[NCCL](../index.html)

[2.25](https://docs.nvidia.com/deeplearning/sdk/nccl-archived/index.html)

* [Communicator Creation and Management Functions](comms.html)
  * [ncclGetLastError](comms.html#ncclgetlasterror)
  * [ncclGetErrorString](comms.html#ncclgeterrorstring)
  * [ncclGetVersion](comms.html#ncclgetversion)
  * [ncclGetUniqueId](comms.html#ncclgetuniqueid)
  * [ncclCommInitRank](comms.html#ncclcomminitrank)
  * [ncclCommInitAll](comms.html#ncclcomminitall)
  * [ncclCommInitRankConfig](comms.html#ncclcomminitrankconfig)
  * [ncclCommInitRankScalable](comms.html#ncclcomminitrankscalable)
  * [ncclCommSplit](comms.html#ncclcommsplit)
  * [ncclCommFinalize](comms.html#ncclcommfinalize)
  * [ncclCommDestroy](comms.html#ncclcommdestroy)
  * [ncclCommAbort](comms.html#ncclcommabort)
  * [ncclCommGetAsyncError](comms.html#ncclcommgetasyncerror)
  * [ncclCommCount](comms.html#ncclcommcount)
  * [ncclCommCuDevice](comms.html#ncclcommcudevice)
  * [ncclCommUserRank](comms.html#ncclcommuserrank)
  * [ncclCommRegister](comms.html#ncclcommregister)
  * [ncclCommDeregister](comms.html#ncclcommderegister)
  * [ncclMemAlloc](comms.html#ncclmemalloc)
  * [ncclMemFree](comms.html#ncclmemfree)
* Collective Communication Functions
  * ncclAllReduce
  * ncclBroadcast
  * ncclReduce
  * ncclAllGather
  * ncclReduceScatter

MPI](../mpi.html#inter-gpu-communication-with-cuda-aware-mpi)

* [Environment Variables](../env.html)

  * [System configuration](../env.html#system-configuration)

    * [NCCL_SOCKET_IFNAME](../env.html#nccl-socket-ifname)

      * [Values accepted](../env.html#values-accepted)

    * [NCCL_SOCKET_FAMILY](../env.html#nccl-socket-family)

      * [Values accepted](../env.html#id2)

    * [NCCL_SOCKET_RETRY_CNT](../env.html#nccl-socket-retry-cnt)

      * [Values accepted](../env.html#id3)

    * [NCCL_SOCKET_RETRY_SLEEP_MSEC](../env.html#nccl-socket-retry-sleep-msec)

      * [Values accepted](../env.html#id4)

    * [NCCL_SOCKET_NTHREADS](../env.html#nccl-socket-nthreads)

      * [Values accepted](../env.html#id5)

    * [NCCL_NSOCKS_PERTHREAD](../env.html#nccl-nsocks-perthread)

      * [Values accepted](../env.html#id6)

    * [NCCL_CROSS_NIC](../env.html#nccl-cross-nic)

      * [Values accepted](../env.html#id7)

    * [NCCL_IB_HCA](../env.html#nccl-ib-hca)

      * [Values accepted](../env.html#id8)

    * [NCCL_IB_TIMEOUT](../env.html#nccl-ib-timeout)

      * [Values accepted](../env.html#id9)

    * [NCCL_IB_RETRY_CNT](../env.html#nccl-ib-retry-cnt)

      * [Values accepted](../env.html#id10)

    * [NCCL_IB_GID_INDEX](../env.html#nccl-ib-gid-index)

      * [Values accepted](../env.html#id11)

    * [NCCL_IB_ADDR_FAMILY](../env.html#nccl-ib-addr-family)

      * [Values accepted](../env.html#id12)

* [NCCL_IB_ADDR_RANGE](../env.html#nccl-ib-addr-range)
  * [Values accepted](../env.html#id13)
* [NCCL_IB_ROCE_VERSION_NUM](../env.html#nccl-ib-roce-version-num)
  * [Values accepted](../env.html#id14)
* [NCCL_IB_SL](../env.html#nccl-ib-sl)
  * [Values accepted](../env.html#id15)
* [NCCL_IB_TC](../env.html#nccl-ib-tc)
  * [Values accepted](../env.html#id16)
* [NCCL_IB_FIFO_TC](../env.html#nccl-ib-fifo-tc)
  * [Values accepted](../env.html#id17)
* [NCCL_IB_RETURN_ASYNC_EVENTS](../env.html#nccl-ib-return-async-events)
  * [Values accepted](../env.html#id18)
* [NCCL_OOB_NET_ENABLE](../env.html#nccl-oob-net-enable)
  * [Values accepted](../env.html#id19)
* [NCCL_OOB_NET_IFNAME](../env.html#nccl-oob-net-ifname)
  * [Values accepted](../env.html#id20)
* [NCCL_UID_STAGGER_THRESHOLD](../env.html#nccl-uid-stagger-threshold)
  * [Values accepted](../env.html#id21)
* [NCCL_UID_STAGGER_RATE](../env.html#nccl-uid-stagger-rate)
  * [Values accepted](../env.html#id22)
* [NCCL_NET](../env.html#nccl-net)
  * [Values accepted](../env.html#id23)
* [NCCL_NET_PLUGIN](../env.html#nccl-net-plugin)
  * [Values accepted](../env.html#id24)
* [NCCL_TUNER_PLUGIN](../env.html#nccl-tuner-plugin)
  * [Values accepted](../env.html#id25)
* [NCCL_PROFILER_PLUGIN](../env.html#nccl-profiler-plugin)

* [NCCL_P2P_DIRECT_DISABLE](../env.html#nccl-p2p-direct-disable)
  * [Values accepted](../env.html#id39)
* [NCCL_SHM_DISABLE](../env.html#nccl-shm-disable)
  * [Values accepted](../env.html#id40)
* [NCCL_BUFFSIZE](../env.html#nccl-buffsize)
  * [Values accepted](../env.html#id41)
* [NCCL_NTHREADS](../env.html#nccl-nthreads)
  * [Values accepted](../env.html#id42)
* [NCCL_MAX_NCHANNELS](../env.html#nccl-max-nchannels)
  * [Values accepted](../env.html#id43)
* [NCCL_MIN_NCHANNELS](../env.html#nccl-min-nchannels)
  * [Values accepted](../env.html#id44)
* [NCCL_CHECKS_DISABLE](../env.html#nccl-checks-disable)
  * [Values accepted](../env.html#id45)
* [NCCL_CHECK_POINTERS](../env.html#nccl-check-pointers)
  * [Values accepted](../env.html#id46)
* [NCCL_LAUNCH_MODE](../env.html#nccl-launch-mode)
  * [Values accepted](../env.html#id47)
* [NCCL_IB_DISABLE](../env.html#nccl-ib-disable)
  * [Values accepted](../env.html#id48)
* [NCCL_IB_AR_THRESHOLD](../env.html#nccl-ib-ar-threshold)
  * [Values accepted](../env.html#id49)
* [NCCL_IB_QPS_PER_CONNECTION](../env.html#nccl-ib-qps-per-connection)
  * [Values accepted](../env.html#id50)
* [NCCL_IB_SPLIT_DATA_ON_QPS](../env.html#nccl-ib-split-data-on-qps)
  * [Values accepted](../env.html#id51)
* [NCCL_IB_CUDA_SUPPORT](../env.html#nccl-ib-cuda-support)

  * [Values accepted](../env.html#id52)
* [NCCL_IB_PCI_RELAXED_ORDERING](../env.html#nccl-ib-pci-relaxed-ordering)
  * [Values accepted](../env.html#id53)
* [NCCL_IB_ADAPTIVE_ROUTING](../env.html#nccl-ib-adaptive-routing)
  * [Values accepted](../env.html#id54)
* [NCCL_IB_ECE_ENABLE](../env.html#nccl-ib-ece-enable)
  * [Values accepted](../env.html#id55)
* [NCCL_MEM_SYNC_DOMAIN](../env.html#nccl-mem-sync-domain)
  * [Values accepted](../env.html#id56)
* [NCCL_CUMEM_ENABLE](../env.html#nccl-cumem-enable)
  * [Values accepted](../env.html#id57)
* [NCCL_CUMEM_HOST_ENABLE](../env.html#nccl-cumem-host-enable)
  * [Values accepted](../env.html#id58)
* [NCCL_NET_GDR_LEVEL (formerly NCCL_IB_GDR_LEVEL)](../env.html#nccl-net-gdr-level-formerly-nccl-ib-gdr-level)
  * [Values accepted](../env.html#id59)
  * [Integer Values (Legacy)](../env.html#id60)
* [NCCL_NET_GDR_READ](../env.html#nccl-net-gdr-read)
  * [Values accepted](../env.html#id61)
* [NCCL_NET_SHARED_BUFFERS](../env.html#nccl-net-shared-buffers)
  * [Value accepted](../env.html#id62)
* [NCCL_NET_SHARED_COMMS](../env.html#nccl-net-shared-comms)
  * [Value accepted](../env.html#id63)
* [NCCL_SINGLE_RING_THRESHOLD](../env.html#nccl-single-ring-threshold)
  * [Values accepted](../env.html#id64)
* [NCCL_LL_THRESHOLD](../env.html#nccl-ll-threshold)
  * [Values accepted](../env.html#id65)

* [NCCL_TREE_THRESHOLD](../env.html#nccl-tree-threshold)
  * [Values accepted](../env.html#id66)
* [NCCL_ALGO](../env.html#nccl-algo)
  * [Values accepted](../env.html#id67)
* [NCCL_PROTO](../env.html#nccl-proto)
  * [Values accepted](../env.html#id68)
* [NCCL_NVB_DISABLE](../env.html#nccl-nvb-disable)
  * [Value accepted](../env.html#id69)
* [NCCL_PXN_DISABLE](../env.html#nccl-pxn-disable)
  * [Value accepted](../env.html#id70)
* [NCCL_P2P_PXN_LEVEL](../env.html#nccl-p2p-pxn-level)
  * [Value accepted](../env.html#id71)
* [NCCL_RUNTIME_CONNECT](../env.html#nccl-runtime-connect)
  * [Value accepted](../env.html#id72)
* [NCCL_GRAPH_REGISTER](../env.html#nccl-graph-register)
  * [Value accepted](../env.html#id74)
* [NCCL_LOCAL_REGISTER](../env.html#nccl-local-register)
  * [Value accepted](../env.html#id75)
* [NCCL_LEGACY_CUDA_REGISTER](../env.html#nccl-legacy-cuda-register)
  * [Value accepted](../env.html#id76)
* [NCCL_SET_STACK_SIZE](../env.html#nccl-set-stack-size)
  * [Value accepted](../env.html#id77)
* [NCCL_GRAPH_MIXING_SUPPORT](../env.html#nccl-graph-mixing-support)
  * [Value accepted](../env.html#id79)
* [NCCL_DMABUF_ENABLE](../env.html#nccl-dmabuf-enable)
  * [Value accepted](../env.html#id80)
* [NCCL_P2P_NET_CHUNKSIZE](../env.html#nccl-p2p-net-chunksize)

__[NCCL](../index.html)

* * *

Collective Communication FunctionsÂ¶

The following NCCL APIs provide some commonly used collective operations.

ncclAllReduceÂ¶

[ncclResult_t](types.html#c.ncclResult_t "ncclResult_t") `ncclAllReduce`(const
void*_ sendbuff_, void*_ recvbuff_, size_t _ count_,
[ncclDataType_t](types.html#c.ncclDataType_t "ncclDataType_t") _ datatype_,
[ncclRedOp_t](types.html#c.ncclRedOp_t "ncclRedOp_t") _ op_,
[ncclComm_t](types.html#c.ncclComm_t "ncclComm_t") _ comm_, cudaStream_t _
stream_)Â¶

Reduces data arrays of length `count` in `sendbuff` using the `op` operation
and leaves identical copies of the result in each `recvbuff`.

In-place operation will happen if `sendbuff == recvbuff`.

Related links: [AllReduce](../usage/collectives.html#allreduce).

ncclBroadcastÂ¶

[ncclResult_t](types.html#c.ncclResult_t "ncclResult_t") `ncclBroadcast`(const

void*_ sendbuff_, void*_ recvbuff_, size_t _ count_,

[ncclDataType_t](types.html#c.ncclDataType_t "ncclDataType_t") _ datatype_,

int _ root_, [ncclComm_t](types.html#c.ncclComm_t "ncclComm_t") _ comm_,

cudaStream_t _ stream_)Â¶

Copies `count` elements from `sendbuff` on the `root` rank to all ranksâ€™
`recvbuff`. `sendbuff` is only used on rank `root` and ignored for other
ranks.

In-place operation will happen if `sendbuff == recvbuff`.

[ncclResult_t](types.html#c.ncclResult_t "ncclResult_t") `ncclBcast`(void*_
buff_, size_t _ count_, [ncclDataType_t](types.html#c.ncclDataType_t
"ncclDataType_t") _ datatype_, int _ root_,
[ncclComm_t](types.html#c.ncclComm_t "ncclComm_t") _ comm_, cudaStream_t _
stream_)Â¶

Legacy in-place version of `ncclBroadcast` in a similar fashion to MPI_Bcast.
A call to

```
ncclBcast(buff, count, datatype, root, comm, stream)
```

is equivalent to

```
ncclBroadcast(buff, buff, count, datatype, root, comm, stream)
```

Related links: [Broadcast](../usage/collectives.html#broadcast)

ncclReduceÂ¶

[ncclResult_t](types.html#c.ncclResult_t "ncclResult_t") `ncclReduce`(const
void*_ sendbuff_, void*_ recvbuff_, size_t _ count_,
[ncclDataType_t](types.html#c.ncclDataType_t "ncclDataType_t") _ datatype_,
[ncclRedOp_t](types.html#c.ncclRedOp_t "ncclRedOp_t") _ op_, int _ root_,
[ncclComm_t](types.html#c.ncclComm_t "ncclComm_t") _ comm_, cudaStream_t _
stream_)Â¶

Reduce data arrays of length `count` in `sendbuff` into `recvbuff` on the
`root` rank using the `op` operation. `recvbuff` is only used on rank `root`
and ignored for other ranks.

In-place operation will happen if `sendbuff == recvbuff`.

Related links: [Reduce](../usage/collectives.html#reduce).

ncclAllGatherÂ¶

[ncclResult_t](types.html#c.ncclResult_t "ncclResult_t") `ncclAllGather`(const void*_ sendbuff_, void*_ recvbuff_, size_t _ sendcount_, [ncclDataType_t](types.html#c.ncclDataType_t "ncclDataType_t") _ datatype_, [ncclComm_t](types.html#c.ncclComm_t "ncclComm_t") _ comm_, cudaStream_t _ stream_)Â¶

Gathers `sendcount` values from all GPUs and leaves identical copies of the result in each `recvbuff`, receiving data from rank `i` at offset `i*sendcount`.

Note: This assumes the receive count is equal to `nranks*sendcount`, which means that `recvbuff` should have a size of at least `nranks*sendcount` elements.

In-place operation will happen if `sendbuff == recvbuff + rank * sendcount`.

Related links: [AllGather](../usage/collectives.html#allgather), [In-place Operations](../usage/inplace.html#in-place-operations).

ncclReduceScatterÂ¶

[ncclResult_t](types.html#c.ncclResult_t "ncclResult_t")
`ncclReduceScatter`(const void*_ sendbuff_, void*_ recvbuff_, size_t _
recvcount_, [ncclDataType_t](types.html#c.ncclDataType_t "ncclDataType_t") _
datatype_, [ncclRedOp_t](types.html#c.ncclRedOp_t "ncclRedOp_t") _ op_,
[ncclComm_t](types.html#c.ncclComm_t "ncclComm_t") _ comm_, cudaStream_t _
stream_)Â¶

Reduce data in `sendbuff` from all GPUs using the `op` operation and leave the
reduced result scattered over the devices so that the `recvbuff` on rank `i`
will contain the i-th block of the result.

Note: This assumes the send count is equal to `nranks*recvcount`, which means
that `sendbuff` should have a size of at least `nranks*recvcount` elements.

In-place operation will happen if `recvbuff == sendbuff + rank * recvcount`.

Related links: [ReduceScatter](../usage/collectives.html#reducescatter), [In-
place Operations](../usage/inplace.html#in-place-operations).

[Next](group.html "Group Calls") [Previous](comms.html "Communicator
Creation and Management Functions")

* * *