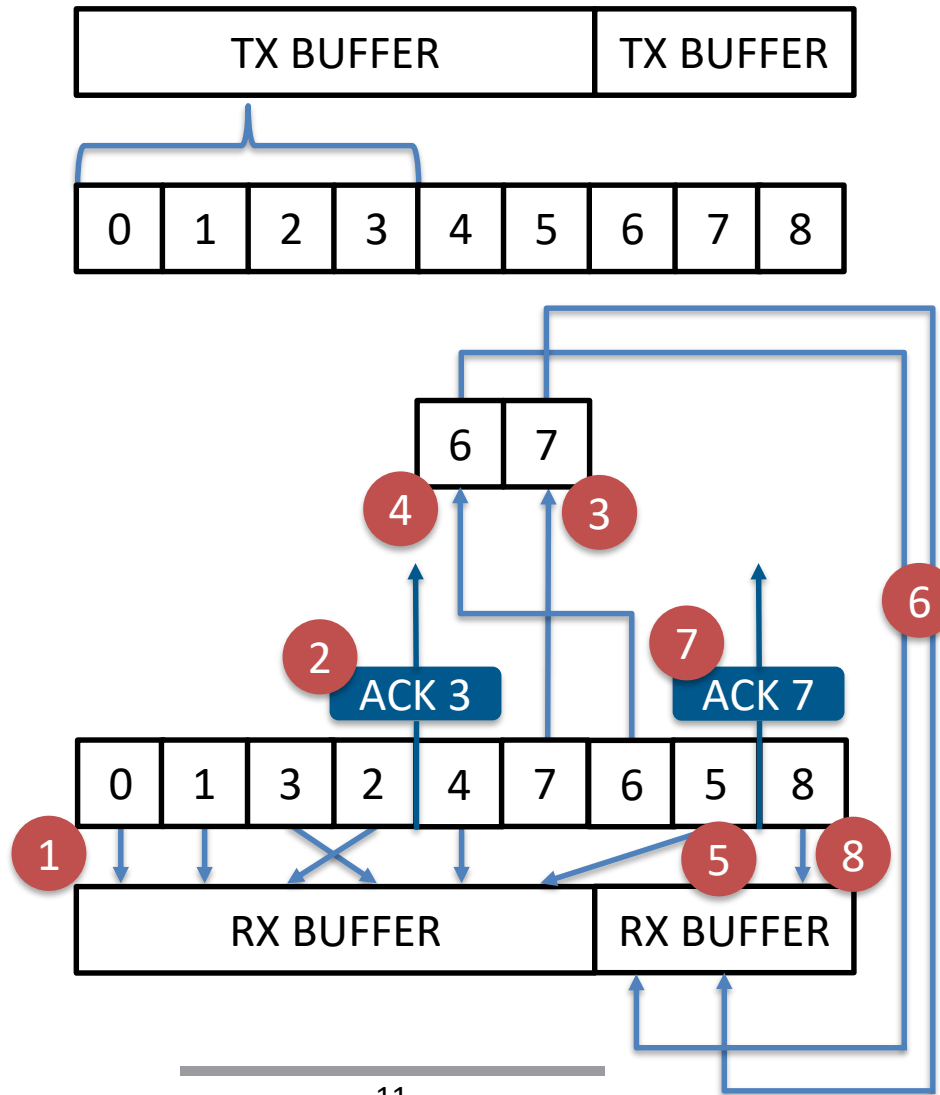
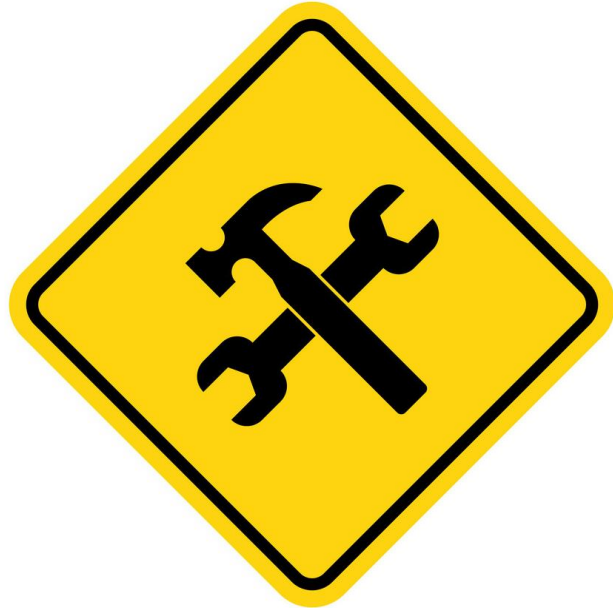
Time and retry  
packets if  
unacked

## SAR





# OUT OF ORDER REASSEMBLY



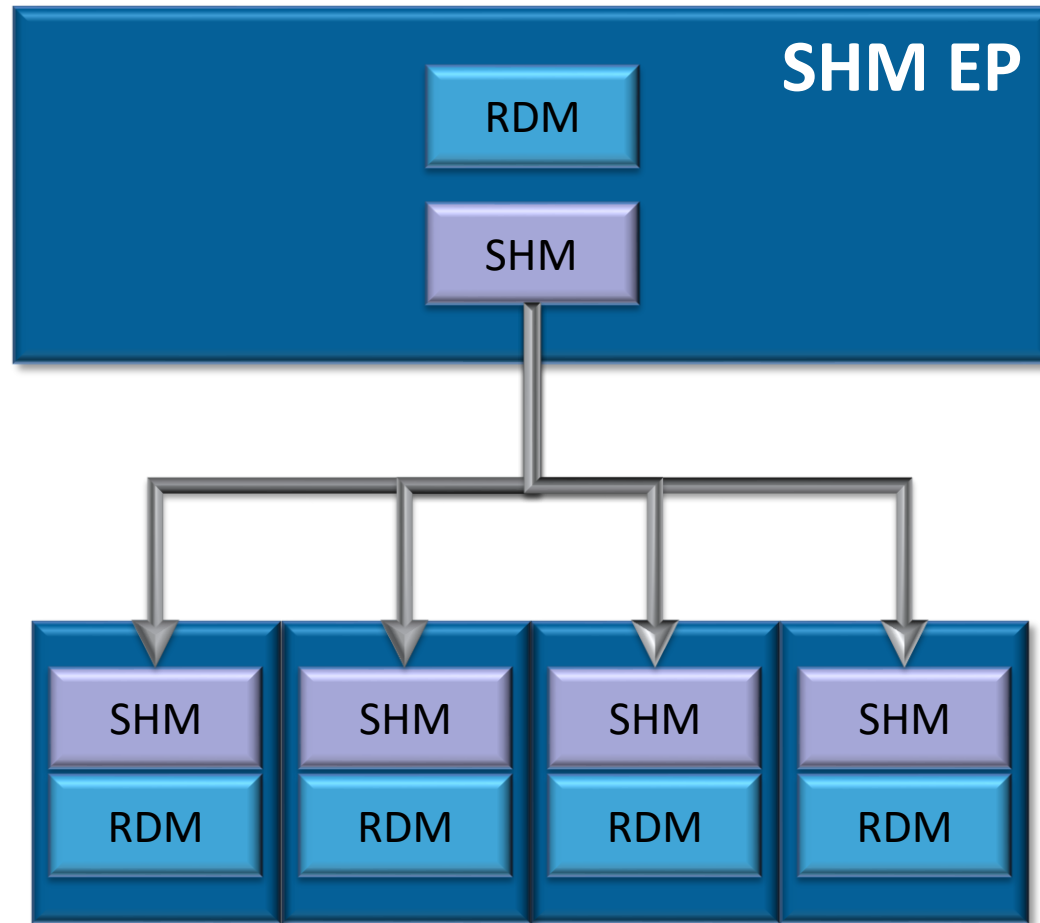


# SHM PROVIDER



# SHM PROVIDER

RDM EP with full msg,  
tagged, RMA, and  
atomic functionality



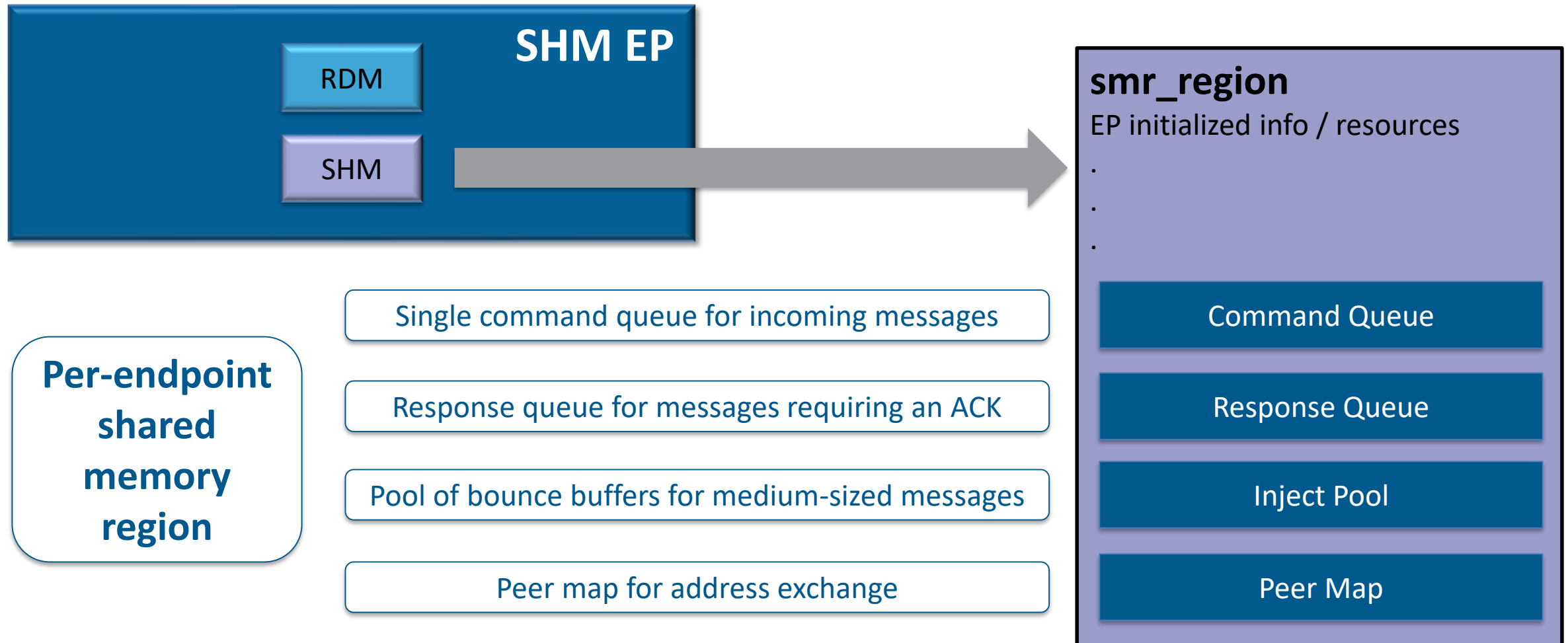
Stand alone  
provider for local  
communication

Per-endpoint  
shared memory  
region

Disabled on non-  
Linux platforms  
(CMA)

xpmem in  
progress

# SHARED MEMORY STRUCTURE





# SHM ADDRESSING

Endpoint name: 11111:0:0

. . .

0	22222	UNSPEC
1	33333	UNSPEC
2	44444	UNSPEC

fi\_av\_insert

fi\_av\_insert

fi\_av\_insert

fi\_av\_insert

Endpoint name: 22222:0:0

. . .

0	11111	0
1	33333	1
2	44444	1

Endpoint name: 33333:0:0

. . .

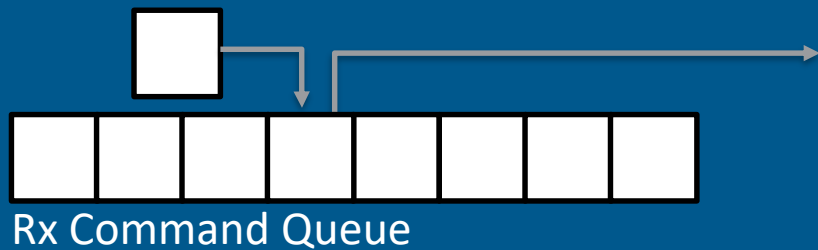
0	11111	1
1	22222	UNSPEC
2	44444	2

Endpoint name: 44444:0:0

. . .

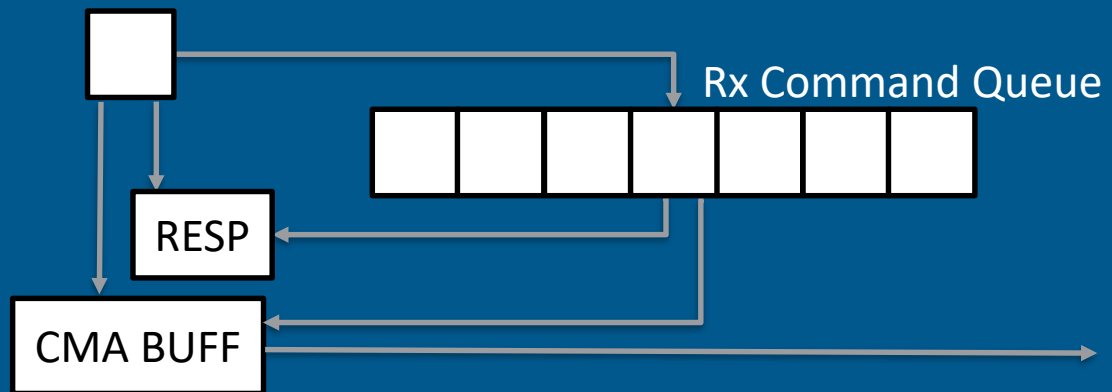
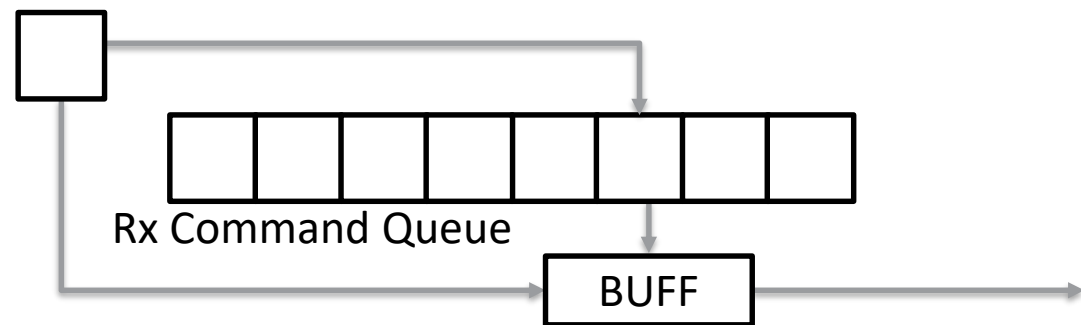
0	11111	2
1	22222	UNSPEC
2	33333	UNSPEC

# SHM PROTOCOLS



**Inline**

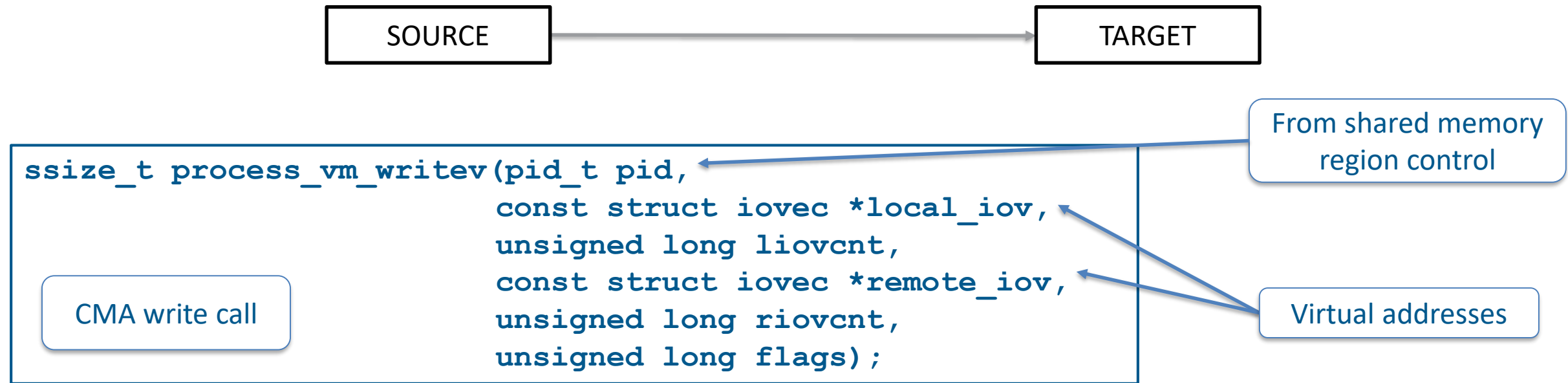
**Inject**



**IOV**



# SHM RMA FAST PATH



Requirements for fast path:

- ✓ `mr_mode: FI_MR_VIRT_ADDR`
- ✓ no RMA ordering requested (RAR, RAW, RAS, WAR, WAW, WAS, SAR, SAW)
- ✓ message is not sending remote CQ data

# FUTURE ENHANCEMENTS

## RXM

Reduce header overhead

MR cache improvements

Improve start-up times

Add HW-specific optimizations

Rendezvous read -> write

## mrail

Under development

Threading, progress

Performance

Non-host memory transfers

xpmem

## SHM

## RXD

Threading, progress

Performance

Packet buffering

Scaling, flow control

Reliable reassembly





15<sup>th</sup> ANNUAL WORKSHOP 2019

# THANK YOU

Alexia Ingerson

Intel® Corporation



# LEGAL DISCLAIMER & OPTIMIZATION NOTICE

- Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information visit [www.intel.com/benchmarks](http://www.intel.com/benchmarks).
- INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS”. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. INTEL ASSUMES NO LIABILITY WHATSOEVER AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO THIS INFORMATION INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT.
- Copyright © 2018, Intel Corporation. All rights reserved. Intel, Pentium, Xeon, Xeon Phi, Core, VTune, Cilk, and the Intel logo are trademarks of Intel Corporation in the U.S. and other countries.

## Optimization Notice

Intel’s compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Notice revision #20110804