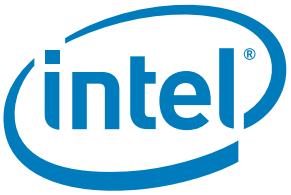


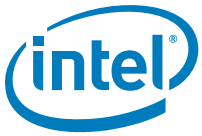
# Intel® Core™ X-Series Processor Families

## Datasheet, Volume 1 of 2

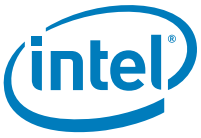
***Supporting Intel® Core™ X-Series Processor Families – 7740X, 7640X***

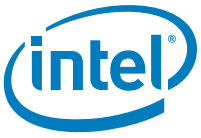
***May 2017***

Document Number: 335898-001



# Contents





## Figures

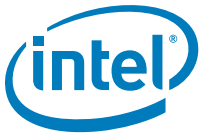


## Tables

# Revision History

| Revision Number | Description | Revision Date |
|---|---|---|
| 001 | • Initial release | May 2017 |

§ §



# 1 Introduction

The Intel® Core™ X-Series processor families are 64-bit, multi-core processors built on 14-nanometer process technology.

The High End Desktop (HEDT) X-Processors are offered in a 2-Chip Platform. The X-Processor Line is connected to a discrete Intel® X299 Series Chipset Family Platform Controller Hub (PCH). See the following figure.

The following table describes the processor line covered in this document.

**Table 1-1. Processor Lines**

| Processor Line | Package | Base TDP | Processor IA Cores | Graphics Configuration | On-Package Cache | Platform Type |
|---|---|---|---|---|---|---|
| X-Processor Line (HEDT) | LGA2066 | 112W | 4 | GT0 | N/A | 2-Chip |

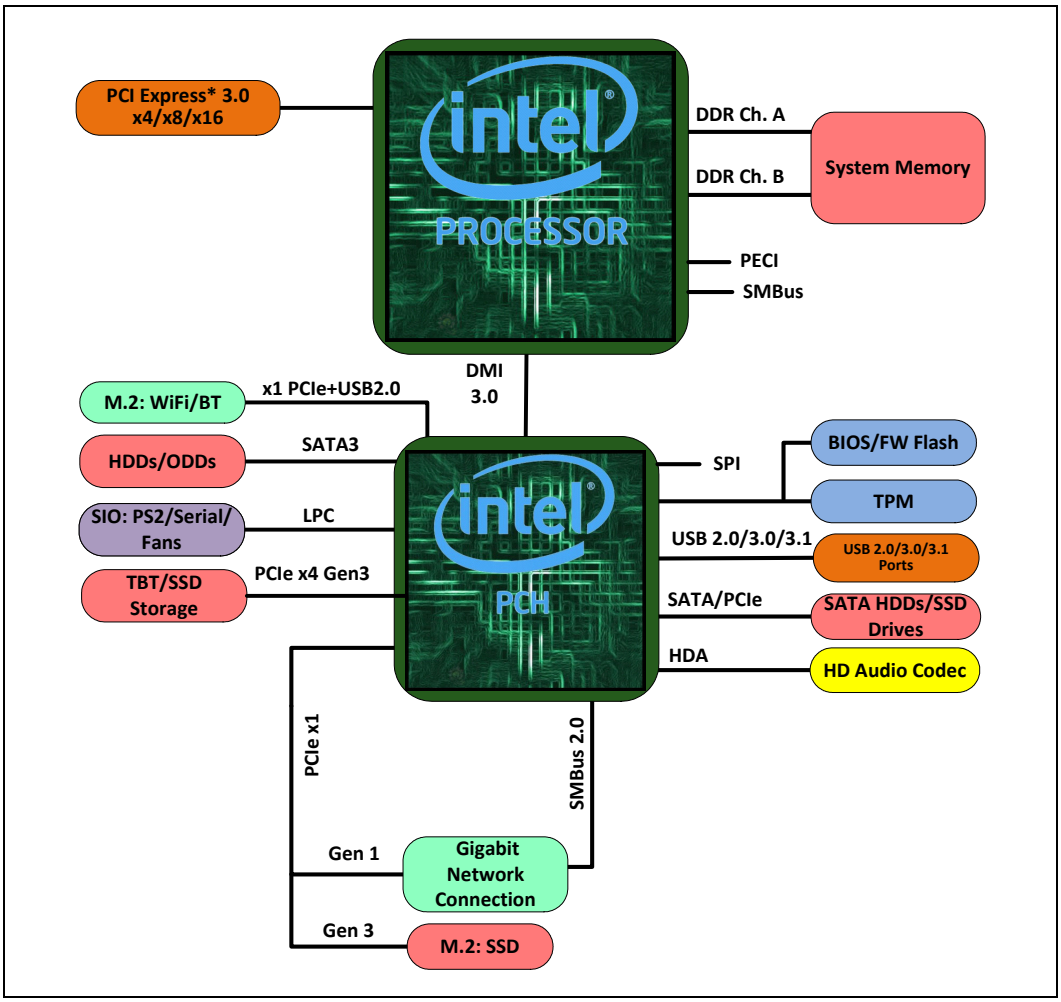
Throughout this document, the Intel® Core™ X-Series processor families may be referred to simply as "processor". The Intel® X299 Series Chipset Family Platform Controller Hub (PCH) may be referred to simply as "PCH".

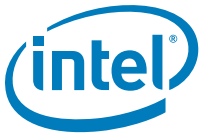
This document is for the following Intel® Core™ X-Series processor families:

- 7740X, 7640X

Refer to the processor Specification Update for additional SKU details.

**Figure 1-1. X-Processor Line Platforms**





## 1.1 Supported Technologies

- Intel® Virtualization Technology (Intel® VT)
- Intel® Active Management Technology 11.0 (Intel® AMT 11.0)
- Intel® Streaming SIMD Extensions 4.2 (Intel® SSE4.2)
- Intel® Hyper-Threading Technology (Intel® HT Technology)
- Intel® 64 Architecture
- Execute Disable Bit
- Intel® Turbo Boost Technology 2.0
- Intel® Advanced Vector Extensions 2 (Intel® AVX2)
- Intel® Advanced Encryption Standard New Instructions (Intel® AES-NI)
- PCLMULQDQ (Perform Carry-Less Multiplication Quad word) Instruction
- Intel® Secure Key
- Intel® Transactional Synchronization Extensions (Intel® TSX-NI)
- PAIR – Power Aware Interrupt Routing
- SMEP – Supervisor Mode Execution Protection
- Intel® Memory Protection Extensions (Intel® MPX)
- GMM Scoring Accelerator
- Intel® Processor Trace
- High-bandwidth Digital Content Protection (HDCP)

***Note:*** The availability of the features may vary between processor SKUs.

Refer to Chapter 3, "Technologies" for more information.

## 1.2 Power Management Support

### 1.2.1 Processor Core Power Management

- Full support of ACPI C-states as implemented by the following processor C-states:
  - C0, C1, C1E, C3, C6, C7, C8
- Enhanced Intel SpeedStep® Technology

Refer to Section 4.2, "Processor IA Core Power Management" for more information.

### 1.2.2 System Power Management

- S0/S0ix, S3, S4, S5

Refer to Chapter 4, "Power Management" for more information.

### 1.2.3 Memory Controller Power Management

- Disabling Unused System Memory Outputs
- DRAM Power Management and Initialization

- Initialization Role of CKE
- Conditional Self-Refresh
- Dynamic Power Down
- DRAM I/O Power Management
- DDR Electrical Power Gating (EPG)
- Power training

Refer to Section 4.3, "Integrated Memory Controller (IMC) Power Management" for more information.

## 1.3 Thermal Management Support

- Digital Thermal Sensor
- Intel Adaptive Thermal Monitor
- THERMTRIP# and PROCHOT# support
- On-Demand Mode
- Memory Open and Closed Loop Throttling
- Memory Thermal Throttling
- External Thermal Sensor (TS-on-DIMM and TS-on-Board)
- Render Thermal Throttling
- Fan speed control with DTS
- Intel Turbo Boost Technology 2.0 Power Control

Refer to Chapter 5, "Thermal Management" for more information.

## 1.4 Package Support

The processor is available in the following packages:

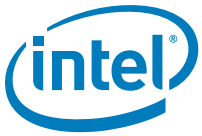
- A 58.5 mm x 51 mm LGA package (LGA2066) for X-Processor Line

## 1.5 Processor Testability

An XDP on-board connector is warmly recommended to enable full debug capabilities. For the processor SKUs, a merged XDP connector is highly recommended to enable lower C-state debug.

***Note:*** When separate XDP connectors will be used at C8 state, the processor will need to be waked up using the PCH.

The processor includes boundary-scan for board and system level testability.



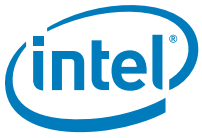
# 1.6 Terminology

## Table 1-2. Terminology (Sheet 1 of 2)

| Term | Description |
|---|---|
| 4K | Ultra High Definition (UHD) |
| AES | Advanced Encryption Standard |
| AGC | Adaptive Gain Control |
| BLT | Block Level Transfer |
| BPP | Bits per pixel |
| CDR | Clock and Data Recovery |
| CTLE | Continuous Time Linear Equalizer |
| DDR4/DDR4-RS | Fourth-Generation Double Data Rate SDRAM Memory Technology<br>RS - Reduced Standby Power |
| DFE | decision feedback equalizer |
| DMA | Direct Memory Access |
| DMI | Direct Media Interface |
| DP | DisplayPort* |
| DTS | Digital Thermal Sensor |
| eDP* | embedded DisplayPort* |
| GSA | Graphics in System Agent |
| HDCP | High-bandwidth Digital Content Protection |
| HDMI* | High Definition Multimedia Interface |
| IMC | Integrated Memory Controller |
| Intel® 64 Technology | 64-bit memory extensions to the IA-32 architecture |
| Intel® PTT | Intel Platform Trust Technology |
| Intel® TSX-NI | Intel Transactional Synchronization Extensions |
| Intel® VT | Intel Virtualization Technology. Processor virtualization, when used in conjunction with Virtual Machine Monitor software, enables multiple, robust independent software environments inside a single platform. |
| Intel® VT-d | Intel Virtualization Technology (Intel VT) for Directed I/O. Intel VT-d is a hardware assist, under system software (Virtual Machine Manager or OS) control, for enabling I/O device virtualization. Intel VT-d also brings robust security by providing protection from errant DMAs by using DMA remapping, a key feature of Intel VT-d. |
| IOV | I/O Virtualization DR3 |
| ISP | Image Signal Processor |
| LFM | Low Frequency Mode. corresponding to the Enhanced Intel SpeedStep® Technology's lowest voltage/frequency pair. It can be read at MSR CEh [47:40]. |
| LLC | Last Level Cache |
| LPM | Low-Power Mode.The LPM Frequency is less than or equal to the LFM Frequency. The LPM TDP is lower than the LFM TDP as the LPM configuration limits the processor to single thread operation |
| LPSP | Low-Power Single Pipe |
| LSF | Lowest Supported Frequency.This frequency is the lowest frequency where manufacturing confirms logical functionality under the set of operating conditions. |
| MCP | Multi Chip Package - includes the processor and the PCH. |
| MFM | Minimum Frequency Mode. MFM is the minimum ratio supported by the processor and can be read from MSR CEh [55:48]. |

## Table 1-2. Terminology (Sheet 2 of 2)

| Term | Description |
|---|---|
| MLC | Mid-Level Cache |
| NCTF | Non-Critical to Function. NCTF locations are typically redundant ground or non-critical reserved balls/lands, so the loss of the solder joint continuity at end of life conditions will not affect the overall product functionality. |
| PAG | Platform Power Architecture Guide (formerly PDDG) |
| PCH | Platform Controller Hub. The chipset with centralized platform capabilities including the main I/O interfaces along with display connectivity, audio features, power management, manageability, security, and storage features. The PCH may also be referred as "chipset". |
| PECI | Platform Environment Control Interface |
| PEG | PCI Express Graphics |
| PL1, PL2, PL3 | Power Limit 1, Power Limit 2, Power Limit 3 |
| Processor | The 64-bit multi-core component (package) |
| Processor Core | The term "processor core" refers to Si die itself, which can contain multiple execution cores. Each execution core has an instruction cache, data cache, and 256-KB L2 cache. All execution cores share the LLC. |
| Processor Graphics | Intel Processor Graphics |
| PSR | Panel Self-Refresh |
| Rank | A unit of DRAM corresponding to four to eight devices in parallel. |
| SCI | System Control Interrupt. SCI is used in the ACPI protocol. |
| SHA | Secure Hash Algorithm |
| SSC | Spread Spectrum Clock |
| Storage Conditions | A non-operational state. The processor may be installed in a platform, in a tray, or loose. Processors may be sealed in packaging or exposed to free air. Under these conditions, processor landings should not be connected to any supply voltages, have any I/Os biased, or receive any clocks. Upon exposure to "free air" (that is, unsealed packaging or a device removed from packaging material), the processor should be handled in accordance with moisture sensitivity labeling (MSL) as indicated on the packaging material. |
| STR | Suspend to RAM |
| TAC | Thermal Averaging Constant |
| TCC | Thermal Control Circuit |
| TDP | Thermal Design Power |
| TOB | Tolerance Budget |
| TTV TDP | Thermal Test Vehicle TDP |
| $V_{CC}$ | Processor core power supply |
| $V_{CCIO}$ | I/O Power Supply |
| $V_{CCSA}$ | System Agent Power Supply |
| $V_{CCST}$ | Vcc Sustain Power Supply |
| $V_{DDQ}$ | DDR Power Supply |
| VLD | Variable Length Decoding |
| VPID | Virtual Processor ID |
| $V_{SS}$ | Processor Ground |



## 1.7 Related Documents

### Table 1-3. Related Documents

| Document | Document Number / Location |
|---|---|
| 7th Generation Intel® Processor Families for S Platforms and Intel® Core™ X-Series Processor Family, Volume 2 of 2 | 335196 |
| 7th Generation Intel® Processor Families Specification Update | 334663 |
| Intel® 200 Series (including X299) Chipset Family Platform Controller Hub (PCH) Datasheet Volume 1 of 2 | 335192 |
| Intel® 200 Series (including X299) Chipset Family Platform Controller Hub (PCH) Datasheet Volume 2 of 2 | 335193 |
| Intel® 200 Series (including X299) Chipset Family Platform Controller Hub (PCH) Specification Update | 335194 |
| Advanced Configuration and Power Interface 3.0 | http://www.acpi.info/ |
| DDR4 Specification | http://www.jedec.org |
| High Definition Multimedia Interface specification revision 1.4 | http://www.hdmi.org/manufacturer/specification.aspx |
| PCI Express* Base Specification Revision 3.0 | http://www.pcisig.com/specifications |
| Intel® 64 and IA-32 Architectures Software Developer's Manuals | http://www.intel.com/products/processor/manuals/index.htm |

§ §

# 2 Interfaces

## 2.1 System Memory Interface

- Two channels of  DDR4 memory with a maximum of two DIMMs per channel. DDR technologies, number of DIMMs per channel, number of ranks per channel are SKU dependent.
- UDIMM support
- Single-channel and dual-channel memory organization modes
- Data burst length of eight for all memory organization modes
- DDR4 I/O Voltage of 1.2V
- 64-bit wide channels
- Non-ECC UDIMM DDR4 support
- Theoretical maximum memory bandwidth of:
  - 37.5 GB/s in dual-channel mode assuming 2400 MT/s
  - 41.6 GB/s in dual-channel mode assuming 2666 MT/s

### 2.1.1 System Memory Technology Supported

The Integrated Memory Controller (IMC) supports DDR4 protocols with two independent, 64-bit wide channels.

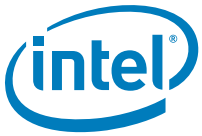
**Table 2-1. Processor DRAM Support Matrix**

| Processor Line | DPC[1] | DDR3L/-RS [MT/s] | DDR4 [MT/s] | LPDDR3 [MT/s] |
|---|---|---|---|---|
| **X-Processor Line** | 2 | N/A | 2400/2666[2] | N/A |

***Notes:***
1. DPC = DIMM Per Channel
2. DDR4 2666 [MT/S] support for 1 DPC only, 2400[MT/S] supported for 2 DPC.

- DDR4 Data Transfer Rates:
  - 2400 MT/s (PC4-2400)
  - 2666 MT/s (PC4-2666)

  There is no support for memory modules with different technologies or capacities on opposite sides of the same memory module. If one side of a memory module is populated, the other side is either identical or empty.



#### 2.1.1.1 DDR4 Supported Memory Modules and Devices

**Table 2-2. Supported DDR4 Non-ECC UDIMM Module Configurations**

| Raw Card Version | DIMM Capacity | DRAM Device Technology | DRAM Organization | # of DRAM Devices | # of Ranks | # of Row/Col Address Bits | # of Banks Inside DRAM | Page Size |
|---|---|---|---|---|---|---|---|---|
| A | 4GB | 4Gb | 512M x 8 | 8 | 1 | 15/10 | 16 | 8K |
| A | 8GB | 8Gb | 1024M x 8 | 8 | 1 | 16/10 | 16 | 8K |
| B | 8GB | 4Gb | 512M x 8 | 16 | 2 | 15/10 | 16 | 8K |
| B | 16GB | 8Gb | 1024M x 8 | 16 | 2 | 16/10 | 16 | 8K |

### 2.1.2 System Memory Timing Support

The IMC supports the following DDR Speed Bin, CAS Write Latency (CWL), and command signal mode timings on the main memory interface:

- tCL = CAS Latency
- tRCD = Activate Command to READ or WRITE Command delay
- tRP = PRECHARGE Command Period
- CWL = CAS Write Latency
- Command Signal modes:
  - 1N indicates a new DDR4 command may be issued every clock
  - 2N indicates a new DDR4 command may be issued every 2 clocks

**Table 2-3. DRAM System Memory Timing Support**

| DRAM Device | Transfer Rate (MT/s) | tCL (tCK) | tRCD (tCK) | tRP (tCK) | CWL (tCK) | CMD Mode |
|---|---|---|---|---|---|---|
| DDR4 | 2400 | 17 | 17 | 17 | 12/16/16 | 1N/2N |

### 2.1.3 System Memory Organization Modes

The IMC supports two memory organization modes, single-channel and dual-channel. Depending upon how the DDR Schema and DIMM Modules are populated in each memory channel, a number of different configurations can exist.

**Single-Channel Mode**

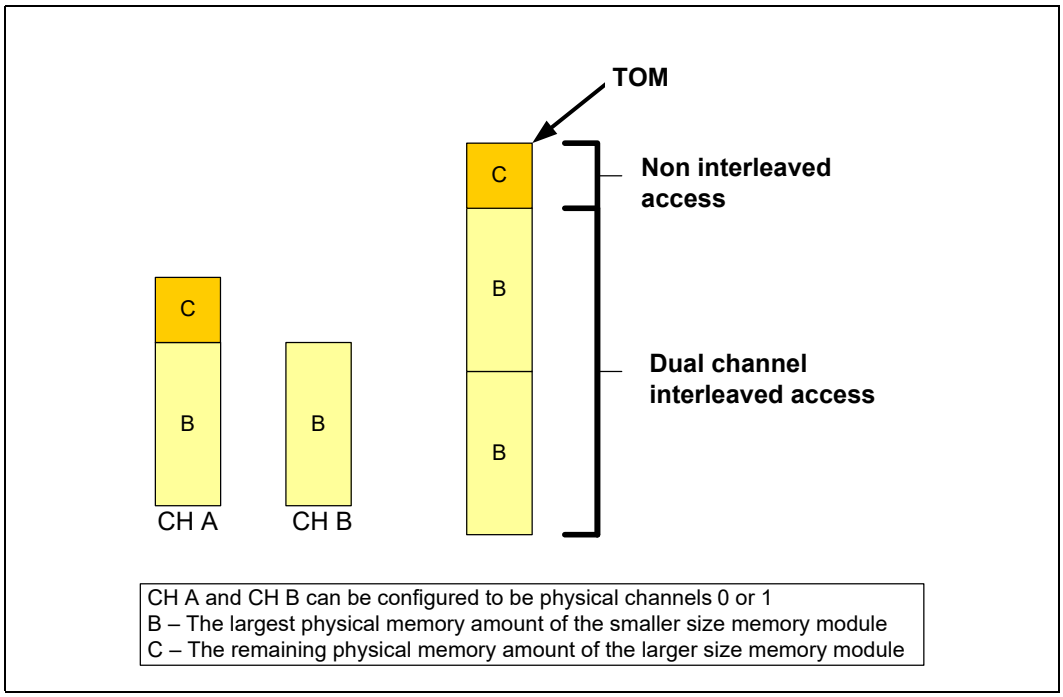
In this mode, all memory cycles are directed to a single channel. Single-Channel mode is used when either the Channel A or Channel B DIMM connectors are populated in any order, but not both.

**Dual-Channel Mode – Intel® Flex Memory Technology Mode**

The IMC supports Intel Flex Memory Technology Mode. Memory is divided into a symmetric and asymmetric zone. The symmetric zone starts at the lowest address in each channel and is contiguous until the asymmetric zone begins or until the top address of the channel with the smaller capacity is reached. In this mode, the system runs with one zone of dual-channel mode and one zone of single-channel mode, simultaneously, across the whole memory array.

***Note:*** Channels A and B can be mapped for physical channel 0 and 1 respectively or vice versa. However, channel A size should be greater or equal to channel B size.

**Figure 2-1. Intel® Flex Memory Technology Operations**




**Dual-Channel Symmetric Mode (Interleaved Mode)**

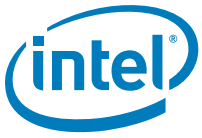
Dual-Channel Symmetric mode, also known as interleaved mode, provides maximum performance on real world applications. Addresses are ping-ponged between the channels after each cache line (64-byte boundary). If there are two requests, and the second request is to an address on the opposite channel from the first, that request can be sent before data from the first request has returned. If two consecutive cache lines are requested, both may be retrieved simultaneously, since they are ensured to be on opposite channels. Use Dual-Channel Symmetric mode when both Channel A and Channel B DIMM connectors are populated in any order, with the total amount of memory in each channel being the same.

When both channels are populated with the same memory capacity and the boundary between the dual channel zone and the single channel zone is the top of memory, IMC operates completely in Dual-Channel Symmetric mode.

***Note:*** The DRAM device technology and width may vary from one channel to the other.

## 2.1.4 System Memory Frequency

In all modes, the frequency of system memory is the lowest frequency of all memory modules placed in the system, as determined through the SPD registers on the memory modules. The system memory controller supports up to two DIMM connectors per channel. If DIMMs with different latency are populated across the channels, the BIOS will use the slower of the two latencies for both channels. For Dual-Channel modes both channels should have a DIMM connector populated. For Single-Channel mode, only a single channel can have a DIMM connector populated.



### 2.1.5 Technology Enhancements of Intel® Fast Memory Access (Intel® FMA)

The following sections describe the Just-in-Time Scheduling, Command Overlap, and Out-of-Order Scheduling Intel FMA technology enhancements.

#### Just-in-Time Command Scheduling

The memory controller has an advanced command scheduler where all pending requests are examined simultaneously to determine the most efficient request to be issued next. The most efficient request is picked from all pending requests and issued to system memory Just-in-Time to make optimal use of Command Overlapping. Thus, instead of having all memory access requests go individually through an arbitration mechanism forcing requests to be executed one at a time, they can be started without interfering with the current request allowing for concurrent issuing of requests. This allows for optimized bandwidth and reduced latency while maintaining appropriate command spacing to meet system memory protocol.

#### Command Overlap

Command Overlap allows the insertion of the DRAM commands between the Activate, Pre-charge, and Read/Write commands normally used, as long as the inserted commands do not affect the currently executing command. Multiple commands can be issued in an overlapping manner, increasing the efficiency of system memory protocol.

#### Out-of-Order Scheduling

While leveraging the Just-in-Time Scheduling and Command Overlap enhancements, the IMC continuously monitors pending requests to system memory for the best use of bandwidth and reduction of latency. If there are multiple requests to the same open page, these requests would be launched in a back to back manner to make optimum use of the open memory page. This ability to reorder requests on the fly allows the IMC to further reduce latency and increase bandwidth efficiency.

### 2.1.6 Data Scrambling

The system memory controller incorporates a Data Scrambling feature to minimize the impact of excessive di/dt on the platform system memory VRs due to successive 1s and 0s on the data bus. Past experience has demonstrated that traffic on the data bus is not random and can have energy concentrated at specific spectral harmonics creating high di/dt which is generally limited by data patterns that excite resonance between the package inductance and on die capacitances. As a result, the system memory controller uses a data scrambling feature to create pseudo-random patterns on the system memory data bus to reduce the impact of any excessive di/dt.

### 2.1.7 DDR I/O Interleaving

The processor supports I/O interleaving, which has the ability to swap DDR bytes for routing considerations. BIOS configures the I/O interleaving mode before DDR training.

There are 2 supported modes:

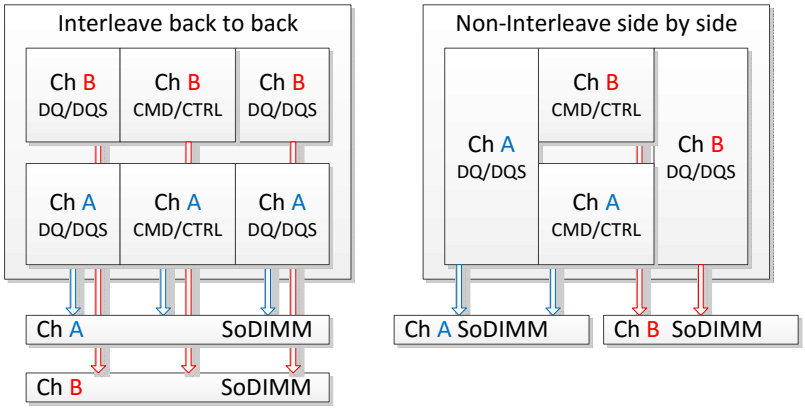
- Interleave (IL)
- Non-Interleave (NIL)

The following table and figure describe the pin mapping between the IL and NIL modes.

**Table 2-4. Interleave (IL) and Non-Interleave (NIL) Modes Pin Mapping**

| IL | | NIL | |
|---|---|---|---|
| **Channel** | **Byte** | **Channel** | **Byte** |
| DDR0 | Byte0 | DDR0 | Byte0 |
| DDR0 | Byte1 | DDR0 | Byte1 |
| DDR0 | Byte2 | DDR0 | Byte4 |
| DDR0 | Byte3 | DDR0 | Byte5 |
| DDR0 | Byte4 | DDR1 | Byte0 |
| DDR0 | Byte5 | DDR1 | Byte1 |
| DDR0 | Byte6 | DDR1 | Byte4 |
| DDR0 | Byte7 | DDR1 | Byte5 |
| DDR1 | Byte0 | DDR0 | Byte2 |
| DDR1 | Byte1 | DDR0 | Byte3 |
| DDR1 | Byte2 | DDR0 | Byte6 |
| DDR1 | Byte3 | DDR0 | Byte7 |
| DDR1 | Byte4 | DDR1 | Byte2 |
| DDR1 | Byte5 | DDR1 | Byte3 |
| DDR1 | Byte6 | DDR1 | Byte6 |
| DDR1 | Byte7 | DDR1 | Byte7 |

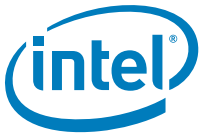
**Figure 2-2. Interleave (IL) and Non-Interleave (NIL) Modes Mapping**




## 2.1.8 Data Swapping

By default, the processor supports on-board data swapping in two manners (for all segments and DRAM technologies):

- byte (DQ+DQS) swapping between bytes in the same channel.
- bit swapping within specific byte.



### 2.1.9 DRAM Clock Generation

Every supported rank has a differential clock pair. There are a total of four clock pairs driven directly by the processor to DRAM.

### 2.1.10 DRAM Reference Voltage Generation

The memory controller has the capability of generating the DDR4 Reference Voltage (VREF) internally for both read and write operations. The generated VREF can be changed in small steps, and an optimum VREF value is determined for both during a cold boot through advanced training procedures in order to provide the best voltage to achieve the best signal margins.

## 2.2 PCI Express* Graphics Interface (PEG)

This section describes the PCI Express* interface capabilities of the processor. See the *PCI Express Base* Specification 3.0* for details on PCI Express*.

### 2.2.1 PCI Express* Support

The processor's PCI Express* interface is a 16-lane (x16) port that can also be configured as multiple ports at narrower widths (see Table 2-5, Table 2-6).

The processor supports the configurations shown in the following table.

### Table 2-5. PCI Express* Bifurcation and Lane Reversal Mapping

| Bifurcation | Link Width | | | CFG Signals | | | Lanes | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0:1:0 | 0:1:1 | 0:1:2 | CFG [6] | CFG [5] | CFG [2] | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1x16 | x16 | N/A | N/A | 1 | 1 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| 1x16 **Reversed** | x16 | N/A | N/A | 1 | 1 | 0 | 15 | 14 | 13 | 12 | 11 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| 2x8 | x8 | x8 | N/A | 1 | 0 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2x8 **Reversed** | x8 | x8 | N/A | 1 | 0 | 0 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |
| 1x8+2x4 | x8 | x4 | x4 | 0 | 0 | 1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 0 | 1 | 2 | 3 | 0 | 1 | 2 | 3 |
| 1x8+2x4 **Reversed** | x8 | x4 | x4 | 0 | 0 | 0 | 3 | 2 | 1 | 0 | 3 | 2 | 1 | 0 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | 0 |

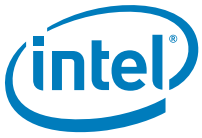
***Notes:***

1. For CFG bus details, refer to Section 6.4.
2. Support is also provided for narrow width and use devices with lower number of lanes (that is, usage on x4 configuration), however further bifurcation is not supported.
3. In case that more than one device is connected, the device with the highest lane count, should always be connected to the lower lanes, as follows:
   - Connect lane 0 of 1st device to lane 0.
   - Connect lane 0 of 2nd device to lane 8.
   - Connect lane 0 of 3rd device to lane 12.

   For example:
   a. When using 1x8 + 2x4, the 8 lane device should use lanes 0:7.
   b. When using 1x4 + 1x2, the 4 lane device should use lanes 0:3, and other 2 lanes device should use lanes 8:9.
   c. When using 1x4 + 1x2 + 1x1, 4 lane device should use lanes 0:3, two lane device should use lanes 8:9, one lane device should use lane 12.
4. for reversal lanes, for example:
   When using 1x8, the 8 lane device should use lanes 8:15, so lane 15 will be connected to lane 0 of the Device.
5. For X-Processor platform, use 1x8+2x4 Bifurcation

The processor supports the following:

- Hierarchical PCI-compliant configuration mechanism for downstream devices
- Traditional PCI style traffic (asynchronous snooped, PCI ordering)
- PCI Express* extended configuration space. The first 256 bytes of configuration space aliases directly to the PCI Compatibility configuration space. The remaining portion of the fixed 4-KB block of memory-mapped space above that (starting at 100h) is known as extended configuration space.
- PCI Express* Enhanced Access Mechanism. Accessing the device configuration space in a flat memory mapped fashion
- Automatic discovery, negotiation, and training of link out of reset.
- Peer segment destination posted write traffic (no peer-to-peer read traffic) in Virtual Channel 0: DMI -> PCI Express* Port 0
- 64-bit downstream address format, but the processor never generates an address above 512 GB (Bits 63:39 will always be zeros)
- 64-bit upstream address format, but the processor responds to upstream read transactions to addresses above 512 GB (addresses where any of Bits 63:39 are nonzero) with an Unsupported Request response. Upstream write transactions to addresses above 512 GB will be dropped.
- Re-issues Configuration cycles that have been previously completed with the Configuration Retry status
- PCI Express* reference clock is 100-MHz differential clock



- Power Management Event (PME) functions
- Dynamic width capability
- Message Signaled Interrupt (MSI and MSI-X) messages
- Lane reversal

The following table summarizes the transfer rates and theoretical bandwidth of PCI Express* link.

**Table 2-6. PCI Express* Maximum Transfer Rates and Theoretical Bandwidth**

| PCI Express* Generation | Encoding | Maximum Transfer Rate [GT/s] | Theoretical Bandwidth [GB/s] | | | | |
|---|---|---|---|---|---|---|---|
| | | | x1 | x2 | x4 | x8 | x16 |
| Gen 1 | 8b/10b | 2.5 | 0.25 | 0.5 | 1.0 | 2.0 | 4.0 |
| Gen 2 | 8b/10b | 5 | 0.5 | 1.0 | 2.0 | 4.0 | 8.0 |
| Gen 3 | 128b/130b | 8 | 1.0 | 2.0 | 3.9 | 7.9 | 15.8 |

***Note:*** The processor has limited support for Hot-Plug. For details, refer to Section 4.4.

## 2.2.2 PCI Express* Architecture

Compatibility with the PCI addressing model is maintained to ensure that all existing applications and drivers operate unchanged.

The PCI Express* configuration uses standard mechanisms as defined in the PCI Plug and-Play specification. The processor PCI Express* ports support Gen 3.
At 8 GT/s, Gen3 operation results in twice as much bandwidth per lane as compared to Gen 2 operation. The 16 lanes port can operate at 2.5 GT/s, 5 GT/s, or 8 GT/s.

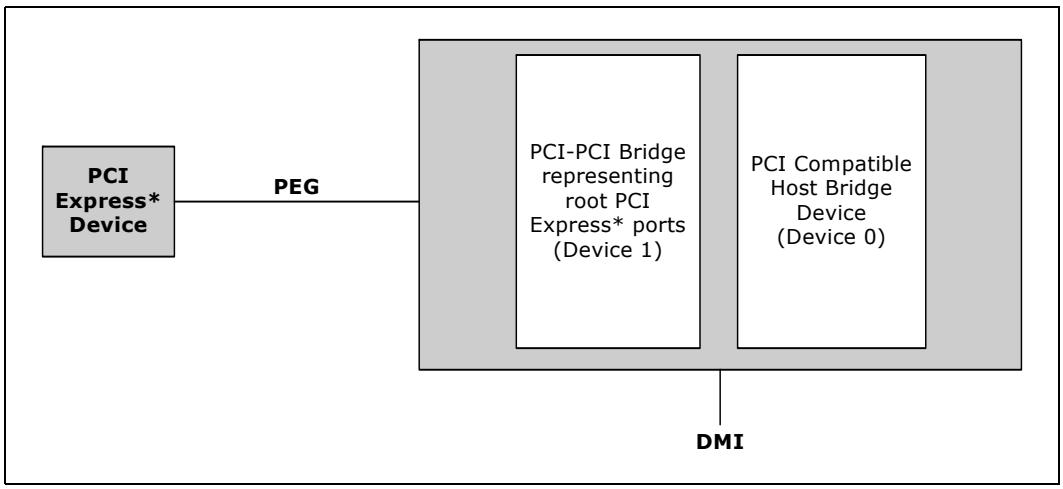
Gen 3 PCI Express* uses a 128b/130b encoding which is about 23% more efficient than the 8b/10b encoding used in Gen 1 and Gen 2.

The PCI Express* architecture is specified in three layers – Transaction Layer, Data Link Layer, and Physical Layer. See the *PCI Express Base Specification 3.0* for details of PCI Express* architecture.

## 2.2.3 PCI Express* Configuration Mechanism

The PCI Express* (external graphics) link is mapped through a PCI-to-PCI bridge structure.

**Figure 2-3. PCI Express* Related Register Structures in the Processor**




PCI Express* extends the configuration space to 4096 bytes per-device/function, as compared to 256 bytes allowed by the conventional PCI specification. PCI Express* configuration space is divided into a PCI-compatible region (that consists of the first 256 bytes of a logical device's configuration space) and an extended PCI Express* region (that consists of the remaining configuration space). The PCI-compatible region can be accessed using either the mechanisms defined in the PCI specification or using the enhanced PCI Express* configuration access mechanism described in the PCI Express* Enhanced Configuration Mechanism section.

The PCI Express* Host Bridge is required to translate the memory-mapped PCI Express* configuration space accesses from the host processor to PCI Express* configuration cycles. To maintain compatibility with PCI configuration addressing mechanisms, it is recommended that system software access the enhanced configuration space using 32-bit operations (32-bit aligned) only. See the PCI Express Base Specification for details of both the PCI-compatible and PCI Express* Enhanced configuration mechanisms and transaction rules.

## 2.2.4 PCI Express* Equalization Methodology

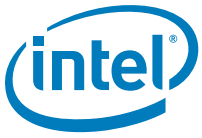
The equalization of link requires equalization for both TX and RX sides for the processor and for the End point device.

Adjusting transmitter and receiver of the lanes is done to improve signal reception quality and for improving link robustness and electrical margin.

The link timing margins and voltage margins are strongly dependent on equalization of the link.

The processor supports the following:

- Full TX Equalization: Three Taps Linear Equalization (Pre, Current and Post cursors), with FS/LF (Full Swing /Low Frequency) 24/8 values respectively.



- Full RX Equalization and acquisition for: AGC (Adaptive Gain Control), CDR (Clock and Data Recovery), adaptive DFE (decision feedback equalizer) and adaptive CTLE peaking (continuous time linear equalizer).
- Full adaptive phase 3 EQ compliant with PCI Express* Gen 3 specification

See the *PCI Express* Base Specification 3.0* for details on PCI Express* equalization.

## 2.3 Direct Media Interface (DMI)

Direct Media Interface (DMI) connects the processor and the PCH.

Main characteristics:

- 4 lanes Gen 3 DMI support
- 8 GT/s point-to-point DMI interface to PCH
- DC coupling - no capacitors between the processor and the PCH
- PCH end-to-end lane reversal across the link
- Half-Swing support (low-power/low-voltage)

***Note:*** Only DMI x4 configuration is supported.

***Note:*** Polarity Inversion and Lane Reversal on DMI Link are not allowed.

### 2.3.1 DMI Error Flow

DMI can only generate SERR in response to errors; never SCI, SMI, MSI, PCI INT, or GPE. Any DMI related SERR activity is associated with Device 0.

### 2.3.2 DMI Link Down

The DMI link going down is a fatal, unrecoverable error. If the DMI data link goes to data link down, after the link was up, then the DMI link hangs the system by not allowing the link to retrain to prevent data corruption. This link behavior is controlled by the PCH.

Downstream transactions that had been successfully transmitted across the link prior to the link going down may be processed as normal. No completions from downstream, non-posted transactions are returned upstream over the DMI link after a link down event.

## 2.4 Platform Environmental Control Interface (PECI)

PECI is an Intel proprietary interface that provides a communication channel between Intel processors and external components like Super IO (SIO) and Embedded Controllers (EC) to provide processor temperature, Turbo, and memory throttling control mechanisms and many other services. PECI is used for platform thermal management and real time control and configuration of processor features and performance.

### 2.4.1 PECI Bus Architecture

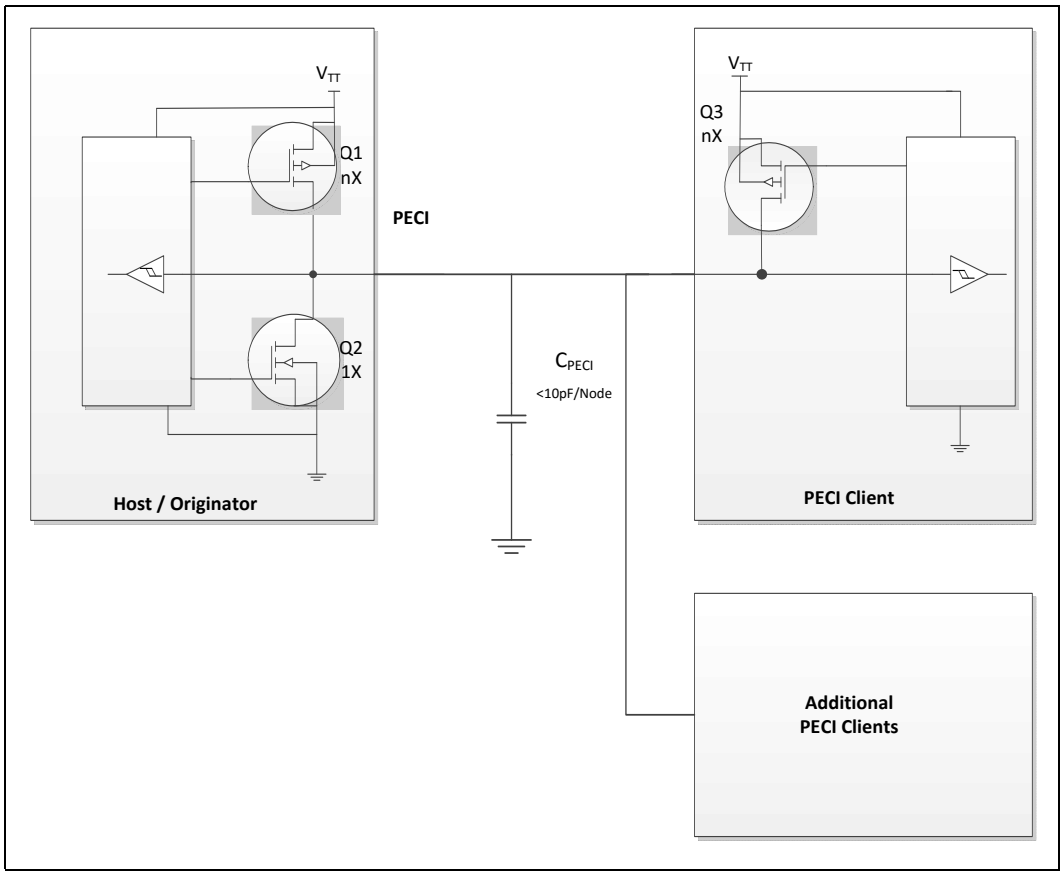
The PECI architecture is based on a wired OR bus that the clients (as processor PECI) can pull up (with strong drive).

The idle state on the bus is near zero.

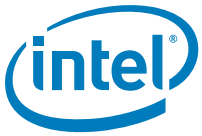
The following figures demonstrates PECI design and connectivity:

- PECI Host-Clients Connection: While the host/originator can be third party PECI host and one of the PECI client is a processor PECI device.
- PECI EC Connection.

**Figure 2-4. Example for PECI Host-Clients Connection**




§ §



# 3 Technologies

This chapter provides a high-level description of Intel technologies implemented in the processor.

The implementation of the features may vary between the processor SKUs.

Details on the different technologies of Intel processors and other relevant external notes are located at the Intel technology web site: http://www.intel.com/technology/

## 3.1 Intel® Virtualization Technology (Intel® VT)

Intel® Virtualization Technology (Intel® VT) makes a single system appear as multiple independent systems to software. This allows multiple, independent operating systems to run simultaneously on a single system. Intel VT comprises technology components to support virtualization of platforms based on Intel architecture microprocessors and chipsets.

Intel Virtualization Technology (Intel VT) for IA-32, Intel 64 and Intel Architecture (Intel VT-x) added hardware support in the processor to improve the virtualization performance and robustness. Intel Virtualization Technology for Directed I/O (Intel VT-d) extends Intel VT-x by adding hardware assisted support to improve I/O device virtualization performance.

Intel VT-x specifications and functional descriptions are included in the *Intel 64 and IA-32 Architectures Software Developer's Manual, Volume 3*. Available at:

http://www.intel.com/products/processor/manuals/index.htm

The Intel VT-d specification and other VT documents can be referenced at:

http://www.intel.com/technology/virtualization/index.htm

https://sharedspaces.intel.com/sites/PCDC/SitePages/Ingredients/ingredient.aspx?ing=VT

### 3.1.1 Intel® Virtualization Technology (Intel® VT) for IA-32, Intel® 64 and Intel® Architecture (Intel® VT-X)

**Intel® VT-x Objectives**

Intel VT-x provides hardware acceleration for virtualization of IA platforms. Virtual Machine Monitor (VMM) can use Intel VT-x features to provide an improved reliable virtualized platform. By using Intel VT-x, a VMM is:

- **Robust:** VMMs no longer need to use para-virtualization or binary translation. This means that VMMs will be able to run off-the-shelf operating systems and applications without any special steps.
- **Enhanced:** Intel VT enables VMMs to run 64-bit guest operating systems on IA x86 processors.
- **More reliable:** Due to the hardware support, VMMs can now be smaller, less complex, and more efficient. This improves reliability and availability and reduces the potential for software conflicts.

- **More secure:** The use of hardware transitions in the VMM strengthens the isolation of VMs and further prevents corruption of one VM from affecting others on the same system.

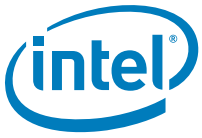
### Intel® VT-x Key Features

The processor supports the following added new Intel VT-x features:

- Extended Page Table (EPT) Accessed and Dirty Bits
  - EPT A/D bits enabled VMMs to efficiently implement memory management and page classification algorithms to optimize VM memory operations, such as de-fragmentation, paging, live migration, and check-pointing. Without hardware support for EPT A/D bits, VMMs may need to emulate A/D bits by marking EPT paging-structures as not-present or read-only, and incur the overhead of EPT page-fault VM exits and associated software processing.
- EPTP (EPT pointer) switching
  - EPTP switching is a specific VM function. EPTP switching allows guest software (in VMX non-root operation, supported by EPT) to request a different EPT paging-structure hierarchy. This is a feature by which software in VMX non-root operation can request a change of EPTP without a VM exit. Software will be able to choose among a set of potential EPTP values determined in advance by software in VMX root operation.
- Pause loop exiting
  - Support VMM schedulers seeking to determine when a virtual processor of a multiprocessor virtual machine is not performing useful work. This situation may occur when not all virtual processors of the virtual machine are currently scheduled and when the virtual processor in question is in a loop involving the PAUSE instruction. The new feature allows detection of such loops and is thus called PAUSE-loop exiting.

The processor IA core supports the following Intel VT-x features:

- **Mode based (XU/XS) EPT execute control - New Feature for this processor**
  - A new mode of EPT operation which enables different controls for executability of GPA based on Guest specified mode (User/Supervisor) of linear address translating to the GPA. When the mode is enabled, the executability of a GPA is defined by two bits in EPT entry. One bit for accesses to user pages and other one for accesses to supervisor pages.
  - The new mode requires changes in VMCS, and EPT entries. VMCS includes a bit "mode based EPT execute control" which is used to enable/disable the mode. An additional bit in EPT entry is defined as "supervisor-execute access"; the original execute control bit is considered as "user-execute access". If the "mode based EPT execute control" is disabled the additional bit is ignored and the system works with one bit execute control for both user pages and supervisor pages.
  - Behavioral changes - Behavioral changes are across three areas:
    - **Access to GPA-** If the "mode-based EPT execute control" VM-execution control is 1, treatment of guest-physical accesses by instruction fetches depends on the linear address from which an instruction is being fetched
      1. If the translation of the linear address specifies user mode (the U/S bit was set in every paging structure entry used to translate the linear address), the resulting guest-physical address is executable under EPT only if the XU bit (at position 2) is set in every EPT paging-structure entry used to translate the guest-physical address.
      2. If the translation of the linear address specifies supervisor mode (the U/S bit was clear in at least one of the paging-structure entries used



to translate the linear address), the resulting guest-physical address is executable under EPT only if the XS bit is set in every EPT paging-structure entry used to translate the guest-physical address

—The XU and XS bits are used only when translating linear addresses for guest code fetches. They do not apply to guest page walks, data accesses, or A/D-bit updates

- **VMEntry -** If the "activate secondary controls" and "mode-based EPT execute control" VM-execution controls are both 1, VM entries ensure that the "enable EPT" VM-execution control is 1. VM entry fails if this check fails. When such a failure occurs, control is passed to the next instruction,
- **VMExit -** The exit qualification due to EPT violation reports clearly whether the violation was due to User mode access or supervisor mode access.

— Capability Querying: IA32_VMX_PROCBASED_CTLS2 has bit to indicate the capability, RDMSR can be used to read and query whether the processor supports the capability or not.

- Extended Page Tables (EPT)
  - — EPT is hardware assisted page table virtualization
  - — It eliminates VM exits from guest OS to the VMM for shadow page-table maintenance
- Virtual Processor IDs (VPID)
  - — Ability to assign a VM ID to tag processor IA core hardware structures (such as TLBs)
  - — This avoids flushes on VM transitions to give a lower-cost VM transition time and an overall reduction in virtualization overhead.
- Guest Preemption Timer
  - — Mechanism for a VMM to preempt the execution of a guest OS after an amount of time specified by the VMM. The VMM sets a timer value before entering a guest
  - — The feature aids VMM developers in flexibility and Quality of Service (QoS) guarantees
- Descriptor-Table Exiting
  - — Descriptor-table exiting allows a VMM to protect a guest OS from internal (malicious software based) attack by preventing relocation of key system data structures like IDT (interrupt descriptor table), GDT (global descriptor table), LDT (local descriptor table), and TSS (task segment selector).
  - — A VMM using this feature can intercept (by a VM exit) attempts to relocate these data structures and prevent them from being tampered by malicious software.

## 3.1.2 Intel® Virtualization Technology (Intel® VT) for Directed I/O (Intel® VT-d)

**Intel® VT-d Objectives**

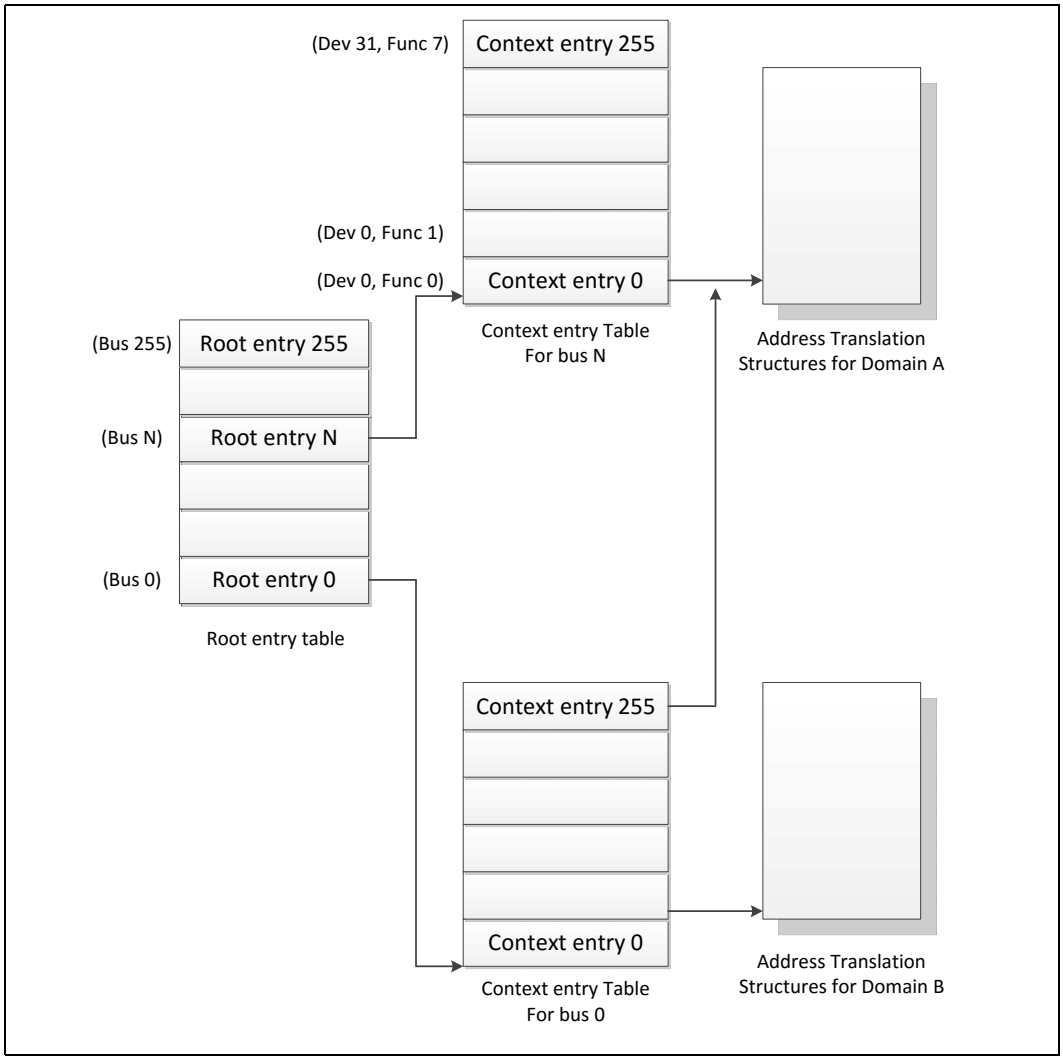
The key Intel VT-d objectives are domain-based isolation and hardware-based virtualization. A domain can be abstractly defined as an isolated environment in a platform to which a subset of host physical memory is allocated. Intel VT-d provides accelerated I/O performance for a virtualized platform and provides software with the following capabilities:

- I/O device assignment and security: for flexibly assigning I/O devices to VMs and extending the protection and isolation properties of VMs for I/O operations.

- DMA remapping: for supporting independent address translations for Direct Memory Accesses (DMA) from devices.
- Interrupt remapping: for supporting isolation and routing of interrupts from devices and external interrupt controllers to appropriate VMs.
- Reliability: for recording and reporting to system software DMA and interrupt errors that may otherwise corrupt memory or impact VM isolation.

Intel VT-d accomplishes address translation by associating transaction from a given I/O device to a translation table associated with the Guest to which the device is assigned. It does this by means of the data structure in the following illustration. This table creates an association between the device's PCI Express* Bus/Device/Function (B/D/F) number and the base address of a translation table. This data structure is populated by a VMM to map devices to translation tables in accordance with the device assignment restrictions above, and to include a multi-level translation table (VT-d Table) that contains Guest specific address translations.

**Figure 3-1. Device to Domain Mapping Structures**



Intel VT-d functionality, often referred to as an Intel VT-d Engine, has typically been implemented at or near a PCI Express* host bridge component of a computer system. This might be in a chipset component or in the PCI Express functionality of a processor with integrated I/O. When one such VT-d engine receives a PCI Express transaction from a PCI Express bus, it uses the B/D/F number associated with the transaction to search for an Intel VT-d translation table. In doing so, it uses the B/D/F number to traverse the data structure shown in the above figure. If it finds a valid Intel VT-d table in this data structure, it uses that table to translate the address provided on the PCI Express bus. If it does not find a valid translation table for a given translation, this results in an Intel VT-d fault. If Intel VT-d translation is required, the Intel VT-d engine performs an N-level table walk.

For more information, refer to *Intel Virtualization Technology for Directed I/O Architecture Specification* http://www.intel.com/content/dam/www/public/us/en/documents/product-specifications/vt-directed-io-spec.pdf

### Intel® VT-d Key Features

The processor supports the following Intel VT-d features:

- Memory controller and processor graphics comply with the Intel VT-d 2.1 Specification.
- Two Intel VT-d DMA remap engines.
  - — iGFX DMA remap engine
  - — Default DMA remap engine (covers all devices except iGFX)
- Support for root entry, context entry, and default context
- 39-bit guest physical address and host physical address widths
- Support for 4K page sizes only
- Support for register-based fault recording only (for single entry only) and support for MSI interrupts for faults
- Support for both leaf and non-leaf caching
- Support for boot protection of default page table
- Support for non-caching of invalid page table entries
- Support for hardware based flushing of translated but pending writes and pending reads, on IOTLB invalidation
- Support for Global, Domain specific and Page specific IOTLB invalidation
- MSI cycles (MemWr to address FEEx_xxxxh) not translated
  - — Translation faults result in cycle forwarding to VBIOS region (byte enables masked for writes). Returned data may be bogus for internal agents, PEG/DMI interfaces return unsupported request status
- Interrupt Remapping is supported
- Queued invalidation is supported
- Intel VT-d translation bypass address range is supported (Pass Through)

The processor supports the following added new Intel VT-d features:

- 4-level Intel VT-d Page walk – both default Intel VT-d engine as well as the IGD VT-d engine are upgraded to support 4-level Intel VT-d tables (adjusted guest address width of 48 bits)

- Intel VT-d superpage – support of Intel VT-d superpage (2 MB, 1 GB) for default Intel VT-d engine (that covers all devices except IGD)
  IGD Intel VT-d engine does not support superpage and BIOS should disable superpage in default Intel VT-d engine when iGfx is enabled.

***Note:*** Intel VT-d Technology may not be available on all SKUs.

# 3.2 Security Technologies

## 3.2.1 Intel® Advanced Encryption Standard New Instructions (Intel® AES-NI)

The processor supports Intel Advanced Encryption Standard New Instructions (Intel AES-NI) that are a set of Single Instruction Multiple Data (SIMD) instructions that enable fast and secure data encryption and decryption based on the Advanced Encryption Standard (AES). Intel AES-NI are valuable for a wide range of cryptographic applications, such as applications that perform bulk encryption/decryption, authentication, random number generation, and authenticated encryption. AES is broadly accepted as the standard for both government and industry applications, and is widely deployed in various protocols.

Intel AES-NI consists of six Intel SSE instructions. Four instructions, AESENC, AESENCLAST, AESDEC, and AESDELAST facilitate high performance AES encryption and decryption. The other two, AESIMC and AESKEYGENASSIST, support the AES key expansion procedure. Together, these instructions provide full hardware for supporting AES; offering security, high performance, and a great deal of flexibility.

***Note:*** Intel AES-NI Technology may not be available on all SKUs.

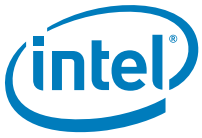
## 3.2.2 PCLMULQDQ (Perform Carry-Less Multiplication Quad word) Instruction

The processor supports the carry-less multiplication instruction, PCLMULQDQ. PCLMULQDQ is a Single Instruction Multiple Data (SIMD) instruction that computes the 128-bit carry-less multiplication of two 64-bit operands without generating and propagating carries. Carry-less multiplication is an essential processing component of several cryptographic systems and standards. Hence, accelerating carry-less multiplication can significantly contribute to achieving high speed secure computing and communication.

## 3.2.3 Intel® Secure Key

The processor supports Intel Secure Key (formerly known as Digital Random Number Generator (DRNG)), a software visible random number generation mechanism supported by a high quality entropy source. This capability is available to programmers through the RDRAND instruction. The resultant random number generation capability is designed to comply with existing industry standards in this regard (ANSI X9.82 and NIST SP 800-90).

Some possible usages of the RDRAND instruction include cryptographic key generation as used in a variety of applications, including communication, digital signatures, secure storage, and so on.



### 3.2.4 Execute Disable Bit

The Execute Disable Bit allows memory to be marked as non executable when combined with a supporting operating system. If code attempts to run in non-executable memory, the processor raises an error to the operating system. This feature can prevent some classes of viruses or worms that exploit buffer overrun vulnerabilities and can, thus, help improve the overall security of the system.

See the *Intel 64 and IA-32 Architectures Software Developer's Manuals* for more detailed information.

### 3.2.5 Intel® Supervisor Mode Execution Protection (SMEP)

Intel® Supervisor Mode Execution Protection (SMEP) is a mechanism that provides the next level of system protection by blocking malicious software attacks from user mode code when the system is running in the highest privilege level. This technology helps to protect from virus attacks and unwanted code from harming the system. For more information, refer to *Intel® 64 and IA-32 Architectures Software Developer's Manual, Volume 3A* at: http://www.intel.com/Assets/PDF/manual/253668.pdf

### 3.2.6 Intel® Supervisor Mode Access Protection (SMAP)

Intel® Supervisor Mode Access Protection (SMAP) is a mechanism that provides next level of system protection by blocking a malicious user from tricking the operating system into branching off user data. This technology shuts down very popular attack vectors against operating systems.

For more information, refer to the *Intel ® 64 and IA-32 Architectures Software Developer's Manual, Volume 3A*: http://www.intel.com/Assets/PDF/manual/253668.pdf

### 3.2.7 Intel® Memory Protection Extensions (Intel® MPX)

Intel® MPX provides hardware accelerated mechanism for memory testing (heap and stack) buffer boundaries in order to identify buffer overflow attacks.

An Intel MPX enabled compiler inserts new instructions that tests memory boundaries prior to a buffer access. Other Intel MPX commands are used to modify a database of memory regions used by the boundary checker instructions.

The Intel MPX ISA is designed for backward compatibility and will be treated as no-operation instructions (NOPs) on older processors.

Intel MPX can be used for:

- Efficient runtime memory boundary checks for security-sensitive portions of the application.
- As part of a memory checker tool for finding difficult memory access errors. Intel MPX is significantly of magnitude faster than software implementations.

Intel MPX emulation (without hardware acceleration) is available with the Intel C++ Compiler 13.0 or newer.

For more information, refer to the Intel MPX documentation.

### 3.2.8 Intel® Virtualization Technology (Intel® VT) for Directed I/O (Intel® VT-d)

Refer to Section 3.1.2 Intel VT-d for detail.

## 3.3 Power and Performance Technologies

### 3.3.1 Intel® Hyper-Threading Technology (Intel® HT Technology)

The processor supports Intel® Hyper-Threading Technology (Intel® HT Technology) that allows an execution processor IA core to function as two logical processors. While some execution resources such as caches, execution units, and buses are shared, each logical processor has its own architectural state with its own set of general-purpose registers and control registers. This feature should be enabled using the BIOS and requires operating system support.

***Note:*** Intel HT Technology may not be available on all SKUs.

### 3.3.2 Intel® Turbo Boost Technology 2.0

The Intel® Turbo Boost Technology 2.0 allows the processor IA core to opportunistically and automatically run faster than the processor IA core base frequency if it is operating below power, temperature, and current limits. The Intel Turbo Boost Technology 2.0 feature is designed to increase performance of both multi-threaded and single-threaded workloads.

Compared with previous generation products, Intel Turbo Boost Technology 2.0 will increase the ratio of application power towards TDP and also allows to increase power above TDP as high as PL2 for short periods of time. Thus, thermal solutions and platform cooling that are designed to less than thermal design guidance might experience thermal and performance issues since more applications will tend to run at the maximum power limit for significant periods of time.

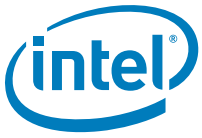
***Note:*** Intel Turbo Boost Technology 2.0 may not be available on all SKUs.

#### 3.3.2.1 Intel® Turbo Boost Technology 2.0 Frequency

To determine the highest performance frequency amongst active processor IA cores, the processor takes the following into consideration:

- The number of processor IA cores operating in the C0 state.
- The estimated processor IA core current consumption and $I_{CCMax}$ register settings.
- The estimated package prior and present power consumption and turbo power limits.
- The package temperature.
- Sustained turbo residencies at high voltages and temperature.

Any of these factors can affect the maximum frequency for a given workload. If the power, current, Voltage or thermal limit is reached, the processor will automatically reduce the frequency to stay within the PL1 value. Turbo processor frequencies are only



active if the operating system is requesting the P0 state. If turbo frequencies are limited the cause is logged in IA_PERF_LIMIT_REASONS register. For more information on P-states and C-states, see Power Management.

### 3.3.3 Intel® Advanced Vector Extensions 2 (Intel® AVX2)

Intel® Advanced Vector Extensions 2.0 (Intel® AVX2) is the latest expansion of the Intel instruction set. Intel AVX2 extends the Intel Advanced Vector Extensions (Intel AVX) with 256-bit integer instructions, floating-point fused multiply add (FMA) instructions, and gather operations. The 256-bit integer vectors benefit math, codec, image, and digital signal processing software. FMA improves performance in face detection, professional imaging, and high performance computing. Gather operations increase vectorization opportunities for many applications. In addition to the vector extensions, this generation of Intel processors adds new bit manipulation instructions useful in compression, encryption, and general purpose software.
For more information on Intel AVX, see http://www.intel.com/software/avx

Intel Advanced Vector Extensions (Intel AVX) are designed to achieve higher throughput to certain integer and floating point operation. Due to varying processor power characteristics, utilizing AVX instructions may cause a) parts to operate below the base frequency b) some parts with Intel Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software and system configuration and you should consult your system manufacturer for more information. Intel Advanced Vector Extensions refers to Intel AVX, Intel AVX2 or Intel AVX-512.
For more information on Intel AVX, see http://www-ssl.intel.com/content/www/us/en/architecture-and-technology/turbo-boost/turbo-boost-technology.html

***Note:*** Intel AVX2 Technology may not be available on all SKUs.

### 3.3.4 Intel® 64 Architecture x2APIC

The x2APIC architecture extends the xAPIC architecture that provides key mechanisms for interrupt delivery. This extension is primarily intended to increase processor addressability.

Specifically, x2APIC:

- Retains all key elements of compatibility to the xAPIC architecture:
  - — Delivery modes
  - — Interrupt and processor priorities
  - — Interrupt sources
  - — Interrupt destination types
- Provides extensions to scale processor addressability for both the logical and physical destination modes
- Adds new features to enhance performance of interrupt delivery
- Reduces complexity of logical destination mode interrupt delivery on link based architectures

The key enhancements provided by the x2APIC architecture over xAPIC are the following:

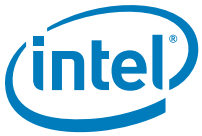
- Support for two modes of operation to provide backward compatibility and extensibility for future platform innovations:
  - In xAPIC compatibility mode, APIC registers are accessed through memory mapped interface to a 4K-Byte page, identical to the xAPIC architecture.
  - In x2APIC mode, APIC registers are accessed through Model Specific Register (MSR) interfaces. In this mode, the x2APIC architecture provides significantly increased processor addressability and some enhancements on interrupt delivery.
- Increased range of processor addressability in x2APIC mode:
  - Physical xAPIC ID field increases from 8 bits to 32 bits, allowing for interrupt processor addressability up to 4G-1 processors in physical destination mode. A processor implementation of x2APIC architecture can support fewer than 32-bits in a software transparent fashion.
  - Logical xAPIC ID field increases from 8 bits to 32 bits. The 32-bit logical x2APIC ID is partitioned into two sub-fields – a 16-bit cluster ID and a 16-bit logical ID within the cluster. Consequently, ((2^20) - 16) processors can be addressed in logical destination mode. Processor implementations can support fewer than 16 bits in the cluster ID sub-field and logical ID sub-field in a software agnostic fashion.
- More efficient MSR interface to access APIC registers:
  - To enhance inter-processor and self-directed interrupt delivery as well as the ability to virtualize the local APIC, the APIC register set can be accessed only through MSR-based interfaces in x2APIC mode. The Memory Mapped IO (MMIO) interface used by xAPIC is not supported in x2APIC mode.
- The semantics for accessing APIC registers have been revised to simplify the programming of frequently-used APIC registers by system software. Specifically, the software semantics for using the Interrupt Command Register (ICR) and End Of Interrupt (EOI) registers have been modified to allow for more efficient delivery and dispatching of interrupts.
- The x2APIC extensions are made available to system software by enabling the local x2APIC unit in the "x2APIC" mode. To benefit from x2APIC capabilities, a new operating system and a new BIOS are both needed, with special support for x2APIC mode.
- The x2APIC architecture provides backward compatibility to the xAPIC architecture and forward extendible for future Intel platform innovations.

***Note:*** Intel x2APIC Technology may not be available on all SKUs.

For more information, see the Intel® *64 Architecture x2APIC Specification* at http://www.intel.com/products/processor/manuals/.

## 3.3.5 Power Aware Interrupt Routing (PAIR)

The processor includes enhanced power-performance technology that routes interrupts to threads or processor IA cores based on their sleep states. As an example, for energy savings, it routes the interrupt to the active processor IA cores without waking the deep idle processor IA cores. For performance, it routes the interrupt to the idle (C1) processor IA cores without interrupting the already heavily loaded processor IA cores. This enhancement is mostly beneficial for high-interrupt scenarios like Gigabit LAN, WLAN peripherals, and so on.



### 3.3.6 Intel® Transactional Synchronization Extensions (Intel® TSX-NI)

Intel® Transactional Synchronization Extensions (Intel® TSX-NI) provides a set of instruction set extensions that allow programmers to specify regions of code for transactional synchronization. Programmers can use these extensions to achieve the performance of fine-grain locking while actually programming using coarse-grain locks. Details on Intel TSX-NI may be found in *Intel® Architecture Instruction Set Extensions Programming Reference*.

***Note:*** Intel TSX-NI may not be available on all SKUs.

## 3.4 Debug Technologies

### 3.4.1 Intel® Processor Trace

Intel® Processor Trace (Intel® PT) is a new tracing capability added to Intel Architecture, for use in software debug and profiling. Intel PT provides the capability for more precise software control flow and timing information, with limited impact to software execution. This provides enhanced ability to debug software crashes, hangs, or other anomalies, as well as responsiveness and short-duration performance issues.

Intel VTune™ Amplifier for Systems and the Intel System Debugger are part of Intel System Studio 2015, which includes updates for new debug and trace features on this latest platform, including Intel PT and Intel Trace Hub.

§ §



# 4 Power Management

This chapter provides information on the following power management topics:

- Advanced Configuration and Power Interface (ACPI) States
- Processor IA Core Power Management
- Integrated Memory Controller (IMC) Power Management
- PCI Express* Power Management
- Direct Media Interface (DMI) Power Management

## Figure 4-1. Processor Power States

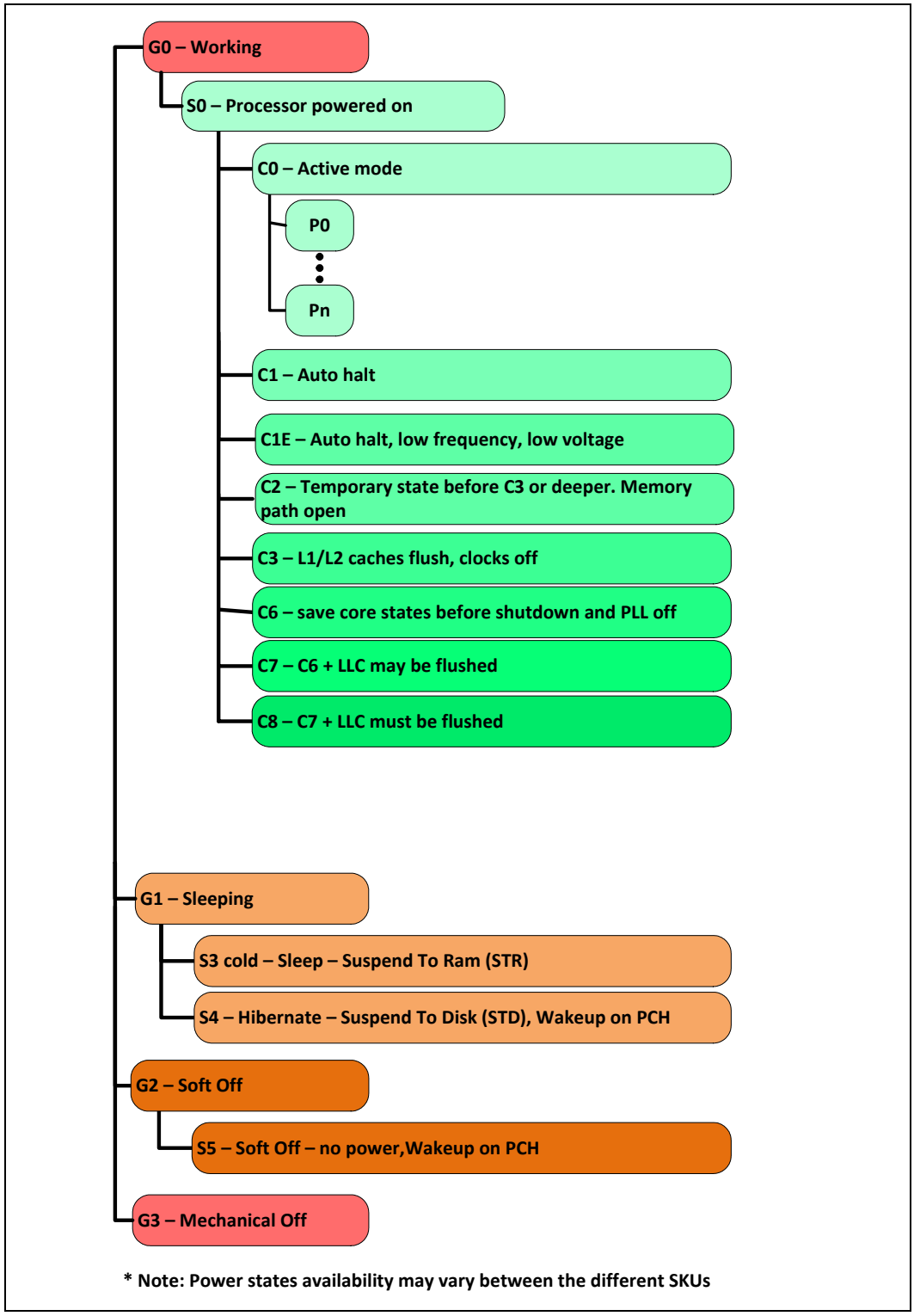
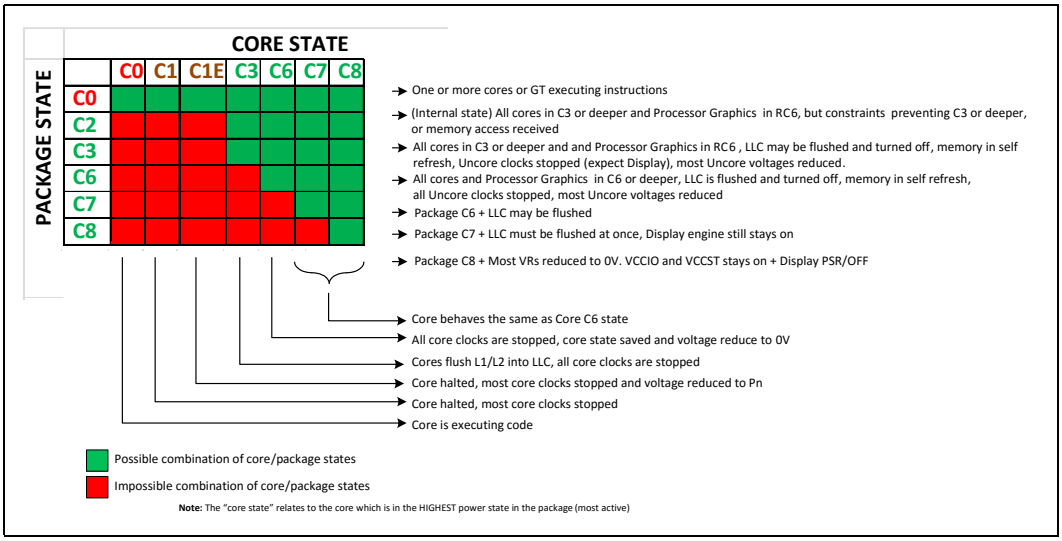


**Figure 4-2. Processor Package and IA Core C-States**

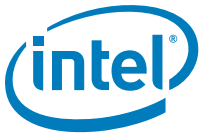



# 4.1 Advanced Configuration and Power Interface (ACPI) States Supported

This section describes the ACPI states supported by the processor.

**Table 4-1. System States**

| State | Description |
|---|---|
| G0/S0 | Full On |
| G1/S3-Cold | Suspend-to-RAM (STR). Context saved to memory (S3-Hot is not supported by the processor). |
| G1/S4 | Suspend-to-Disk (STD). All power lost (except wake-up on PCH). |
| G2/S5 | Soft off. All power lost (except wake-up on PCH). Total reboot. |
| G3 | Mechanical off. All power removed from system. |



### Table 4-2. Processor IA Core / Package State Support

| State | Description |
|---|---|
| C0 | Active mode, processor executing code. |
| C1 | AutoHALT processor IA core state (package C0 state). |
| C1E | AutoHALT processor IA core state with lowest frequency and voltage operating point (package C0 state). |
| C2 | All processor IA cores in C3 or deeper. Memory path open. Temporary state before Package C3 or deeper. |
| C3 | Processor IA execution cores in C3 or deeper, flush their L1 instruction cache, L1 data cache, and L2 cache to the LLC shared cache. LLC may be flushed. Clocks are shut off to each core. |
| C6 | Processor IA execution cores in this state save their architectural state before removing core voltage. BCLK is off. |
| C7 | Processor IA execution cores in this state behave similarly to the C6 state. If all execution cores request C7, LLC ways may be flushed until it is cleared. If the entire LLC is flushed, voltage will be removed from the LLC. |
| C8 | C7 plus LLC should be flushed. |

### Table 4-3. Integrated Memory Controller (IMC) States

| State | Description |
|---|---|
| Power up | CKE asserted. Active mode. |
| Pre-charge Power down | CKE de-asserted (not self-refresh) with all banks closed. |
| Active Power down | CKE de-asserted (not self-refresh) with minimum one bank active. |
| Self-Refresh | CKE de-asserted using device self-refresh. |

### Table 4-4. PCI Express* Link States

| State | Description |
|---|---|
| L0 | Full on – Active transfer state. |
| L1 | Lowest Active Power Management – Longer exit latency |
| L3 | Lowest power state (power-off) – Longest exit latency |

### Table 4-5. Direct Media Interface (DMI) States

| State | Description |
|---|---|
| L0 | Full on – Active transfer state |
| L1 | Lowest Active Power Management – Longer exit latency |
| L3 | Lowest power state (power-off) – Longest exit latency |

**Table 4-6. G, S, and C Interface State Combinations**

| Global (G) State | Sleep (S) State | Processor Package (C) State | Processor State | System Clocks | Description |
|---|---|---|---|---|---|
| G0 | S0 | C0 | Full On | On | Full On |
| G0 | S0 | C1/C1E | Auto-Halt | On | Auto-Halt |
| G0 | S0 | C3 | Deep Sleep | On | Deep Sleep |
| G0 | S0 | C6/C7 | Deep Power Down | On | Deep Power Down |
| G0 | S0 | C8 | Off | On | Deeper Power Down |
| G1 | S3 | Power off | Off | Off, except RTC | Suspend to RAM |
| G1 | S4 | Power off | Off | Off, except RTC | Suspend to Disk |
| G2 | S5 | Power off | Off | Off, except RTC | Soft Off |
| G3 | N/A | Power off | Off | Power off | Hard off |

## 4.2 Processor IA Core Power Management

While executing code, Enhanced Intel SpeedStep Technology and Intel Speed Shift® Technology optimizes the processor's IA core frequency and voltage based on workload. Each frequency and voltage operating point is defined by ACPI as a P-state. When the processor is not executing code, it is idle. A low-power idle state is defined by ACPI as a C-state. In general, deeper power C-states have longer entry and exit latencies.

### 4.2.1 OS/HW controlled P-states

#### 4.2.1.1 Enhanced Intel® SpeedStep® Technology

Enhanced Intel® SpeedStep® Technology enables OS to control and select P-state. The following are the key features of Enhanced Intel SpeedStep Technology:

- Multiple frequency and voltage points for optimal performance and power efficiency. These operating points are known as P-states.
- Frequency selection is software controlled by writing to processor MSRs. The voltage is optimized based on the selected frequency and the number of active processor IA cores.
  - — Once the voltage is established, the PLL locks on to the target frequency.
  - — All active processor IA cores share the same frequency and voltage. In a multi-core processor, the highest frequency P-state requested among all active IA cores is selected.
  - — Software-requested transitions are accepted at any time. If a previous transition is in progress, the new transition is deferred until the previous transition is completed.
- The processor controls voltage ramp rates internally to ensure glitch-free transitions.
- Because there is low transition latency between P-states, a significant number of transitions per-second are possible.

#### 4.2.1.2 Intel® Speed Shift Technology

Intel Speed Shift Technology is an energy efficient method of frequency control by the hardware rather than relying on OS control. OS is aware of available hardware P-states and request a desired P-state or it can let Hardware determine the P-state. The OS request is based on its workload requirements and awareness of processor capabilities. Processor decision is based on the different system constraints for example: Workload demand, thermal limits while taking into consideration the minimum and maximum levels and activity window of performance requested by the operating system.

For more details, refer to the following document (see related documents section):

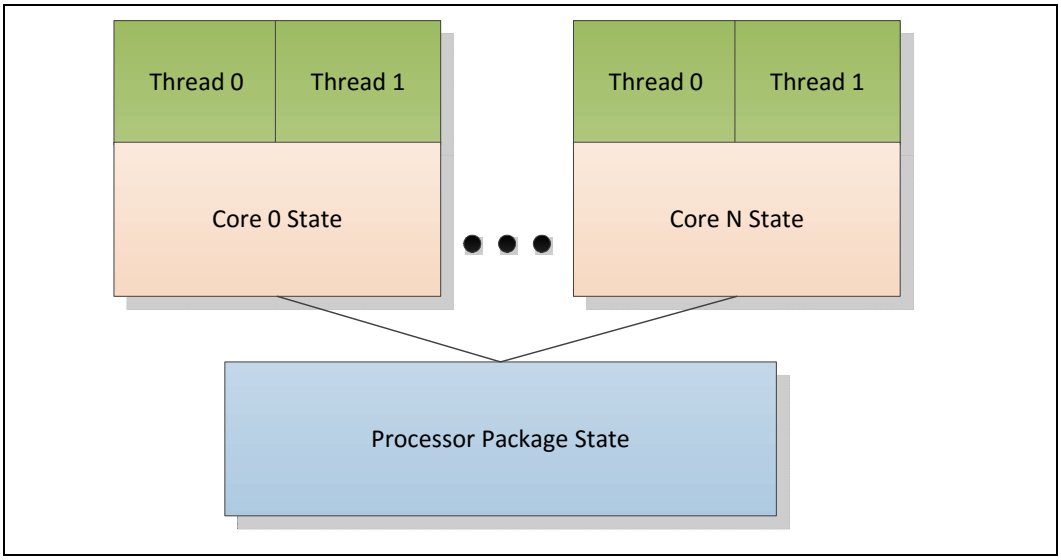
- Intel® 64 and IA-32 Architectures Software Developer's Manual (SDM), volume 3B.

### 4.2.2 Low-Power Idle States

When the processor is idle, low-power idle states (C-states) are used to save power. More power savings actions are taken for numerically higher C-states. However, deeper C-states have longer exit and entry latencies. Resolution of C-states occur at the thread, processor IA core, and processor package level. Thread-level C-states are available if Intel Hyper-Threading Technology is enabled.

***Caution:*** Long term reliability cannot be assured unless all the Low-Power Idle States are enabled.

**Figure 4-3. Idle Power Management Breakdown of the Processor IA Cores**




While individual threads can request low-power C-states, power saving actions only take place once the processor IA core C-state is resolved. processor IA core C-states are automatically resolved by the processor. For thread and processor IA core C-states, a transition to and from C0 state is required before entering any other C-state.

### 4.2.3 Requesting Low-Power Idle States

The primary software interfaces for requesting low-power idle states are through the MWAIT instruction with sub-state hints and the HLT instruction (for C1 and C1E). However, software may make C-state requests using the legacy method of I/O reads from the ACPI-defined processor clock control registers, referred to as P_LVLx. This method of requesting C-states provides legacy support for operating systems that initiate C-state transitions using I/O reads.

For legacy operating systems, P_LVLx I/O reads are converted within the processor to the equivalent MWAIT C-state request. Therefore, P_LVLx reads do not directly result in I/O reads to the system. The feature, known as I/O MWAIT redirection, should be enabled in the BIOS.

The BIOS can write to the C-state range field of the PMG_IO_CAPTURE MSR to restrict the range of I/O addresses that are trapped and emulate MWAIT like functionality. Any P_LVLx reads outside of this range do not cause an I/O redirection to MWAIT(Cx) like request. They fall through like a normal I/O instruction.

When P_LVLx I/O instructions are used, MWAIT sub-states cannot be defined. The MWAIT sub-state is always zero if I/O MWAIT redirection is used. By default, P_LVLx I/O redirections enable the MWAIT 'break on EFLAGS.IF' feature that triggers a wake up on an interrupt, even if interrupts are masked by EFLAGS.IF.

### 4.2.4 Processor IA Core C-State Rules

The following are general rules for all processor IA core C-states, unless specified otherwise:

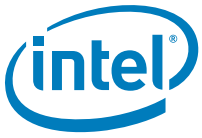
- A processor IA core C-State is determined by the lowest numerical thread state (such as Thread 0 requests C1E while Thread 1 requests C3 state, resulting in a processor IA core C1E state). See the *G, S, and C Interface State Combinations* table.
- A processor IA core transitions to C0 state when:
  - An interrupt occurs
  - There is an access to the monitored address if the state was entered using an MWAIT/Timed MWAIT instruction
  - The deadline corresponding to the Timed MWAIT instruction expires
- An interrupt directed toward a single thread wakes up only that thread.
- If any thread in a processor IA core is active (in C0 state), the core's C-state will resolve to C0.
- Any interrupt coming into the processor package may wake any processor IA core.
- A system reset re-initializes all processor IA cores.

**processor IA core C0 State**

The normal operating state of a processor IA core where code is being executed.

**processor IA core C1/C1E State**

C1/C1E is a low-power state entered when all threads within a processor IA core execute a HLT or MWAIT(C1/C1E) instruction.



A System Management Interrupt (SMI) handler returns execution to either Normal state or the C1/C1E state. See the *Intel 64 and IA-32 Architectures Software Developer's Manual* for more information.

While a processor IA core is in C1/C1E state, it processes bus snoops and snoops from other threads. For more information on C1E, see Section 4.2.5.

### processor IA core C3 State

Individual threads of a processor IA core can enter the C3 state by initiating a P_LVL2 I/O read to the P_BLK or an MWAIT(C3) instruction. A processor IA core in C3 state flushes the contents of its L1 instruction cache, L1 data cache, and L2 cache to the shared LLC, while maintaining its architectural state. All processor IA core clocks are stopped at this point. Because the processor IA core's caches are flushed, the processor does not wake any processor IA core that is in the C3 state when either a snoop is detected or when another processor IA core accesses cacheable memory.

### processor IA core C6 State

Individual threads of a processor IA core can enter the C6 state by initiating a P_LVL3 I/O read or an MWAIT(C6) instruction. Before entering processor IA core C6 state, the processor IA core will save its architectural state to a dedicated SRAM. Once complete, a processor IA core will have its voltage reduced to zero volts. During exit, the processor IA core is powered on and its architectural state is restored.

### processor IA core C7-C8 States

Individual threads of a processor IA core can enter the C7, C8 state by initiating a P_LVL4, P_LVL5, P_LVL6, P_LVL7 I/O read (respectively) to the P_BLK or by an MWAIT(C7/C8) instruction. The processor IA core C7-C8 state exhibits the same behavior as the processor IA core C6 state.

### C-State Auto-Demotion

In general, deeper C-states, such as C6 or C7, have long latencies and have higher energy entry/exit costs. The resulting performance and energy penalties become significant when the entry/exit frequency of a deeper C-state is high. Therefore, incorrect or inefficient usage of deeper C-states have a negative impact on battery life and idle power. To increase residency and improve battery life and idle power in deeper C-states, the processor supports C-state auto-demotion.

There are two C-State auto-demotion options:

- C7/C6 to C3
- C7/C6/C3 To C1

The decision to demote a processor IA core from C6/C7 to C3 or C3/C6/C7 to C1 is based on each processor IA core's immediate residency history. Upon each processor IA core C6/C7 request, the processor IA core C-state is demoted to C3 or C1 until a sufficient amount of residency has been established. At that point, a processor IA core is allowed to go into C3/C6 or C7. Each option can be run concurrently or individually. If the interrupt rate experienced on a processor IA core is high and the processor IA core is rarely in a deep C-state between such interrupts, the processor IA core can be demoted to a C3 or C1 state. A higher interrupt pattern is required to demote a processor IA core to C1 as compared to C3.

This feature is disabled by default. BIOS should enable it in the PMG_CST_CONFIG_CONTROL register. The auto-demotion policy is also configured by this register.

### 4.2.5 Package C-States

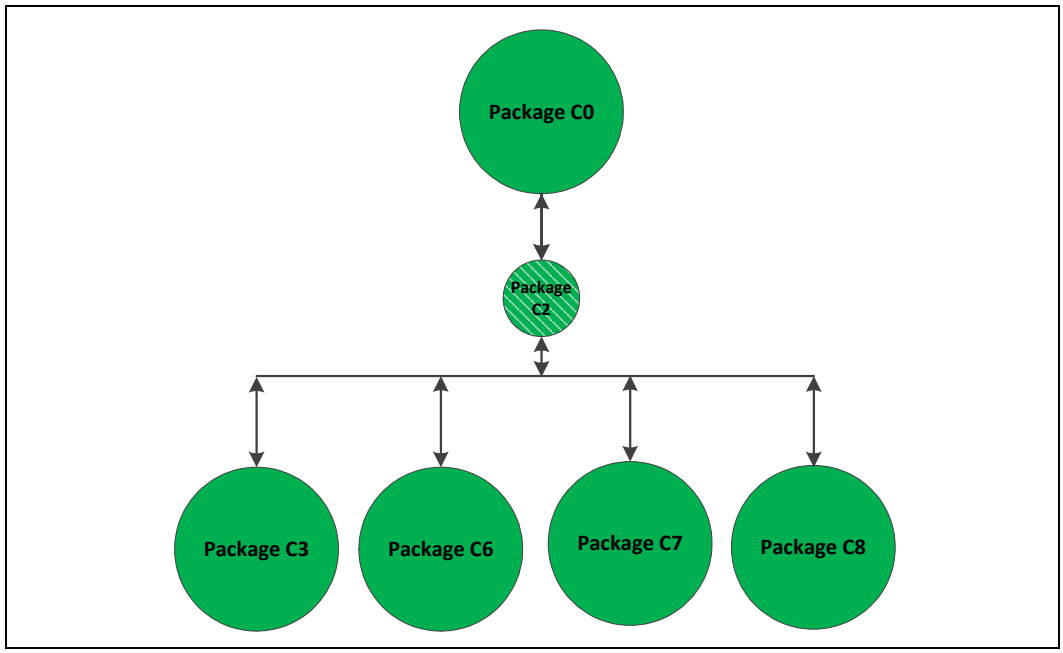
The processor supports C0, C2, C3, C6, C7, C8 package states. The following is a summary of the general rules for package C-state entry. These apply to all package C-states, unless specified otherwise:

- A package C-state request is determined by the lowest numerical processor IA core C-state amongst all processor IA cores.
- A package C-state is automatically resolved by the processor depending on the processor IA core idle power states and the status of the platform components.
  - Each processor IA core can be at a lower idle power state than the package if the platform does not grant the processor permission to enter a requested package C-state.
  - The platform may allow additional power savings to be realized in the processor.
  - For package C-states, the processor is not required to enter C0 before entering any other C-state.
  - Entry into a package C-state may be subject to auto-demotion – that is, the processor may keep the package in a deeper package C-state then requested by the operating system if the processor determines, using heuristics, that the deeper C-state results in better power/performance.

The processor exits a package C-state when a break event is detected. Depending on the type of break event, the processor does the following:

- If a processor IA core break event is received, the target processor IA core is activated and the break event message is forwarded to the target processor IA core.
  - If the break event is not masked, the target processor IA core enters the processor IA core C0 state and the processor enters package C0.
  - If the break event is masked, the processor attempts to re-enter its previous package state.
- If the break event was due to a memory access or snoop request,
  - But the platform did not request to keep the processor in a higher package C-state, the package returns to its previous C-state.
  - And the platform requests a higher power C-state, the memory access or snoop request is serviced and the package remains in the higher power C-state.

**Figure 4-4. Package C-State Entry and Exit**




### Package C0

This is the normal operating state for the processor. The processor remains in the normal state when at least one of its processor IA cores is in the C0 or C1 state or when the platform has not granted permission to the processor to go into a low-power state. Individual processor IA cores may be in deeper power idle states while the package is in C0 state.

### Package C2 State

Package C2 state is an internal processor state that cannot be explicitly requested by software. A processor enters Package C2 state when either:

- All processor IA cores have requested a C3 or deeper power state, but constraints (LTR, programmed timer events in the near future, and so forth) prevent entry to any state deeper than C2 state.
- Or, all processor IA cores have requested a C3 or deeper power state and a memory access request is received. Upon completion of all outstanding memory requests, the processor transitions back into a deeper package C-state.

### Package C3 State

A processor enters the package C3 low-power state when:

- At least one processor IA core is in the C3 state.
- The other processor IA cores are in a C3 or deeper power state, and the processor has been granted permission by the platform.
- The platform has not granted a request to a package C6/C7 state or deeper state but has allowed a package C3 state.

In package C3-state, the LLC shared cache is valid.

### Package C6 State

A processor enters the package C6 low-power state when:

- At least one processor IA core is in the C6 state.
- The other processor IA cores are in a C6 or deeper power state, and the processor has been granted permission by the platform.
- The platform has not granted a package C7 or deeper request but has allowed a C6 package state.

In package C6 state, all processor IA cores have saved their architectural state and have had their voltages reduced to zero volts. It is possible the LLC shared cache is flushed and turned off in package C6 state.

### Package C7 State

The processor enters the package C7 low-power state when all processor IA cores are in the C7 or deeper state and the operating system may request that the LLC will be flushed.

processor IA core break events are handled the same way as in package C3 or C6.

Upon exit of the package C7 state, the LLC will be partially enabled once a processor IA core wakes up if it was fully flushed, and will be fully enabled once the processor has stayed out of C7 for a preset amount of time. Power is saved since this prevents the LLC from being re-populated only to be immediately flushed again. Some VRs are reduce to 0V.

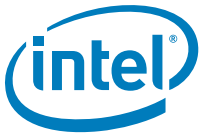
### Package C8 State

The processor enters C8 states when the processor IA cores lower numerical state is C8.

The C8 state is similar to C7 state, but in addition, the LLC is flushed in a single step, Vcc are reduced to 0V.

### Dynamic LLC Sizing

When all processor IA cores request C7 or deeper C-state, internal heuristics dynamically flushes the LLC. Once the processor IA cores enter a deep C-state, depending on their MWAIT sub-state request, the LLC is either gradually flushed N-ways at a time or flushed all at once. Upon the processor IA cores exiting to C0 state, the LLC is gradually expanded based on internal heuristics.



## 4.3 Integrated Memory Controller (IMC) Power Management

The main memory is power managed during normal operation and in low-power ACPI C-states.

### 4.3.1 Disabling Unused System Memory Outputs

Any system memory (SM) interface signal that goes to a memory in which it is not connected to any actual memory devices is tri-stated. The benefits of disabling unused SM signals are:

- Reduced power consumption.
- Reduced possible overshoot/undershoot signal quality issues seen by the processor I/O buffer receivers caused by reflections from potentially un-terminated transmission lines.

When a given rank is not populated, the corresponding control signals (CLK_P/CLK_N/CKE/ODT/CS) are not driven.

At reset, all rows should be assumed to be populated, until it can be proven that they are not populated. This is due to the fact that when CKE is tri-stated with a DRAMs present, the DRAMs are not ensured to maintain data integrity. CKE tri-state should be enabled by BIOS where appropriate, since at reset all rows should be assumed to be populated.

### 4.3.2 DRAM Power Management and Initialization

The processor implements extensive support for power management on the memory interface.Each channel drives 4 CKE pins, one per rank.

The CKE is one of the power-saving means. When CKE is off, the internal DDR clock is disabled and the DDR power is reduced. The power-saving differs according to the selected mode and the DDR type used. For more information, refer to the IDD table in the DDR specification.

The processor supports four different types of power-down modes in package C0 state. The different power-down modes can be enabled through configuring PM PDWN configuration register. The type of CKE power-down can be configured through PDWN_mode (bits 15:12) and the idle timer can be configured through PDWN_idle_counter (bits 11:0). The different power-down modes supported are:

- **No power-down** (CKE disable)
- **Active power-down (APD):** This mode is entered if there are open pages when de-asserting CKE. In this mode the open pages are retained. Power-saving in this mode is the lowest. Power consumption of DDR is defined by IDD3P. Exiting this mode is fined by tXP – small number of cycles. For this mode, DRAM DLL should be on.
- **PPD/DLL-off:** In this mode the data-in DLLs on DDR are off. Power-saving in this mode is the best among all power modes. Power consumption is defined by IDD2P. Exiting this mode is defined by tXP, but also tXPDLL (10–20 according to DDR type) cycles until first data transfer is allowed. For this mode, DRAM DLL should be off.
- **Precharged power-down (PPD):** This mode is entered if all banks in DDR are precharged when de-asserting CKE. Power-saving in this mode is intermediate –

better than APD, but less than DLL-off. Power consumption is defined by IDD2P. Exiting this mode is defined by tXP. The difference from APD mode is that when waking-up, all page-buffers are empty.) The LPDDR does not have a DLL. As a result, the power savings are as good as PPD/DDL-off but will have lower exit latency and higher performance.

The CKE is determined per rank, whenever it is inactive. Each rank has an idle counter. The idle-counter starts counting as soon as the rank has no accesses, and if it expires, the rank may enter power-down while no new transactions to the rank arrives to queues. The idle-counter begins counting at the last incoming transaction arrival.

It is important to understand that since the power-down decision is per rank, the IMC can find many opportunities to power down ranks, even while running memory intensive applications; the savings are significant (may be few Watts, according to DDR specification). This is significant when each channel is populated with more ranks.

Selection of power modes should be according to power-performance or thermal trade-off of a given system:

- When trying to achieve maximum performance and power or thermal consideration is not an issue: use no power-down
- In a system which tries to minimize power-consumption, try using the deepest power-down mode possible – PPD/DLL-off with a low idle timer value
- In high-performance systems with dense packaging (that is, tricky thermal design) the power-down mode should be considered in order to reduce the heating and avoid DDR throttling caused by the heating.

The default value that BIOS configures in PM PDWN configuration register is 6080 – that is, PPD/DLL-off mode with idle timer of 0x80, or 128 DCLKs. This is a balanced setting with deep power-down mode and moderate idle timer value.

The idle timer expiration count defines the # of DCLKs that a rank is idle that causes entry to the selected power mode. As this timer is set to a shorter time the IMC will have more opportunities to put the DDR in power-down. There is no BIOS hook to set this register. Customers choosing to change the value of this register can do it by changing it in the BIOS. For experiments, this register can be modified in real time if BIOS does not lock the IMC registers.

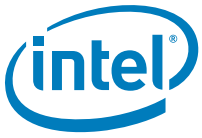
#### 4.3.2.1 Initialization Role of CKE

During power-up, CKE is the only input to the SDRAM that has its level recognized (other than the reset pin) once power is applied. It should be driven LOW by the DDR controller to make sure the SDRAM components float DQ and DQS during power-up. CKE signals remain LOW (while any reset is active) until the BIOS writes to a configuration register. Using this method, CKE is ensured to remain inactive for much longer than the specified 200 micro-seconds after power and clocks to SDRAM devices are stable.

#### 4.3.2.2 Conditional Self-Refresh

During S0 idle state, system memory may be conditionally placed into self-refresh state when the processor is in package C3 or deeper power state.

When entering the S3 – Suspend-to-RAM (STR) state or S0 conditional self-refresh, the processor IA core flushes pending cycles and then enters SDRAM ranks. The CKE signals remain LOW so the SDRAM devices perform self-refresh.



The target behavior is to enter self-refresh for package C3 or deeper power states as long as there are no memory requests to service.

**Table 4-7. Targeted Memory State Conditions**

| State | Memory State with External Graphics |
|---|---|
| C0, C1, C1E | Dynamic memory rank power-down based on idle conditions. |
| C3, C6, C7 or deeper | If there are no memory requests, then enter self-refresh. Otherwise use dynamic memory rank power-down based on idle conditions. |
| S3 | Self-Refresh Mode |
| S4 | Memory power-down (contents lost) |

#### 4.3.2.3 Dynamic Power-Down

Dynamic power-down of memory is employed during normal operation. Based on idle conditions, a given memory rank may be powered down. The IMC implements aggressive CKE control to dynamically put the DRAM devices in a power-down state. The processor IA core controller can be configured to put the devices in active power-down (CKE de-assertion with open pages) or precharge power-down (CKE de-assertion with all pages closed). Precharge power-down provides greater power savings but has a bigger performance impact, since all pages will first be closed before putting the devices in power-down mode.

If dynamic power-down is enabled, all ranks are powered up before doing a refresh cycle and all ranks are powered down at the end of refresh.

#### 4.3.2.4 DRAM I/O Power Management

Unused signals should be disabled to save power and reduce electromagnetic interference. This includes all signals associated with an unused memory channel. Clocks, CKE, ODT and CS signals are controlled per DIMM rank and will be powered down for unused ranks.

The I/O buffer for an unused signal should be tri-stated (output driver disabled), the input receiver (differential sense-amp) should be disabled, and any DLL circuitry related ONLY to unused signals should be disabled. The input path should be gated to prevent spurious results due to noise on the unused signals (typically handled automatically when input receiver is disabled).

### 4.3.3 DDR Electrical Power Gating (EPG)

The DDR I/O of the processor supports Electrical Power Gating (DDR-EPG) while the processor is at C3 or deeper power state.

In C3 or deeper power state, the processor internally gates VDDQ for the majority of the logic to reduce idle power while keeping all critical DDR pins such as CKE and VREF in the appropriate state.

In C7 or deeper power state, the processor internally gates $V_{CCIO}$ for all non-critical state to reduce idle power.

In S3 or C-state transitions, the DDR does not go through training mode and will restore the previous training information.

### 4.3.4 Power Training

BIOS MRC performing Power Training steps to reduce DDR I/O power while keeping reasonable operational margins, still ensuring platform operation. The algorithms attempt to weaken ODT, driver strength and the related buffers parameters both on the MC and the DRAM side and find the best possible trade-off between the total I/O power and the operational margins using advanced mathematical models.

## 4.4 PCI Express* Power Management

- Active power management support using L1 state.
- All inputs and outputs disabled in L2/L3 Ready state.

***Note:*** Processor PEG-PCIe interface does not support Hot-Plug.

Hot Plug like* is only supported at Processor PEG-PCIe using Thunderbolt Device.

* Turning Thunderbolt™ power on and Off electrically RTD3 Like

An increase in power consumption may be observed when PCI Express* ASPM capabilities are disabled.

**Table 4-8. Package C-States with PCIe* Link States dependencies**

| PEG/DMI | L-State | Description | Package C-State |
|---|---|---|---|
| DMI | L1 | Higher latency, lower power “standby” state | PC6-PC8 |
| PEG | L1, L2, Disabled, NDA (no device attached) | L1- Higher latency, lower power “standby” state<br>L2 – Auxiliary-powered Link, deep-energy-saving state.<br>Disabled - The intent of the Disabled state is to allow a configured Link to be disabled until directed or Electrical Idle is exited (i.e., due to a hot removal and insertion) after entering Disabled.<br>NDA- no physical device is attached on PEG port | PC6-PC7 |
| PEG | L2, Disabled, NDA (no device attached) | L2 – Auxiliary-powered Link, deep-energy-saving state.<br>Disabled - The intent of the Disabled state is to allow a configured Link to be disabled until directed or Electrical Idle is exited (i.e., due to a hot removal and insertion) after entering Disabled.<br>NDA- no physical device is attached on PEG port | PC8-PC8 |

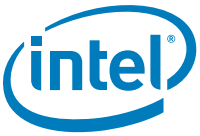
## 4.5 Direct Media Interface (DMI) Power Management

- Active power management support using L1 state.

## 4.6 Voltage Optimization

Voltage Optimization opportunistically provides reduction in power consumption, that is, a boost in performance at a given PL1. Over time the benefit is reduced. There is no change to base frequency or turbo frequency. During system validation and tuning, this feature should be disabled to reflect processor power and performance that is expected over time.

§ §



# 5 Thermal Management

## 5.1 Processor Thermal Management

The thermal solution provides both component-level and system-level thermal management. To allow optimal operation and long-term reliability of Intel processor-based systems, the system/processor thermal solution should be designed so that the processor:

- Bare Die Parts: Remains below the maximum junction temperature ($Tj_{MAX}$) specification at the maximum thermal design power (TDP).
- Lidded Parts: Remains below the maximum case temperature (Tcmax) specification at the maximum thermal design power.
- Conforms to system constraints, such as system acoustics, system skin-temperatures, and exhaust-temperature requirements.

***Caution:*** Thermal specifications given in this chapter are on the component and package level and apply specifically to the processor. Operating the processor outside the specified limits may result in permanent damage to the processor and potentially other components in the system.

### 5.1.1 Thermal Considerations

The processor TDP is the maximum sustained power that should be used for design of the processor thermal solution. TDP is a power dissipation and component temperature operating condition limit, specified in this document, that is validated during manufacturing for the base configuration when executing a near worst case commercially available workload as specified by Intel for the SKU segment. TDP may be exceeded for short periods of time or if running a very high power workload.

To allow the optimal operation and long-term reliability of Intel processor-based systems, the processor must remain within the minimum and maximum component temperature specifications. For lidded parts, the appropriate case temperature ($T_{CASE}$) specifications is defined by the applicable thermal profile. For bare die parts, the component temperature specification is the applicable $Tj_{MAX}$.

Thermal solutions not designed to provide this level of thermal capability may affect the long-term reliability of the processor and system.

The processor integrates multiple processing IA cores. This may result in power distribution differences across the package and should be considered when designing the thermal solution.

Intel Turbo Boost Technology 2.0 allows processor IA cores to run faster than the base frequency. It is invoked opportunistically and automatically as long as the processor is conforming to its temperature, voltage, power delivery and current control limits. When Intel Turbo Boost Technology 2.0 is enabled:

- Applications are expected to run closer to TDP more often as the processor will attempt to maximize performance by taking advantage of estimated available energy budget in the processor package.

- The processor may exceed the TDP for short durations to use any available thermal capacitance within the thermal solution. The duration and time of such operation can be limited by platform runtime configurable registers within the processor.
- Thermal solutions and platform cooling that are designed to less than thermal design guidance may experience thermal and performance issues.

***Note:*** Intel Turbo Boost Technology 2.0 availability may vary between the different SKUs.

## 5.1.2 Intel® Turbo Boost Technology 2.0 Power Monitoring

When operating in turbo mode, the processor monitors its own power and adjusts the processor frequencies to maintain the average power within limits over a thermally significant time period. The processor estimates the package power for all components on package. In the event that a workload causes the temperature to exceed program temperature limits, the processor will protect itself using the Adaptive Thermal Monitor.

## 5.1.3 Intel® Turbo Boost Technology 2.0 Power Control

Illustration of Intel® Turbo Boost Technology 2.0 power control is shown in the following sections and figures. Multiple controls operate simultaneously allowing customization for multiple system thermal and power limitations. These controls allow for turbo optimizations within system constraints and are accessible using MSR, MMIO, or PECI interfaces.

### 5.1.3.1 Package Power Control

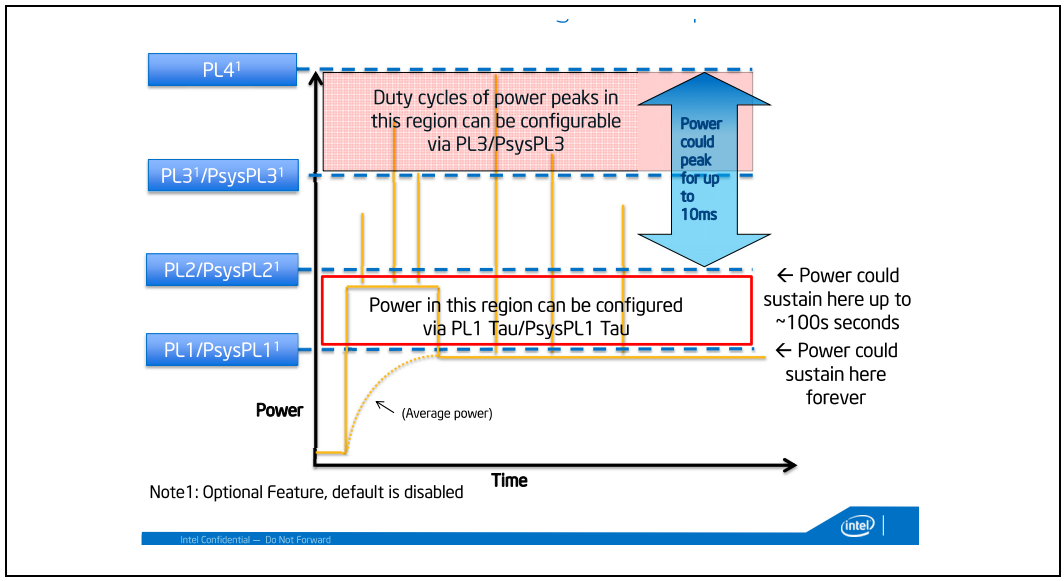
The package power control settings of PL1, PL2, PL3, PL4 and Tau allow the designer to configure Intel Turbo Boost Technology 2.0 to match the platform power delivery and package thermal solution limitations.

- Power Limit 1 (PL1): A threshold for average power that will not exceed - recommend to set to equal TDP power. PL1 should not be set higher than thermal solution cooling limits.
- Power Limit 2 (PL2): A threshold that if exceeded, the PL2 rapid power limiting algorithms will attempt to limit the spike above PL2.
- Power Limit 3 (PL3): A threshold that if exceeded, the PL3 rapid power limiting algorithms will attempt to limit the duty cycle of spikes above PL3 by reactively limiting frequency. This is an optional setting
- Power Limit 4 (PL4): A limit that will not be exceeded, the PL4 power limiting algorithms will preemptively limit frequency to prevent spikes above PL4.
- Turbo Time Parameter (Tau): An averaging constant used for PL1 exponential weighted moving average (EWMA) power calculation.

***Note:*** Implementation of Intel Turbo Boost Technology 2.0 only requires configuring PL1, PL1 Tau, and PL2.

***Note:*** PL3 and PL4 are disabled by default.

**Figure 5-1. Package Power Control**




### 5.1.3.2 Platform Power Control

The processor supports Psys (Platform Power) to enhance processor power management. The Psys signal needs to be sourced from a compatible charger circuit and routed to the IMVP8 (voltage regulator). This signal will provide the total thermally relevant platform power consumption (processor and rest of platform) via SVID to the processor.

When the Psys signal is properly implemented, the system designer can utilize the package power control settings of PsysPL1/Tau, PsysPL2 and PsysPL3 for additional manageability to match the platform power delivery and platform thermal solution limitations for Intel Turbo Boost Technology 2.0. The operation of the PsysPL1/tau, PsysPL2 and PsysPL3 is analogous to the processor power limits described in Section 5.1.3.1.

- Platform Power Limit 1 (PsysPL1): A threshold for average platform power that will not be exceeded - recommend to set to equal platform thermal capability.
- Platform Power Limit 2 (PsysPL2): A threshold that if exceeded, the PsysPL2 rapid power limiting algorithms will attempt to limit the spikes above PsysPL2.
- Platform Power Limit 3 (PsysPL3): A threshold that if exceeded, the PsysPL3 rapid power limiting algorithms will attempt to limit the duty cycle of spikes above PsysPL3 by reactively limiting frequency.
- PsysPL1 Tau: An averaging constant used for PsysPL1 exponential weighted moving average (EWMA) power calculation.
- The Psys signal and associated power limits / Tau are optional for the system designer and disabled by default.
- The Psys data will not include power consumption for charging.

#### 5.1.3.3 Turbo Time Parameter (Tau)

Turbo Time Parameter (Tau) is a mathematical parameter (units of seconds) that controls the Intel Turbo Boost Technology 2.0 algorithm. During a maximum power turbo event, the processor could sustain PL2 for a duration longer than the Turbo Time Parameter. If the power value and/or Turbo Time Parameter is changed during runtime, it may take some time based on the new Turbo Time Parameter level for the algorithm to settle at the new control limits. The time varies depending on the magnitude of the change, power limits, and other factors. There is an individual Turbo Time Parameter associated with Package Power Control and Platform Power Control.

### 5.1.4 Thermal Management Features

Occasionally the processor may operate in conditions that are near to its maximum operating temperature. This can be due to internal overheating or overheating within the platform. In order to protect the processor and the platform from thermal failure, several thermal management features exist to reduce package power consumption and thereby temperature in order to remain within normal operating limits. Furthermore, the processor supports several methods to reduce memory power.

#### 5.1.4.1 Adaptive Thermal Monitor

The purpose of the Adaptive Thermal Monitor is to reduce processor IA core power consumption and temperature until it operates below its maximum operating temperature. Processor IA core power reduction is achieved by:

- Adjusting the operating frequency (using the processor IA core ratio multiplier) and voltage.
- Modulating (starting and stopping) the internal processor IA core clocks (duty cycle).

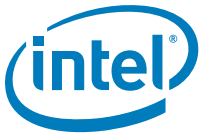
The Adaptive Thermal Monitor can be activated when the package temperature, monitored by any digital thermal sensor (DTS), meets its maximum operating temperature. The maximum operating temperature implies maximum junction temperature $Tj_{MAX}$.

Reaching the maximum operating temperature activates the Thermal Control Circuit (TCC). When activated, the TCC causes the processor IA core to reduce frequency and voltage adaptively. The Adaptive Thermal Monitor will remain active as long as the package temperature remains at its specified limit. Therefore, the Adaptive Thermal Monitor will continue to reduce the package frequency and voltage until the TCC is de-activated.

$Tj_{MAX}$ is factory calibrated and is not user configurable. The default value is software visible in the TEMPERATURE_TARGET (0x1A2) MSR, bits [23:16].

The Adaptive Thermal Monitor does not require any additional hardware, software drivers, or interrupt handling routines. It is not intended as a mechanism to maintain processor thermal control to PL1 = TDP. The system design should provide a thermal solution that can maintain normal operation when PL1 = TDP within the intended usage range.

Adaptive Thermal Monitor protection is always enabled.



#### 5.1.4.1.1 TCC Activation Offset

TCC Activation Offset can be set as an offset from the maximum allowed component temperature to lower the onset of TCC and Adaptive Thermal Monitor. In addition, the processor has added an optional time window (Tau) to manage processor performance at the TCC Activation offset value via an EWMA (Exponential Weighted Moving Average) of temperature.

**TCC Activation Offset with Tau=0**

An offset (degrees Celsius) can be written to the TEMPERATURE_TARGET (0x1A2) MSR, bits [29:24], the offset value will be subtracted from the value found in bits [23:16]. When the time window (Tau) is set to zero, there will be no averaging, the offset, will be subtracted from the $Tj_{MAX}$ value and used as a new max temperature set point for Adaptive Thermal Monitoring. This will have the same behavior as in prior products to have TCC activation and Adaptive Thermal Monitor to occur at this lower target silicon temperature.

If enabled, the offset should be set lower than any other passive protection such as ACPI _PSV trip points

**TCC Activation Offset with Tau**

To manage the processor with the EWMA (Exponential Weighted Moving Average) of temperature, an offset (degrees Celsius) is written to the TEMPERATURE_TARGET (0x1A2) MSR, bits [29:24], and the time window (Tau) is written to the TEMPERATURE_TARGET (0x1A2) MSR [6:0]. The Offset value will be subtracted from the value found in bits [23:16] and be the temperature.

The processor will manage to this average temperature by adjusting the frequency of the various domains. The instantaneous Tj can briefly exceed the average temperature. The magnitude and duration of the overshoot is managed by the time window value (Tau).

This averaged temperature thermal management mechanism is in addition, and not instead of $Tj_{MAX}$ thermal management. That is, whether the TCC activation offset is 0 or not, TCC Activation will occur at $Tj_{MAX}$.

#### 5.1.4.1.2 Frequency / Voltage Control

Upon Adaptive Thermal Monitor activation, the processor attempts to dynamically reduce processor temperature by lowering the frequency and voltage operating point. The operating points are automatically calculated by the processor IA core itself and do not require the BIOS to program them as with previous generations of Intel processors. The processor IA core will scale the operating points such that:

- The voltage will be optimized according to the temperature, the processor IA core bus ratio and number of processor IA cores in deep C-states.
- The processor IA core power and temperature are reduced while minimizing performance degradation.

Once the temperature has dropped below the trigger temperature, the operating frequency and voltage will transition back to the normal system operating point.

Once a target frequency/bus ratio is resolved, the processor IA core will transition to the new target automatically.

- On an upward operating point transition the voltage transition precedes the frequency transition.
- On a downward transition the frequency transition precedes the voltage transition.
- The processor continues to execute instructions. However, the processor will halt instruction execution for frequency transitions.

If a processor load-based Enhanced Intel SpeedStep Technology/P-state transition (through MSR write) is initiated while the Adaptive Thermal Monitor is active, there are two possible outcomes:

- If the P-state target frequency is higher than the processor IA core optimized target frequency, the P-state transition will be deferred until the thermal event has been completed.
- If the P-state target frequency is lower than the processor IA core optimized target frequency, the processor will transition to the P-state operating point.

#### 5.1.4.1.3 Clock Modulation

If the frequency/voltage changes are unable to end an Adaptive Thermal Monitor event, the Adaptive Thermal Monitor will utilize clock modulation. Clock modulation is done by alternately turning the clocks off and on at a duty cycle (ratio between clock "on" time and total time) specific to the processor. The duty cycle is factory configured to 25% on and 75% off and cannot be modified. The period of the duty cycle is configured to 32 microseconds when the Adaptive Thermal Monitor is active. Cycle times are independent of processor frequency. A small amount of hysteresis has been included to prevent excessive clock modulation when the processor temperature is near its maximum operating temperature. Once the temperature has dropped below the maximum operating temperature, and the hysteresis timer has expired, the Adaptive Thermal Monitor goes inactive and clock modulation ceases. Clock modulation is automatically engaged as part of the Adaptive Thermal Monitor activation when the frequency/voltage targets are at their minimum settings. Processor performance will be decreased when clock modulation is active. Snooping and interrupt processing are performed in the normal manner while the Adaptive Thermal Monitor is active.

Clock modulation will not be activated by the Package average temperature control mechanism.

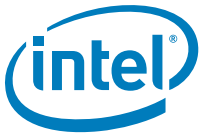
### 5.1.4.2 Digital Thermal Sensor

Each processor has multiple on-die Digital Thermal Sensor (DTS) that detects the processor IA, GT and other areas of interest instantaneous temperature.

Temperature values from the DTS can be retrieved through:

- A software interface using processor Model Specific Register (MSR).
- A processor hardware interface as described in Platform Environmental Control Interface (PECI).

When temperature is retrieved by the processor MSR, it is the instantaneous temperature of the given DTS. When temperature is retrieved using PECI, it is the average of the highest DTS temperature in the package over a 256 ms time window. Intel recommends using the PECI reported temperature for platform thermal control that benefits from averaging, such as fan speed control. The average DTS temperature may not be a good indicator of package Adaptive Thermal Monitor activation or rapid increases in temperature that triggers the Out of Specification status bit within the PACKAGE_THERM_STATUS MSR 1B1h and IA32_THERM_STATUS MSR 19Ch.



Code execution is halted in C1 or deeper C- states. Package temperature can still be monitored through PECI in lower C-states.

Unlike traditional thermal devices, the DTS outputs a temperature relative to the maximum supported operating temperature of the processor ($Tj_{MAX}$), regardless of TCC activation offset. It is the responsibility of software to convert the relative temperature to an absolute temperature. The absolute reference temperature is readable in the TEMPERATURE_TARGET MSR 1A2h. The temperature returned by the DTS is an implied negative integer indicating the relative offset from $Tj_{MAX}$. The DTS does not report temperatures greater than $Tj_{MAX}$. The DTS-relative temperature readout directly impacts the Adaptive Thermal Monitor trigger point. When a package DTS indicates that it has reached the TCC activation (a reading of 0x0, except when the TCC activation offset is changed), the TCC will activate and indicate an Adaptive Thermal Monitor event. A TCC activation will lower the processor IA core frequency, voltage, or both. Changes to the temperature can be detected using two programmable thresholds located in the processor thermal MSRs. These thresholds have the capability of generating interrupts using the processor IA core's local APIC. Refer to the *Intel 64 and IA-32 Architectures Software Developer's Manual* for specific register and programming details.

#### 5.1.4.2.1 Digital Thermal Sensor Accuracy (Taccuracy)

The error associated with DTS measurements will not exceed ±5 °C within the entire operating range.

#### 5.1.4.2.2 Fan Speed Control with Digital Thermal Sensor

Digital Thermal Sensor based fan speed control ($T_{FAN}$) is a recommended feature to achieve optimal thermal performance. At the $T_{FAN}$ temperature, Intel recommends full cooling capability before the DTS reading reaches $Tj_{MAX}$.

### 5.1.4.3 PROCHOT# Signal

PROCHOT# (processor hot) is asserted by the processor when the TCC is active. Only a single PROCHOT# pin exists at a package level. When any DTS temperature reaches the TCC activation temperature, the PROCHOT# signal will be asserted. PROCHOT# assertion policies are independent of Adaptive Thermal Monitor enabling.

### 5.1.4.4 Bi-Directional PROCHOT#

By default, the PROCHOT# signal is set to input only. When configured as an input or bi-directional signal, PROCHOT# can be used for thermally protecting other platform components should they overheat as well. When PROCHOT# is driven by an external device:

- The package will immediately transition to the lowest P-State (Pn) supported by the processor IA cores. This is contrary to the internally-generated Adaptive Thermal Monitor response.
- Clock modulation is not activated.

The processor package will remain at the lowest supported P-state until the system de-asserts PROCHOT#. The processor can be configured to generate an interrupt upon assertion and de-assertion of the PROCHOT# signal.

When PROCHOT# is configured as a bi-directional signal and PROCHOT# is asserted by the processor, it is impossible for the processor to detect a system assertion of PROCHOT#. The system assertion will have to wait until the processor de-asserts PROCHOT# before PROCHOT# action can occur due to the system assertion. While the processor is hot and asserting PROCHOT#, the power is reduced but the reduction rate is slower than the system PROCHOT# response of < 100 us. The processor thermal control is staged in smaller increments over many milliseconds. This may cause several milliseconds of delay to a system assertion of PROCHOT# while the output function is asserted.

#### 5.1.4.5 Voltage Regulator Protection using PROCHOT#

PROCHOT# may be used for thermal protection of voltage regulators (VR). System designers can create a circuit to monitor the VR temperature and assert PROCHOT# and, if enabled, activate the TCC when the temperature limit of the VR is reached. When PROCHOT# is configured as a bi-directional or input only signal, if the system assertion of PROCHOT# is recognized by the processor, it will result in an immediate transition to the lowest P-State (Pn) supported by the processor IA cores. Systems should still provide proper cooling for the VR and rely on bi-directional PROCHOT# only as a backup in case of system cooling failure. Overall, the system thermal design should allow the power delivery circuitry to operate within its temperature specification even while the processor is operating at its TDP.

#### 5.1.4.6 Thermal Solution Design and PROCHOT# Behavior

With a properly designed and characterized thermal solution, it is anticipated that PROCHOT# will only be asserted for very short periods of time when running the most power intensive applications. The processor performance impact due to these brief periods of TCC activation is expected to be so minor that it would be immeasurable. However, an under-designed thermal solution that is not able to prevent excessive assertion of PROCHOT# in the anticipated ambient environment may:

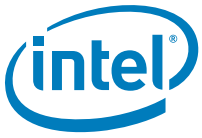
- Cause a noticeable performance loss.
- Result in prolonged operation at or above the specified maximum junction temperature and affect the long-term reliability of the processor.
- May be incapable of cooling the processor even when the TCC is active continuously (in extreme situations).

#### 5.1.4.7 Low-Power States and PROCHOT# Behavior

Depending on package power levels during package C-states, outbound PROCHOT# may de-assert while the processor is idle as power is removed from the signal. Upon wake up, if the processor is still hot, the PROCHOT# will re-assert. Although, typically package idle state residency should resolve any thermal issues. The PECI interface is fully operational during all C-states and it is expected that the platform continues to manage processor IA core and package thermals even during idle states by regularly polling for thermal data over PECI.

#### 5.1.4.8 THERMTRIP# Signal

Regardless of enabling the automatic or on-demand modes, in the event of a catastrophic cooling failure, the package will automatically shut down when the silicon has reached an elevated temperature that risks physical damage to the product. At this point, the THERMTRIP# signal will go active.



#### 5.1.4.9 Critical Temperature Detection

Critical Temperature detection is performed by monitoring the package temperature. This feature is intended for graceful shutdown before the THERMTRIP# is activated. However, the processor execution is not guaranteed between critical temperature and THERMTRIP#. If the Adaptive Thermal Monitor is triggered and the temperature remains high, a critical temperature status and sticky bit are latched in the PACKAGE_THERM_STATUS MSR 1B1h and the condition also generates a thermal interrupt, if enabled. For more details on the interrupt mechanism, refer to the *Intel® 64 and IA-32 Architectures Software Developer's Manual* (see Related Documents section).

#### 5.1.4.10 On-Demand Mode

The processor provides an auxiliary mechanism that allows system software to force the processor to reduce its power consumption using clock modulation. This mechanism is referred to as "On-Demand" mode and is distinct from Adaptive Thermal Monitor and bi-directional PROCHOT#. The processor platforms should not rely on software usage of this mechanism to limit the processor temperature. On-Demand Mode can be accomplished using processor MSR or chipset I/O emulation. On-Demand Mode may be used in conjunction with the Adaptive Thermal Monitor. However, if the system software tries to enable On-Demand mode at the same time the TCC is engaged, the factory configured duty cycle of the TCC will override the duty cycle selected by the On-Demand mode. If the I/O based and MSR-based On-Demand modes are in conflict, the duty cycle selected by the I/O emulation-based On-Demand mode will take precedence over the MSR-based On-Demand Mode.

#### 5.1.4.11 MSR Based On-Demand Mode

If Bit 4 of the IA32_CLOCK_MODULATION MSR is set to 1, the processor will immediately reduce its power consumption using modulation of the internal processor IA core clock, independent of the processor temperature. The duty cycle of the clock modulation is programmable using bits [3:1] of the same IA32_CLOCK_MODULATION MSR. In this mode, the duty cycle can be programmed in either 12.5% or 6.25% increments (discoverable using CPUID). Thermal throttling using this method will modulate each processor IA core's clock independently.

#### 5.1.4.12 I/O Emulation-Based On-Demand Mode

I/O emulation-based clock modulation provides legacy support for operating system software that initiates clock modulation through I/O writes to ACPI defined processor clock control registers on the chipset (PROC_CNT). Thermal throttling using this method will modulate all processor IA cores simultaneously.

### 5.1.5 Intel® Memory Thermal Management

The processor provides thermal protection for system memory by throttling memory traffic when using either DIMM modules or a memory down implementation. Two levels of throttling are supported by the processor, either a warm threshold or hot threshold that is customizable through memory mapped I/O registers. Throttling based on the warm threshold should be an intermediate level of throttling. Throttling based on the hot threshold should be the most severe. The amount of throttling is dynamically controlled by the processor.

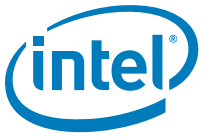
Memory temperature can be acquired through an on-board thermal sensor (TS-on-Board), retrieved by an embedded controller and reports to the processor through the PECI 3.1 interface. This methodology is known as PECI injected temperatures. This is a method of Closed Loop Thermal Management (CLTM). CLTM requires the use of a physical thermal sensor. EXTTS# is another method of CLTM but it is only capable of reporting memory thermal status to the processor. EXTTS# consists of two GPIO pins on the PCH, where the state of the pins is communicated internally to the processor.

When a physical thermal sensor is not available to report temperature, the processor supports Open Loop Thermal Management (OLTM) that estimates the power consumed per rank of the memory using the processor's DRAM power meter. A per rank power is associated with the warm and hot thresholds that, when exceeded, may trigger memory thermal throttling.

## 5.2 Thermal and Power Specifications

The following notes apply only to Table 5-1 and Table 5-2.

| Note | Definition |
|---|---|
| 1 | The TDP values are the average power dissipation in junction temperature operating condition limit, for the SKU Segment and Configuration, for which the processor is validated during manufacturing when executing an associated Intel-specified high-complexity workload at the processor IA core frequency corresponding to the configuration and SKU. |
| 2 | TDP workload may consist of a processor IA core intensive applications. |
| 3 | Can be modified at runtime by MSR writes, with MMIO and with PECI commands. |
| 4 | 'Turbo Time Parameter' is a mathematical parameter (units of seconds) that controls the processor turbo algorithm using a moving average of energy usage. Do not set the Turbo Time Parameter to a value less than 0.1 seconds. refer to Section 5.1.3.2 for further information. |
| 5 | Shown limit is a time averaged power, based upon the Turbo Time Parameter. Absolute product power may exceed the set limits for short durations or under virus or uncharacterized workloads. |
| 6 | Processor will be controlled to specified power limit as described in Section 5.1.2. If the power value and/or 'Turbo Time Parameter' is changed during runtime, it may take a short period of time (approximately 3 to 5 times the 'Turbo Time Parameter') for the algorithm to settle at the new control limits. |
| 7 | This is a hardware default setting and not a behavioral characteristic of the part. The reference BIOS code may override the hardware default power limit values to optimize performance |
| 8 | For controllable turbo workloads, the PL2 limit may be exceeded for up to 10 ms. |
| 9 | N/A |
| 10 | LPM power level is an opportunistic power and is not a guaranteed value as usages and implementations may vary. |
| 11 | N/A |
| 12 | N/A |
| 13 | N/A |
| 14 | May vary based on SKU. |
| 16 | Sustained residencies at high voltages and temperatures may temporarily limit turbo frequency. |



## 5.2.1 Processor Line Thermal and Power Specifications

### Table 5-1. TDP Specifications

| Segment and Package | Processor IA Cores, Graphics Configuration and TDP | Configuration | Processor IA Core Frequency | Graphics core Frequency | Thermal Design Power (TDP) [w] | Notes |
|---|---|---|---|---|---|---|
| X-Processor Line LGA | Quad Core GT0 112W | Base | 3.3 GHz | N/A | 112 | 1,9,10, 11, 12, 16 |
| | | LPM | 800 MHz | N/A | N/A | |

### Table 5-2. Package Turbo Specifications

| Processor IA Cores, Graphics Configuration and TDP | Parameter | Min. | Hardware Default | Max | Units | Notes |
|---|---|---|---|---|---|---|
| Quad Core GT0 112W | Power Limit 1 Time (PL1 Tau)<br>Power Limit 1 (PL1)<br>Power Limit 2 (PL2) | 0.1<br>N/A<br>N/A | 1<br>112<br>1.25*TDP | 8<br>N/A<br>N/A | s<br>W<br>W | 3, 4, 5, 6, 7, 8, 14 |

### Table 5-3. Low Power and TTV Specifications

| Processor IA Cores, Graphics Configuration and TDP | PCG[7] | Max Power Package C7 (W) [1,4,5] | Max Power Package C8 (W) [1,4,5] | TTV TDP (W) [6,7] | Min $T_{CASE}$ (°C) | Max TTV $T_{CASE}$ (°C) |
|---|---|---|---|---|---|---|
| Quad Core GT0 112W | 2017X | N/A | N/A | 112 | 0 | 54.8 |

***Notes:***
1. The package C-state power is the worst case power in the system configured as follows:
   a. Memory configured for DDR4 2400 and populated with two DIMMs per channel.
   b. DMI and PCIe links are at L1
2. Specification at DTS = 50 °C and minimum voltage loadline.
3. Specification at DTS = 35 °C and minimum voltage loadline.
4. These DTS values in Notes 2 - 3 are based on the TCC Activation MSR having a value of 100.
5. These values are specified at $V_{CC\_MAX}$ and $V_{NOM}$ for all other voltage rails for all processor frequencies. Systems should be designed to ensure the processor is not to be subjected to any static $V_{CC}$ and $I_{CC}$ combination wherein $V_{CCP}$ exceeds $V_{CCP\_MAX}$ at specified $I_{CCP}$. See the loadline specifications.
6. Thermal Design Power (TDP) should be used for processor thermal solution design targets. TDP is not the maximum power that the processor can dissipate. TDP is measured at DTS = -1.TDP is achieved with the Memory configured for DDR4 2400 and 2 DIMMs per channel.
7. Platform Compatibility Guide (PCG) (previously known as FMB) provides a design target for meeting all planned processor frequency requirements.
8. Not 100% tested. Specified by design characterization.

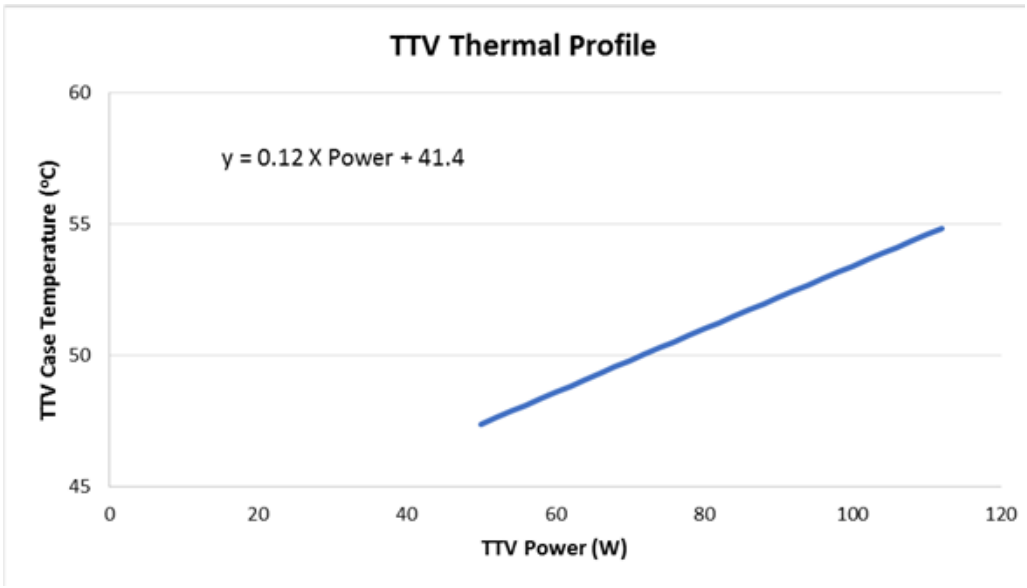
### Table 5-4. $T_{CONTROL}$ Offset Configuration

| Segment | TDP [W] | TEMP_TARGET (TCONTROL) [ºC] |
|---|---|---|
| X-Processor Line-Quad Core GT0 | 112 | 20 |

***Notes:***
1. Digital Thermal Sensor (DTS) based fan speed control is recommended to achieve optimal thermal performance.
2. Intel recommends full cooling capability at approximately the DTS value of -1, to minimize TCC activation risk.
3. For example, if $T_{CONTROL}$ = 20 ºC, Fan acceleration operation will start at 80 ºC (100 ºC - 20 ºC).

### 5.2.1.1 Thermal Profile for PCG 2017X Processor

**Figure 5-2. Thermal Test Vehicle Thermal Profile for PCG 2017X Processor**




***Notes:***
1. Refer to Table 5-5 for discrete points that constitute the thermal profile.

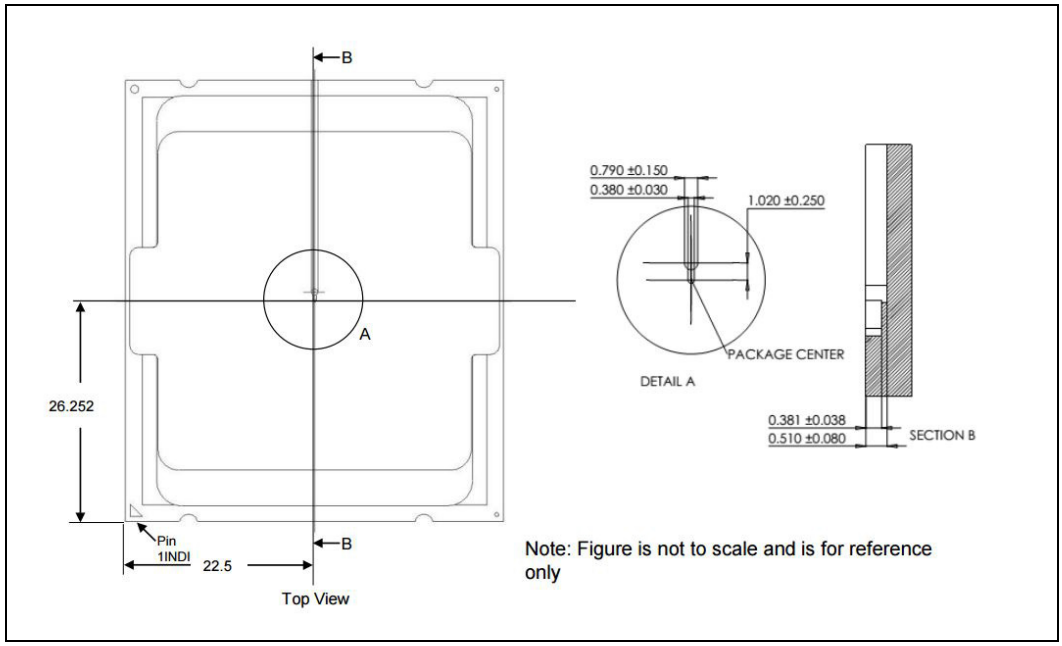
**Table 5-5. Thermal Test Vehicle Thermal Profile for PCG 2017X Processor**

| Power (W) | Tc (oC) | Power (W) | Tc (oC) |
|---|---|---|---|
| 50 | 47.4 | 82 | 51.2 |
| 52 | 47.6 | 84 | 51.5 |
| 54 | 47.9 | 86 | 51.7 |
| 56 | 48.1 | 88 | 52.0 |
| 58 | 48.4 | 90 | 52.2 |
| 60 | 48.6 | 92 | 52.4 |
| 62 | 48.8 | 94 | 52.7 |
| 64 | 49.1 | 96 | 52.9 |
| 66 | 49.3 | 98 | 53.2 |
| 68 | 49.6 | 100 | 53.4 |
| 70 | 49.8 | 102 | 53.6 |
| 72 | 50.0 | 104 | 53.9 |
| 74 | 50.3 | 106 | 54.1 |
| 76 | 50.5 | 108 | 54.4 |
| 78 | 50.8 | 110 | 54.6 |
| 80 | 51.0 | 112 | 54.8 |

### 5.2.1.2 Thermal Metrology

The maximum TTV case temperatures ($T_{CASE-MAX}$) can be derived from the data in the appropriate TTV thermal profile earlier in this chapter. The TTV $T_{CASE}$ is measured at the geometric top center of the TTV integrated heat spreader (IHS). Figure 5-3 illustrates the location where $T_{CASE}$ temperature measurements should be made.

**Figure 5-3. Thermal Test Vehicle (TTV) Case Temperature ($T_{CASE}$) Measurement Location**




The following supplier can machine the groove and attach a thermocouple to the IHS. The following supplier is listed as a convenience to Intel's general customers and may be subject to change without notice. THERM-X OF CALIFORNIA, 3200 Investment Blvd, Hayward, Ca 94544. George Landis +1-510-441-7566 Ext. 368 george@therm-x.com. The vendor part number is XTMS1565.

### 5.2.1.3 Fan Speed Control Scheme with Digital Thermal Sensor (DTS) 1.1

To correctly use DTS 1.1, the designer must first select a worst case scenario $T_{AMBIENT}$, and ensure that the Fan Speed Control (FSC) can provide a $\Psi_{CA}$ that is equivalent or greater than the $\Psi_{CA}$ specification.

The DTS 1.1 implementation consists of two points: a $\Psi_{CA}$ at $T_{CONTROL}$ and a $\Psi_{CA}$ at DTS = -1.

The $\Psi_{CA}$ point at DTS = -1 defines the minimum $\Psi_{CA}$ required at TDP considering the worst case system design $T_{AMBIENT}$ design point:

$$\Psi_{CA} = (T_{CASE-MAX} - T_{AMBIENT-TARGET}) / TDP$$

For example, for a 91 W TDP part, the $T_{CASE}$ maximum is 63.7 °C and at a worst case design point of 40 °C local ambient this will result in:

$$\Psi_{CA} = (63.7 - 40) / 91 = 0.26\ °C/W$$

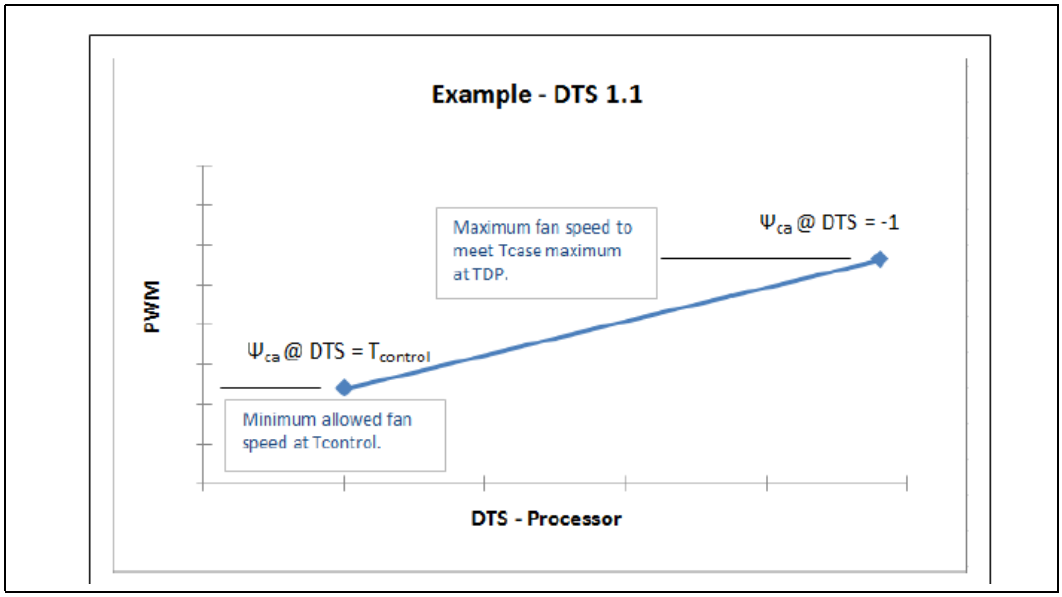
Similarly for a system with a design target of 45 °C ambient, the $\Psi_{CA}$ at DTS = -1 needed will be 0.21 °C/W.

The second point defines the thermal solution performance ($\Psi_{CA}$) at $T_{CONTROL}$.

These two points define the operational limits for the processor for DTS 1.1 implementation.

The fan speed controller must linearly ramp the fan speed from processor DTS = $T_{CONTROL}$ to processor DTS = -1.

**Figure 5-4. Digital Thermal Sensor (DTS) 1.1 Definition Points**



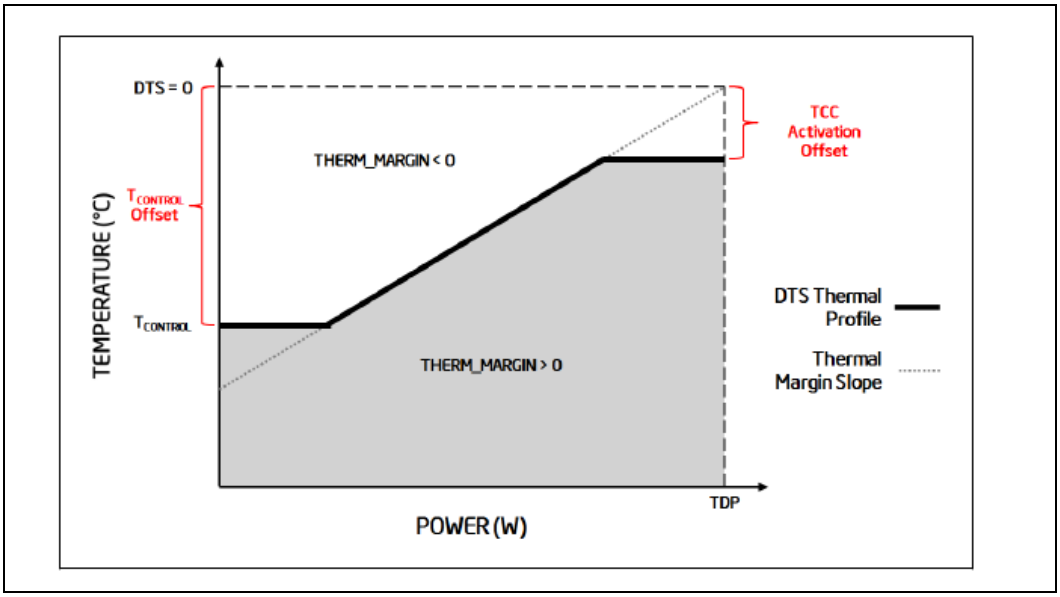

### 5.2.1.4 Fan Speed Control Scheme with Digital Thermal Sensor (DTS) 2.0

To simplify processor thermal specification compliance, the processor calculates the DTS Thermal Profile from $T_{CONTROL}$ Offset, TCC Activation Temperature, TDP, and the Thermal Margin Slope.

***Note:*** TCC Activation Offset is 0 for the processors.

Using the DTS Thermal Profile, the processor can calculate and report the Thermal Margin, where a value less than 0 indicates that the processor needs additional cooling, and a value greater than 0 indicates that the processor is sufficiently cooled.

## Figure 5-5. Digital Thermal Sensor (DTS) 1.1 Definition Points




§ §

# 6 Signal Description

This chapter describes the processor signals. They are arranged in functional groups according to their associated interface or category. The notations in the following table are used to describe the signal type.

The signal description also includes the type of buffer used for the particular signal (see the following table).

**Table 6-1. Signal Tables Terminology**

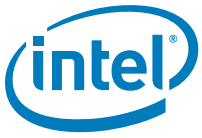
| Notation | Signal Type |
|---|---|
| I | Input pin |
| O | Output pin |
| I/O | Bi-directional Input/Output pin |
| SE | Single Ended Link |
| Diff | Differential Link |
| CMOS | CMOS buffers. 1.05V- tolerant |
| OD | Open Drain buffer |
| DDR4 | DDR4 buffers: 1.2V-tolerant |
| A | Analog reference or output. May be used as a threshold voltage or for buffer compensation |
| GTL | Gunning Transceiver Logic signaling technology |
| Ref | Voltage reference signal |
| Availability | Signal Availability condition - based on segment, SKU, platform type or any other factor |
| Asynchronous [1] | Signal has no timing relationship with any reference clock. |

***Note:***
1. Qualifier for a buffer type.

## 6.1 System Memory Interface

**Table 6-2. DDR4 Memory Interface (Sheet 1 of 2)**

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| DDR0_DQ[63:0]<br>DDR1_DQ[63:0] | **Data Buses:** Data signals interface to the SDRAM data buses. | I/O | DDR4 | SE | All Processor Lines |
| DDR0_DQSP[7:0]<br>DDR0_DQSN[7:0]<br>DDR1_DQSP[7:0]<br>DDR1_DQSN[7:0] | **Data Strobes:** Differential data strobe pairs. The data is captured at the crossing point of DQS during read and write transactions. | I/O | DDR4 | Diff | All Processor Lines |
| DDR0_CKN[3:0]<br>DDR0_CKP[3:0]<br>DDR1_CKN[3:0]<br>DDR1_CKP[3:0] | **SDRAM Differential Clock:** Differential clocks signal pairs, pair per rank. The crossing of the positive edge of DDR0_CKP/DDR1_CKP and the negative edge of their complement DDR0_CKN / DDR1_CKN are used to sample the command and control signals on the SDRAM. | O | DDR4 | Diff | All Processor Lines |



## Table 6-2. DDR4 Memory Interface (Sheet 2 of 2)

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| DDR0_CKE[3:0]<br>DDR1_CKE[3:0] | **Clock Enable:** (1 per rank). These signals are used to:<br>• Initialize the SDRAMs during power-up.<br>• Power-down SDRAM ranks.<br>• Place all SDRAM ranks into and out of self-refresh during STR (Suspend to RAM). | O | DDR4 | SE | All Processor Lines |
| DDR0_CS#[3:0]<br>DDR1_CS#[3:0] | **Chip Select:** (1 per rank). These signals are used to select particular SDRAM components during the active state. There is one Chip Select for each SDRAM rank. | O | DDR4 | SE | All Processor Lines |
| DDR0_ODT[3:0]<br>DDR1_ODT[3:0] | **On Die Termination:** (1 per rank). Active SDRAM Termination Control. | O | DDR4 | SE | All Processor Lines |
| DDR0_MA[16:0]<br>DDR1_MA[16:0] | **Address:** These signals are used to provide the multiplexed row and column address to the SDRAM.<br>• A[16:14] use also as command signals, see ACT# signal description.<br>• A10 is sampled during Read/Write commands to determine whether Autoprecharge should be performed to the accessed bank after the Read/Write operation.<br>HIGH: Autoprecharge;<br>LOW: no Autoprecharge).<br>• A10 is sampled during a Precharge command to determine whether the Precharge applies to one bank (A10 LOW) or all banks (A10 HIGH). If only one bank is to be precharged, the bank is selected by bank addresses.<br>• A12 is sampled during Read and Write commands to determine if burst chop (on-the-fly) will be performed.<br>HIGH, no burst chop;<br>LOW: burst chopped). | O | DDR4 | SE | All Processor Lines |
| DDR0_ACT#<br>DDR1_ACT# | **Activation Command:** ACT# HIGH, along with CS#, determine that the signals addresses below have command functionality.<br>A16: use as RAS# signal<br>A15: use as CAS# signal<br>A14: use as WE# signal | O | DDR4 | SE | All Processor Lines |
| DDR0_BG[1:0]<br>DDR1_BG[1:0] | **Bank Group:** BG[0:1] define which bank group an Active, Read, Write or Precharge command is being applied.<br>BG0 also determines which mode register is to be accessed during a MRS cycle. | O | DDR4 | SE | All processor lines<br>SO-DIMM, x8 DRAMs devices use BG[1:0]. |
| DDR0_BA[1:0]<br>DDR1_BA[1:0] | **Bank Address:** BA[1:0] define which bank an Active, Read, Write or Precharge command is being applied. Bank address also determines which mode register is to be accessed during a MRS cycle. | O | DDR4 | SE | All Processor Lines |
| DDR0_ALERT#<br>DDR1_ALERT# | **Alert:** This signal is used as command training only. It is getting the Command and Address Parity error flag during training. CRC feature is not supported. | I | DDR4 | SE | All Processor Lines |
| DDR0_PAR<br>DDR1_PAR | **Command and Address Parity:** These signals are used for parity check. | O | DDR4 | SE | All Processor Lines |
| DDR_VREF_CA | **Memory Reference Voltage for Command & Address:** | O | A | SE | All Processor Lines |

### Table 6-3. System Memory Reference and Compensation Signals

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| DDR_RCOMP[2:0] | **System Memory Resistance Compensation:** | N/A | A | SE | All Processor Lines |
| DDR_VTT_CNTL | **System Memory Power Gate Control:** When signal is high – Platform memory VTT regulator is enable, output high.<br>When signal is low - Disables the platform memory VTT regulator in C8 and deeper and S3. | O | CMOS | SE | All Processor Lines |

## 6.2 PCI Express* Graphics (PEG) Signals

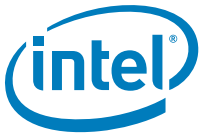
### Table 6-4. PCI Express* Interface

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| PEG_RCOMP | Resistance Compensation for PCI Express channels PEG and DMI. | N/A | A | SE | All Processor Lines |
| PEG_RXP[15:0]<br>PEG_RXN[15:0] | PCI Express Receive Differential Pairs. | I | PCI Express* | Diff | |
| PEG_TXP[15:0]<br>PEG_TXN[15:0] | PCI Express Transmit Differential Pairs. | O | PCI Express* | Diff | |

## 6.3 Direct Media Interface (DMI) Signals

### Table 6-5. DMI Interface Signals

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| DMI_RXP[3:0]<br>DMI_RXN[3:0] | **DMI Input from PCH:** Direct Media Interface receive differential pairs. | I | DMI | Diff | All Processor Lines |
| DMI_TXP[3:0]<br>DMI_TXN[3:0] | **DMI Output to PCH:** Direct Media Interface transmit differential pairs. | O | DMI | Diff | |



# 6.4 Reset and Miscellaneous Signals

## Table 6-6. Reset and Miscellaneous Signals

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| CFG[19:0] | **Configuration Signals:** The CFG signals have a default value of '1' if not terminated on the board.<br>Intel recommends placing test points on the board for CFG pins.<br>• **CFG[0]**: Stall reset sequence after PCU PLL lock until de-asserted:<br>— 1 = (Default) Normal Operation; No stall.<br>— 0 = Stall.<br>• **CFG[1]**: Reserved configuration lane.<br>• **CFG[2]**: PCI Express* Static x16 Lane Numbering Reversal.<br>— 1 = Normal operation<br>— 0 = Lane numbers reversed.<br>• **CFG[3]:** Reserved configuration lane.<br>• **CFG[4]:** eDP enable:<br>— 1 = Disabled.<br>— 0 = Enabled.<br>• **CFG[6:5]:** PCI Express* Bifurcation<br>— 00 = 1 x8, 2 x4 PCI Express*<br>— 01 = reserved<br>— 10 = 2 x8 PCI Express*<br>— 11 = 1 x16 PCI Express*<br>• **CFG[7]**: PEG Training:<br>— 1 = (default) PEG Train immediately following RESET# de assertion.<br>— 0 = PEG Wait for BIOS for training.<br>• **CFG[19:8]:** Reserved configuration lanes. | I | GTL | SE | All Processor Lines |
| CFG_RCOMP | **Configuration Resistance Compensation** | N/A | N/A | SE | All Processor Lines |
| RESET# | Platform Reset pin driven by the PCH. | I | CMOS | SE | All Processor Lines |
| PROC_SELECT# | **Processor Select:** This pin is for compatibility with future platforms. It should be unconnected for this processor. | | | N/A | All Processor Lines |
| PROC_TRIGIN | Debug pin | I | CMOS | SE | All Processor Lines |
| PROC_TRIGOUT | Debug pin | O | CMOS | SE | All Processor Lines |
| PROC_AUDIO_SDI | **Processor Audio Serial Data Input:** This signal is an input to the processor from the PCH. | I | AUD | SE | All Processor Lines |
| PROC_AUDIO_SDO | **Processor Audio Serial Data Output:** This signal is an output from the processor to the PCH. | O | AUD | SE | |
| PROC_AUDIO_CLK | **Processor Audio Clock** | I | AUD | SE | |

## 6.5 Processor Clocking Signals

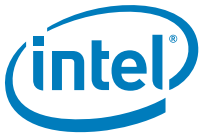
### Table 6-7. Processor Clocking Signals

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| BCLKP<br>BCLKN | 100 MHz Differential bus clock input to the processor | I | | Diff | All Processor Lines |
| CLK24P<br>CLK24N | 24 MHz Differential bus clock input to the processor | I | | Diff | |
| PCI_BCLKP<br>PCI_BCLKN | 100 MHz Clock for PCI Express* logic | I | | Diff | |

## 6.6 Testability Signals

### Table 6-8. Testability Signals

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| BPM#[3:0] | **Breakpoint and Performance Monitor Signals:** Outputs from the processor that indicate the status of breakpoints and programmable counters used for monitoring processor performance. | I/O | GTL | SE | All Processor Lines |
| PROC_PRDY# | **Probe Mode Ready:** PROC_PRDY# is a processor output used by debug tools to determine processor debug readiness. | O | OD | SE | All Processor Lines |
| PROC_PREQ# | **Probe Mode Request:** PROC_PREQ# is used by debug tools to request debug operation of the processor. | I | GTL | SE | All Processor Lines |
| PROC_TCK | **Test Clock:** This signal provides the clock input for the processor Test Bus (also known as the Test Access Port). This signal should be driven low or allowed to float during power on Reset. | I | GTL | SE | All Processor Lines |
| PROC_TDI | **Test Data In:** This signal transfers serial test data into the processor. This signal provides the serial input needed for JTAG specification support. | I | GTL | SE | All Processor Lines |
| PROC_TDO | **Test Data Out:** This signal transfers serial test data out of the processor. This signal provides the serial output needed for JTAG specification support. | O | OD | SE | All Processor Lines |
| PROC_TMS | **Test Mode Select:** A JTAG specification support signal used by debug tools. | I | GTL | SE | All Processor Lines |
| PROC_TRST# | **Test Reset:** Resets the Test Access Port (TAP) logic. This signal should be driven low during power on Reset. | I | GTL | SE | All Processor Lines |



## 6.7 Error and Thermal Protection Signals

### Table 6-9. Error and Thermal Protection Signals

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| CATERR# | **Catastrophic Error:** This signal indicates that the system has experienced a catastrophic error and cannot continue to operate. The processor will set this signal for non-recoverable machine check errors or other unrecoverable internal errors. CATERR# is used for signaling the following types of errors: Legacy MCERRs, CATERR# is asserted for 16 BCLKs. Legacy IERRs, CATERR# remains asserted until warm or cold reset. | O | OD | SE | All Processor Lines |
| PECI | **Platform Environment Control Interface:** A serial sideband interface to the processor. It is used primarily for thermal, power, and error management. Details regarding the PECI electrical specifications, protocols and functions can be found in the RS-Platform Environment Control Interface (PECI) Specification, Revision 3.0. | I/O | PECI, Async | SE | All Processor Lines |
| PROCHOT# | **Processor Hot:** PROCHOT# goes active when the processor temperature monitoring sensor(s) detects that the processor has reached its maximum safe operating temperature. This indicates that the processor Thermal Control Circuit (TCC) has been activated, if enabled. This signal can also be driven to the processor to activate the TCC. | I/O | GTL I OD O | SE | All Processor Lines |
| THERMTRIP# | **Thermal Trip:** The processor protects itself from catastrophic overheating by use of an internal thermal sensor. This sensor is set well above the normal operating temperature to ensure that there are no false trips. The processor will stop all executions when the junction temperature exceeds approximately 130 °C. This is signaled to the system by the THERMTRIP# pin. | O | OD | SE | All Processor Lines |

## 6.8 Power Sequencing Signals

### Table 6-10. Power Sequencing Signals (Sheet 1 of 2)

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| PROCPWRGD | **Processor Power Good:** The processor requires this input signal to be a clean indication that the $V_{CC}$ and $V_{DDQ}$ power supplies are stable and within specifications. This requirement applies regardless of the S-state of the processor. 'Clean' implies that the signal will remain low (capable of sinking leakage current), without glitches, from the time that the power supplies are turned on until they come within specification. The signal should then transition monotonically to a high state. | I | CMOS | SE | All Processor Lines |
| VCCST_PWRGD | **VCCST Power Good:** The processor requires this input signal to be a clean indication that the VCCST and VDDQ power supplies are stable and within specifications. This signal should have a valid level during both S0 and S3 power states. 'Clean' implies that the signal will remain low (capable of sinking leakage current), without glitches, from the time that the power supplies are turned on until they come within specification. The signal should then transition monotonically to a high state. | I | CMOS | SE | All Processor Lines |

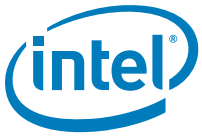
### Table 6-10. Power Sequencing Signals (Sheet 2 of 2)

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| PROC_DETECT# /SKTOCC# | **Processor Detect / Socket Occupied:** Pulled down directly (0 Ohms) on the processor package to the ground. There is no connection to the processor silicon for this signal. System board designers may use this signal to determine if the processor is present. | N/A | N/A | SE | All Processor Lines |
| VIDSOUT<br>VIDSCK<br>VIDALERT# | **VIDSOUT**, **VIDSCK, VIDALERT#:** These signals comprise a three-signal serial synchronous interface used to transfer power management information between the processor and the voltage regulator controllers. | I/O<br>O<br>I | I:GTL/O:OD<br>OD<br>CMOS | SE | All Processor Lines |
| PM_SYNC | **Power Management Sync:** A sideband signal to communicate power management status from the PCH to the processor. The PCH report EXTTS#/EVENT# status to the processor. | I | CMOS | SE | All Processor Lines |
| PM_DOWN | **Power Management Down:** Sideband to PCH. Indicates processor wake up event EXTTS# on PCH. The processor combines the pin status into the OLTM/CLTM. | O | CMOS | SE | All Processor Lines |

## 6.9 Processor Power Rails

### Table 6-11. Processor Power Rails Signals

| Signal Name | Description | Dir. | Buffer Type | Link Type | Availability |
|---|---|---|---|---|---|
| Vcc | Processor IA cores power rail | I | Power | — | All Processor Lines |
| $V_{DDQ}$ | System Memory power rail | I | Power | — | All Processor Lines |
| $Vcc_{SA}$ | Processor System Agent power rail | I | Power | — | All Processor Lines |
| $Vcc_{IO}$ | Processor I/O power rail. Consists of $V_{CCIO}$ and $Vcc_{IO\_DDR}$. $V_{CCIO}$ and $V_{CCIO\_DDR}$ should be isolated from each other. | I | Power | — | All Processor Lines |
| $Vcc_{ST}$ | Sustain voltage for processor standby modes | I | Power | — | All Processor Lines |
| $Vcc_{PLL}$ | Processor PLLs power rails | I | Power | — | All Processor Lines |
| $Vcc_{PLL\_OC}$ | Processor PLLs power rails | I | Power | — | All Processor Lines |
| Vcc_SENSE<br>Vss_SENSE | Isolated, low impedance voltage sense pins. They can be used to sense or measure voltage near the silicon. | N/A | Power | — | All Processor Lines |
| $Vcc_{IO}$_SENSE<br>$Vss_{IO}$_SENSE | Isolated, low impedance voltage sense pins. They can be used to sense or measure voltage near the silicon. | N/A | Power | — | All Processor Lines |
| $Vcc_{SA}$_SENSE<br>$Vss_{SA}$_SENSE | Isolated, low impedance voltage sense pins. They can be used to sense or measure voltage near the silicon. | N/A | Power | — | All Processor Lines |



## 6.10 Ground, Reserved and Non-Critical to Function (NCTF) Signals

The following are the general types of reserved (RSVD) signals and connection guidelines:

- RSVD – these signals should not be connected
- RSVD_TP – these signals should be routed to a test point
- RSVD_NCTF – these signals are non-critical to function and may be left un-connected

Arbitrary connection of these signals to VCC, VDDQ, VSS, or to any other signal (including each other) may result in component malfunction or incompatibility with future processors. See Table 6-12, "GND, RSVD, and NCTF Signals".

For reliable operation, always connect unused inputs or bi-directional signals to an appropriate signal level. Unused active high inputs should be connected through a resistor to ground ($V_{SS}$). Unused outputs may be left unconnected however, this may interfere with some Test Access Port (TAP) functions, complicate debug probing and prevent boundary scan testing. A resistor should be used when tying bi-directional signals to power or ground. When tying any signal to power or ground, the resistor can also be used for system testability.

**Table 6-12. GND, RSVD, and NCTF Signals**

| Signal Name | Description |
|---|---|
| Vss | Processor ground node |
| Vss_NCTF | **Non-Critical To Function:** These signals are for package mechanical reliability. |
| RSVD<br>RSVD_NCTF<br>RSVD_TP | **Reserved:** All signals that are RSVD and RSVD_NCTF should be left unconnected on the board.<br>Intel recommends that all RSVD_TP signals have via test points. |

## 6.11 Processor Internal Pull-Up / Pull-Down Terminations

**Table 6-13. Processor Internal Pull-Up / Pull-Down Terminations**

| Signal Name | Pull Up/Pull Down | Rail | Value |
|---|---|---|---|
| BPM[3:0] | Pull Up / Pull Down | $Vcc_{IO}$ | 16-60 ohms |
| PREQ# | Pull Up | $Vcc_{ST}$ | 3 kohms |
| PROC_TDI | Pull Up | $Vcc_{ST}$ | 3 kohms |
| PROC_TMS | Pull Up | $Vcc_{ST}$ | 3 kohms |
| PROC_TRSN# | Pull Down | - | 3 kohms |
| CFG[19:0] | Pull Up | $Vcc_{IO}$ | 3 kohms |

§ §

# 7 Electrical Specifications

## 7.1 Processor Power Rails

**Table 7-1. Processor Power Rails**

| Power Rail | Description | Control | Availability |
|---|---|---|---|
| $V_{CC}$ | Processor IA Cores Power Rail | SVID | All Processor Lines |
| $Vcc_{SA}$ | System Agent Power Rail | SVID/Fixed (SKU dependent) | All Processor Lines |
| $Vcc_{IO}$ | IO Power Rail | Fixed | All Processor Lines |
| $Vcc_{ST}$ | Sustain Power Rail | Fixed | All Processor Lines |
| $Vcc_{PLL}$ | Processor PLLs power Rail | Fixed | All Processor Lines |
| $Vcc_{PLL\_OC}$ [1] | Processor PLLs OC power Rail | Fixed | All Processor Lines |
| $V_{DDQ}$ | Integrated Memory Controller Power Rail | Fixed (Memory technology dependent) | All Processor Lines |

***Notes:***
1. $Vcc_{PLL\_OC}$ power rail should be sourced from the VDDQ VR. The connection can be direct or through a load switch, depending desired power optimization. In case of direct connection ($Vcc_{PLL\_OC}$ is shorted to $V_{DDQ}$, no load switch), platform should ensure that $Vcc_{ST}$ is ON (high) while $Vcc_{PLL\_OC}$ is ON (high).

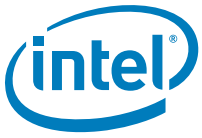
### 7.1.1 Power and Ground Pins

All power pins should be connected to their respective processor power planes, while all VSS pins should be connected to the system ground plane. Use of multiple power and ground planes is recommended to reduce I*R drop.

### 7.1.2 $V_{CC}$ Voltage Identification (VID)

The processor uses three signals for the **S**erial **V**oltage **ID**entification (SVID) interface to support automatic selection of voltages. The following table specifies the voltage level corresponding to the 8-bit VID value transmitted over serial VID. A '1' in this table refers to a high voltage level and a '0' refers to a low voltage level. If the voltage regulation circuit cannot supply the voltage that is requested, the voltage regulator should disable itself. VID signals are CMOS push/pull drivers. See Table 7-11 for the DC specifications for these signals. The VID codes will change due to temperature and/or current load changes in order to minimize the power of the part. A voltage range is provided in Section 7.2. The specifications are set so that one voltage regulator can operate with all supported frequencies.

Individual processor VID values may be set during manufacturing so that two devices at the same processor IA core frequency may have different default VID settings. This is shown in the VID range values in Section 7.2. The processor provides the ability to operate while transitionally to an adjacent VID and its associated voltage. This will represent a DC shift in the loadline.



## 7.2 DC Specifications

The processor DC specifications in this section are defined at the processor signal pins, unless noted otherwise.

- The DC specifications for the DDR4 signals are listed in the *Voltage and Current Specifications* section.
- The *Voltage and Current Specifications* section lists the DC specifications for the processor and are valid only while meeting specifications for junction temperature, clock frequency, and input voltages. Read all notes associated with each parameter.
- AC tolerances for all DC rails include dynamic load currents at switching frequencies up to 1 MHz.

### 7.2.1 Processor Power Rails DC Specifications

#### 7.2.1.1 Vcc DC Specifications

**Table 7-2. Processor IA core (Vcc) Active and Idle Mode DC Voltage and Current Specifications (Sheet 1 of 2)**

| Symbol | Parameter | Segment | Min | Typ | Max | | | Unit | Note[1] |
|---|---|---|---|---|---|---|---|---|---|
| Operating Voltage | Voltage Range for Processor Operating Modes | All | 0 | — | 1.52 | | | V | 1,2,3, 7,12 |
| $Icc_{MAX}$ | Maximum Processor IA Core $I_{CC}$ | X-Processor Line (112W) - Quad Core GT0 | — | — | 115 | | | A | 4, 6, 7, 11 |
| $TOB_{VCC}$ | Voltage Tolerance | PS0, PS1 | — | — | ±20 | | | mV | 3, 6, 8 |
| | | PS2, PS3 | — | — | ±20 | | | | |
| Ripple | Ripple Tolerance | | | | $I_L$<=0.5 | 0.5<$I_L$<$Icc_{TDC}$ | $Icc_{TDC}$<$I_L$<$Icc_{MAX}$ | mV | 3, 6, 8 |
| | | PS0 | — | — | +30/-10 | ±10 | ±15 | | |
| | | PS1 | — | — | +30/-10 | ±15 | ±15 | | |
| | | PS2 | — | — | +30/-10 | +30/-10 | +30/-10 | | |
| | | PS3 | — | — | +30/-10 | +30/-10 | +30/-10 | | |
| DC_LL | Loadline slope within the VR regulation loop capability | X-Processor Line - Quad Core GT0 | — | — | 1.0 | | | mΩ | 10,13,14 |
| AC_LL | AC Loadline | X-Processor Line | — | — | 1.1 (Above 1 KHz) | | | mΩ | 10,13,14 |
| T_OVS_TDP_MAX | Max Overshoot time **TDP/virus mode** | — | — | — | 10/30 | | | µs | |
| V_OVS TDP_MAX/ virus_MAX | Max Overshoot at **TDP/virus mode** | — | — | — | 70/200 | | | mV | |

### Table 7-2. Processor IA core (Vcc) Active and Idle Mode DC Voltage and Current Specifications (Sheet 2 of 2)

| Symbol | Parameter | Segment | Min | Typ | Max | Unit | Note[1] |
|---|---|---|---|---|---|---|---|

***Notes:***
1. Unless otherwise noted, all specifications in this table are based on estimates and simulations or empirical data. These specifications will be updated with characterized data from silicon measurements at a later date.
2. Each processor is programmed with a maximum valid voltage identification value (VID) that is set at manufacturing and cannot be altered. Individual maximum VID values are calibrated during manufacturing such that two processors at the same frequency may have different settings within the VID range. Note that this differs from the VID employed by the processor during a power management event (Adaptive Thermal Monitor, Enhanced Intel SpeedStep Technology, or low-power states).
3. The voltage specification requirements are measured across Vcc_SENSE and Vss_SENSE as near as possible to the processor with an oscilloscope set to 100-MHz bandwidth, 1.5 pF maximum probe capacitance, and 1 MΩ minimum impedance. The maximum length of ground wire on the probe should be less than 5 mm. Ensure external noise from the system is not coupled into the oscilloscope probe.
4. Processor IA core VR to be designed to electrically support this current.
5. Processor IA core VR to be designed to thermally support this current indefinitely.
6. Long term reliability cannot be assured if tolerance, ripple, and core noise parameters are violated.
7. Long term reliability cannot be assured in conditions above or below Max/Min functional limits.
8. PSx refers to the voltage regulator power state as set by the SVID protocol.
9. N/A
10. LL measured at sense points.
11. Typ column represents $Icc_{MAX}$ for commercial application it is NOT a specification - it is a characterization of limited samples using limited set of benchmarks that can be exceeded.
12. Operating voltage range in steady state.
13. LL specification values should not be exceeded. If exceeded, power, performance and reliability penalty are expected.
14. By Improving Load Line (Lower LL than Datasheet values, and reporting it to BIOS), customers may obtain slightly better performance although the frequencies will not be changed.

## 7.2.1.2 $V_{DDQ}$ DC Specifications

### Table 7-3. Memory Controller ($V_{DDQ}$) Supply DC Voltage and Current Specifications

| Symbol | Parameter | Segment | Min | Typ | Max | Unit | Note[1] |
|---|---|---|---|---|---|---|---|
| $V_{DDQ\ (DDR4)}$ | Processor I/O supply voltage for DDR4 | All | Typ-5% | 1.20 | Typ+5% | V | 3, 4, 5 |
| $TOB_{VDDQ}$ | VDDQ Tolerance | All | AC+DC: ± 5 | | | % | 3, 4, 6 |
| $Icc_{MAX\_VDDQ\ (DDR4)}$ | Max Current for $V_{DDQ}$ Rail (DDR4) | X | — | — | 2.8 | A | 2 |

***Notes:***
1. Unless otherwise noted, all specifications in this table are based on estimates and simulations or empirical data. These specifications will be updated with characterized data from silicon measurements at a later date.
2. The current supplied to the DIMM modules is not included in this specification.
3. Includes AC and DC error, where the AC noise is bandwidth limited to under 100 MHz, measured on package pins.
4. No requirement on the breakdown of AC versus DC noise.
5. The voltage specification requirements are measured as near as possible to the processor with an oscilloscope set to 100-MHz bandwidth, 1.5 pF maximum probe capacitance, and 1 MΩ minimum impedance. The maximum length of ground wire on the probe should be less than 5 mm. Ensure external noise from the system is not coupled into the oscilloscope probe.
6. For Voltage less than 1v, TOB will be 50 mv.

### 7.2.1.3 $Vcc_{SA}$ DC Specifications

**Table 7-4. System Agent ($Vcc_{SA}$) Supply DC Voltage and Current Specifications**

| Symbol | Parameter | Segment | Min | Typ | Max | Unit | Note[1,2] |
|---|---|---|---|---|---|---|---|
| $Vcc_{SA}$ | Voltage for the System Agent | X-Processor Line (fixed voltage) | — | 1.05 | — | V | 3,5 |
| $I_{CCMAX\_VCCSA}$ | Max Current for $V_{CCSA}$ Rail | X-Processor Lines | — | — | 11.1 | A | |
| T_OVS_MAX | Max Overshoot time | — | — | — | 10 | µs | |
| V_OVS_MAX | Max Overshoot | — | — | — | 70 | mV | |

***Notes:***
1. Unless otherwise noted, all specifications in this table are based on estimates and simulations or empirical data. These specifications will be updated with characterized data from silicon measurements at a later date.
2. Long term reliability cannot be assured in conditions above or below Max/Min functional limits.
3. The voltage specification requirements are measured across $Vcc_{SA\_SENSE}$ and $Vss_{SA\_SENSE}$ as near as possible to the processor with an oscilloscope set to 100-MHz bandwidth, 1.5 pF maximum probe capacitance, and 1 MΩ minimum impedance. The maximum length of ground wire on the probe should be less than 5 mm. Ensure external noise from the system is not coupled into the oscilloscope probe.
4. PSx refers to the voltage regulator power state as set by the SVID protocol.
5. $Vcc_{SA}$ voltage during boot (Vboot)1.05V for a duration of 2 seconds.
6. LL measured at sense points.
7. LL specification values should not be exceeded. If exceeded, power, performance and reliability penalty are expected.
8. N/A.
9. N/A.
10. For Voltage less than 1v, TOB will be 50 mv.

### 7.2.1.4 $Vcc_{IO}$ DC Specifications

**Table 7-5. Processor I/O ($Vcc_{IO}$) Supply DC Voltage and Current Specifications**

| Symbol | Parameter | Segment | Min | Typ | Max | Unit | Note[1,2] |
|---|---|---|---|---|---|---|---|
| $Vcc_{IO}$ | Voltage for the memory controller and shared cache | X | — | 1.00 | — | V | 3, 4, 5, 6 |
| $TOB_{VCCIO}$ | $Vcc_{IO}$ Tolerance | All | AC+DC: ± 5 | | | % | 3, 8 |
| $Icc_{MAX\_VCCIO}$ | Max Current for $V_{CCIO}$ Rail | X | — | — | 5.5 | A | |
| T_OVS_MAX | Max Overshoot time | All | — | — | 100 | µS | 7 |
| V_OVS_MAX | Max Overshoot at TDP | All | — | — | 20 | mV | 7 |

***Notes:***
1. Unless otherwise noted, all specifications in this table are based on estimates and simulations or empirical data. These specifications will be updated with characterized data from silicon measurements at a later date.
2. Long term reliability cannot be assured in conditions above or below Max/Min functional limits.
3. The voltage specification requirements are measured across $Vcc_{IO\_SENSE}$ and $Vss_{IO\_SENSE}$ as near as possible to the processor with an oscilloscope set to 100-MHz bandwidth, 1.5 pF maximum probe capacitance, and 1 MΩ minimum impedance. The maximum length of ground wire on the probe should be less than 5 mm. Ensure external noise from the system is not coupled into the oscilloscope probe.
4. For low BW bus connection between processor and PCH -> $Vcc_{IO}$=0.85V.
5. For high BW bus connection between processor and PCH -> $Vcc_{IO}$=0.95V.
6. N/A
7. OS occurs during power on only, **not** during normal operation.
8. For Voltage less than 1v, TOB will be 50 mv.

### 7.2.1.5 $Vcc_{ST}$ DC Specifications

**Table 7-6. Vcc Sustain ($Vcc_{ST}$) Supply DC Voltage and Current Specifications (Sheet 1 of**

| Symbol | Parameter | Segment | Min | Typ | Max | Units | Notes [1,2] |
|---|---|---|---|---|---|---|---|
| $Vcc_{ST}$ | Processor Vcc Sustain supply voltage | All | — | 1.0 | — | V | 3 |

### Table 7-6. Vcc Sustain ($Vcc_{ST}$) Supply DC Voltage and Current Specifications (Sheet 2 of

| Symbol | Parameter | Segment | Min | Typ | Max | Units | Notes [1,2] |
|---|---|---|---|---|---|---|---|
| $TOB_{ST}$ | $Vcc_{ST}$ Tolerance | All | AC+DC: ± 5 | | | mV | 3, 4 |
| $Icc_{MAX\_ST}$ | Max Current for $Vcc_{ST}$ | X-Processor Lines | — | — | 60 | mA | |

***Notes:***
1. Unless otherwise noted, all specifications in this table are based on estimates and simulations or empirical data. These specifications will be updated with characterized data from silicon measurements at a later date.
2. Long term reliability cannot be assured in conditions above or below Max/Min functional limits.
3. The voltage specification requirements are measured on package pins as near as possible to the processor with an oscilloscope set to 100-MHz bandwidth, 1.5 pF maximum probe capacitance, and 1 MΩ minimum impedance. The maximum length of ground wire on the probe should be less than 5 mm. Ensure external noise from the system is not coupled into the oscilloscope probe.
4. For Voltage less than 1v, TOB will be 50 mv.

## 7.2.1.6 $Vcc_{PLL}$ DC Specifications

### Table 7-7. Processor PLL ($Vcc_{PLL}$) Supply DC Voltage and Current Specifications

| Symbol | Parameter | Segment | Min | Typ | Max | Unit | Notes[1,2] |
|---|---|---|---|---|---|---|---|
| $Vcc_{PLL}$ | PLL supply voltage (DC + AC specification) | All | — | 1.0 | — | V | 3 |
| $TOB_{CCPLL}$ | $Vcc_{PLL}$ Tolerance | All | AC+DC: ± 5 | | | % | 3,4 |
| $Icc_{MAX\_VCCPLL}$ | Max Current for $Vcc_{PLL}$ Rail | X-Processor Lines | — | — | 150 | mA | |

***Notes:***
1. Unless otherwise noted, all specifications in this table are based on estimates and simulations or empirical data. These specifications will be updated with characterized data from silicon measurements at a later date.
2. Long term reliability cannot be assured in conditions above or below Max/Min functional limits.
3. The voltage specification requirements are measured on package pins as near as possible to the processor with an oscilloscope set to 100-MHz bandwidth, 1.5 pF maximum probe capacitance, and 1 MΩ minimum impedance. The maximum length of ground wire on the probe should be less than 5 mm. Ensure external noise from the system is not coupled into the oscilloscope probe.
4. For Voltage less than 1v, TOB will be 50 mv.

### Table 7-8. Processor PLL_OC ($Vcc_{PLL\_OC}$) Supply DC Voltage and Current Specifications

| Symbol | Parameter | Segment | Min | Typ | Max | Unit | Notes[1,2] |
|---|---|---|---|---|---|---|---|
| $Vcc_{PLL\_OC}$ | PLL_OC supply voltage (DC + AC specification) | All | — | $V_{DDQ}$ | — | V | 3 |
| $TOB_{CCPLL\_OC}$ | $Vcc_{PLL\_OC}$ Tolerance | All | AC+DC: ± 5 | | | % | 3,4 |
| $Icc_{MAX\_VCCPLL\_OC}$ | Max Current for $Vcc_{PLL\_OC}$ Rail | X-Processor Line - Quad Core GT0 | — | — | 130 | mA | |

***Notes:***
1. Unless otherwise noted, all specifications in this table are based on estimates and simulations or empirical data. These specifications will be updated with characterized data from silicon measurements at a later date.
2. Long term reliability cannot be assured in conditions above or below Max/Min functional limits.
3. The voltage specification requirements are measured on package pins as near as possible to the processor with an oscilloscope set to 100-MHz bandwidth, 1.5 pF maximum probe capacitance, and 1 MΩ minimum impedance. The maximum length of ground wire on the probe should be less than 5 mm. Ensure external noise from the system is not coupled into the oscilloscope probe.
4. For Voltage less than 1v, TOB will be 50 mv.

## 7.2.2 Processor Interfaces DC Specifications

### 7.2.2.1 DDR4 DC Specifications

**Table 7-9. DDR4 Signal Group DC Specifications**

| Symbol | Parameter | Min | Typ | Max | Units | Notes[1] |
|---|---|---|---|---|---|---|
| $V_{IL}$ | Input Low Voltage | — | — | VREF(INT) - 0.07*VDDQ | V | 2, 4, 8, 9, 13 |
| $V_{IH}$ | Input High Voltage | VREF(INT) + 0.07*VDDQ | — | — | V | 3, 4, 8, 9, 13 |
| $R_{ON\_UP/DN(DQ)}$ | DDR4 Data Buffer pull-up/ down Resistance | Trainable | | | Ω | 11 |
| $R_{ODT(DQ)}$ | DDR4 On-die termination equivalent resistance for data signals | Trainable | | | Ω | 11 |
| $V_{ODT(DC)}$ | DDR4 On-die termination DC working point (driver set to receive mode) | 0.45*$V_{DDQ}$ | 0.5*$V_{DDQ}$ | 0.55*$V_{DDQ}$ | V | 9 |
| $R_{ON\_UP/DN(CK)}$ | DDR4 Clock Buffer pull-up/ down Resistance | 0.8*Typ | 26 | 1.2*Typ | Ω | 5, 11 |
| $R_{ON\_UP/DN(CMD)}$ | DDR4 Command Buffer pull-up/ down Resistance | 0.8*Typ | 20 | 1.2*Typ | Ω | 11 |
| $R_{ON\_UP/DN(CTL)}$ | DDR4 Control Buffer pull-up/ down Resistance | 0.8*Typ | 20 | 1.2*Typ | Ω | 5, 11 |
| $R_{ON\_UP/DN}$ (DDR_VTT_CNTL) | System Memory Power Gate Control Buffer Pull-Up/ down Resistance | 40 | — | 140 | Ω | - |
| $I_{LI}$ | Input Leakage Current (DQ, CK)<br>0 V<br>0.2*$V_{DDQ}$<br>0.8*$V_{DDQ}$ | — | — | 1 | mA | - |
| DDR0_VREF_DQ<br>DDR1_VREF_DQ<br>DDR_VREF_CA | VREF output voltage | Trainable | $V_{DDQ}$/2 | Trainable | V | 12,14 |
| DDR_RCOMP[0] | ODT resistance compensation | RCOMP values are memory topology dependent. | | | Ω | 6 |
| DDR_RCOMP[1] | Data resistance compensation | | | | Ω | 6 |
| DDR_RCOMP[2] | Command resistance compensation | | | | Ω | 6 |

***Notes:***
1. Unless otherwise noted, all specifications in this table apply to all processor frequencies.
2. $V_{IL}$ is defined as the maximum voltage level at a receiving agent that will be interpreted as a logical low value.
3. $V_{IH}$ is defined as the minimum voltage level at a receiving agent that will be interpreted as a logical high value.
4. $V_{IH}$ and $V_{IL}$ may experience excursions above $V_{DDQ}$. However, input signal drivers should comply with the signal quality specifications.
5. This is the pull up/down driver resistance after compensation. Note that BIOS power training may change these values significantly based on margin/power trade-off.
6. N/A
7. DDR_VREF is defined as $V_{DDQ}$/2 for DDR4
8. $R_{ON}$ tolerance is preliminary and might be subject to change.
9. The value will be set during the MRC boot training within the specified range.
10. Processor may be damaged if $V_{IH}$ exceeds the maximum voltage for extended periods.
11. Final value determined by BIOS power training, values might vary between bytes and/or units.
12. VREF values determined by BIOS training, values might vary between units.
13. VREF(INT) is a trainable parameter whose value is determined by BIOS for margin optimization.
14. DDR0_Vref_DQ - Not in use in DDR4, DDR1_Vref_DQ = DDR4_CA_ch1, DDR_Vref_CA = DD4_CA_ch0

### 7.2.2.2 PCI Express* Graphics (PEG) DC Specifications

**Table 7-10. PCI Express* Graphics (PEG) Group DC Specifications**

| Symbol | Parameter | Min | Typ | Max | Units | Notes[1] |
|---|---|---|---|---|---|---|
| $Z_{TX-DIFF-DC}$ | DC Differential Tx Impedance | 80 | 100 | 120 | Ω | 1, 5 |
| $Z_{RX-DC}$ | DC Common Mode Rx Impedance | 40 | 50 | 60 | Ω | 1, 4 |
| $Z_{RX-DIFF-DC}$ | DC Differential Rx Impedance | 80 | — | 120 | Ω | 1 |
| PEG_RCOMP | resistance compensation | 24.75 | 25 | 25.25 | Ω | 2, 3 |

***Notes:***
1. Refer to the PCI Express Base Specification for more details.
2. Low impedance defined during signaling. Parameter is captured for 5.0 GHz by RLTX-DIFF.
3. PEG_RCOMP resistance should be provided on the system board with 1% resistors. COMP resistors are to $V_{CCIO}$. PEG_RCOMP - Intel allows using 24.9 Ω 1% resistors.
4. DC impedance limits are needed to ensure Receiver detect.
5. The Rx DC Common Mode Impedance should be present when the Receiver terminations are first enabled to ensure that the Receiver Detect occurs properly.Compensation of this impedance can start immediately and the 15 Rx Common Mode Impedance (constrained by RLRX-CM to 50 Ω ±20%) should be within the specified range by the time Detect is entered.

### 7.2.2.3 CMOS DC Specifications

**Table 7-11. CMOS Signal Group DC Specifications**

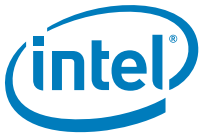
| Symbol | Parameter | Min | Max | Units | Notes[1] |
|---|---|---|---|---|---|
| $V_{IL}$ | Input Low Voltage | — | Vcc * 0.3 | V | 2 |
| $V_{IH}$ | Input High Voltage | Vcc * 0.7 | — | V | 2, 4 |
| $V_{OL}$ | Output Low Voltage | — | Vcc * 0.1 | V | 2 |
| $V_{OH}$ | Output High Voltage | Vcc * 0.9 | — | V | 2, 4 |
| $R_{ON}$ | Buffer on Resistance | 23 | 73 | Ω | - |
| $I_{LI}$ | Input Leakage Current | — | ±150 | μA | 3 |

***Notes:***
1. Unless otherwise noted, all specifications in this table apply to all processor frequencies.
2. The Vcc referred to in these specifications refers to instantaneous Vcc levels.
3. For VIN between "0" V and Vcc Measured when the driver is tri-stated.
4. $V_{IH}$ and $V_{OH}$ may experience excursions above Vcc. However, input signal drivers should comply with the signal quality specifications.

### 7.2.2.4 GTL and OD DC Specifications

**Table 7-12. GTL Signal Group and Open Drain Signal Group DC Specifications (Sheet 1 of 2)**

| Symbol | Parameter | Min | Max | Units | Notes[1] |
|---|---|---|---|---|---|
| $V_{IL}$ | Input Low Voltage (TAP, except PROC_TCK, PROC_TRST#) | — | Vcc * 0.6 | V | 2, 5, 6 |
| $V_{IH}$ | Input High Voltage (TAP, except PROC_TCK, PROC_TRST#) | Vcc * 0.72 | — | V | 2, 4, 5, 6 |
| $V_{IL}$ | Input Low Voltage (PROC_TCK,PROC_TRST#) | — | Vcc * 0.3 | V | 2, 5, 6 |
| $V_{IH}$ | Input High Voltage (PROC_TCK,PROC_TRST#) | Vcc * 0.3 | — | V | 2, 4, 5, 6 |
| $V_{HYSTERESIS}$ | Hysteresis Voltage | Vcc * 0.2 | — | V | - |
| $R_{ON}$ | Buffer on Resistance (TDO) | 7 | 17 | Ω | - |
| $V_{IL}$ | Input Low Voltage (other GTL) | — | Vcc * 0.6 | V | 2, 5, 6 |



### Table 7-12. GTL Signal Group and Open Drain Signal Group DC Specifications (Sheet 2 of 2)

| Symbol | Parameter | Min | Max | Units | Notes[1] |
|---|---|---|---|---|---|
| $V_{IH}$ | Input High Voltage (other GTL) | Vcc * 0.72 | — | V | 2, 4, 5 |
| $R_{ON}$ | Buffer on Resistance (CFG/BPM) | 16 | 24 | Ω | - |
| $R_{ON}$ | Buffer on Resistance (other GTL) | 12 | 28 | Ω | - |
| $I_{LI}$ | Input Leakage Current | — | ±150 | µA | 3 |

***Notes:***
1. Unless otherwise noted, all specifications in this table apply to all processor frequencies.
2. The $Vcc_{ST}$ referred to in these specifications refers to instantaneous $Vcc_{ST/IO}$.
3. For $V_{IN}$ between 0 V and $Vcc_{ST}$. Measured when the driver is tri-stated.
4. $V_{IH}$ and $V_{OH}$ may experience excursions above $Vcc_{ST}$. However, input signal drivers should comply with the signal quality specifications.
5. Those $V_{IL}/V_{IH}$ values are based on ODT disabled (ODT Pull-up not exist).

## 7.2.2.5 PECI DC Characteristics

The PECI interface operates at a nominal voltage set by $Vcc_{ST}$. The set of DC electrical specifications shown in the following table is used with devices normally operating from a $Vcc_{ST}$ interface supply.

$Vcc_{ST}$ nominal levels will vary between processor families. All PECI devices will operate at the $Vcc_{ST}$ level determined by the processor installed in the system.

### Table 7-13. PECI DC Electrical Limits

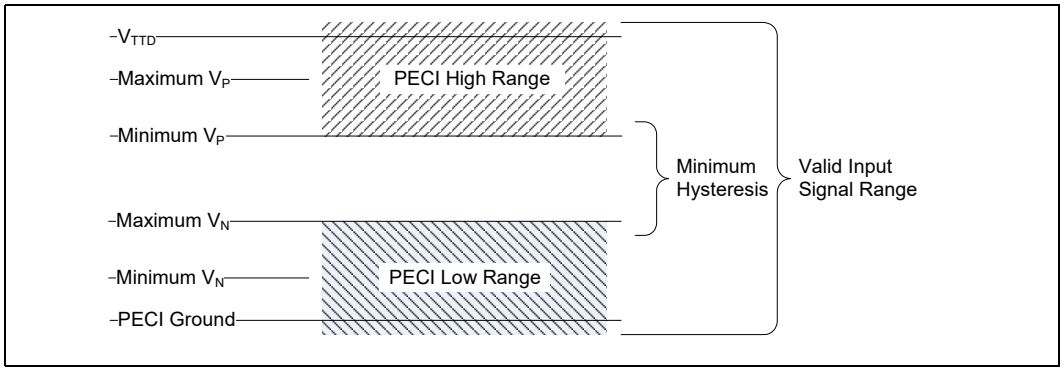
| Symbol | Definition and Conditions | Min | Max | Units | Notes[1] |
|---|---|---|---|---|---|
| $R_{up}$ | Internal pull up resistance | 15 | 45 | Ω | 3 |
| $V_{IN}$ | Input Voltage Range | -0.15 | $Vcc_{ST}$ + 0.15 | V | - |
| $V_{Hysteresis}$ | Hysteresis | 0.15 * $Vcc_{ST}$ | — | V | - |
| $V_{IL}$ | Input Voltage Low- Edge Threshold Voltage | — | 0.3 * $Vcc_{ST}$ | V | - |
| $V_{IH}$ | Input Voltage High-Edge Threshold Voltage | 0.7 * $Vcc_{ST}$ | — | V | - |
| $C_{bus}$ | Bus Capacitance per Node | N/A | 10 | pF | - |
| $C_{pad}$ | Pad Capacitance | 0.7 | 1.8 | pF | - |
| Ileak000 | leakage current @ 0V | — | 0.6 | mA | - |
| Ileak025 | leakage current @ 0.25* $Vcc_{ST}$ | — | 0.4 | mA | - |
| Ileak050 | leakage current @ 0.50* $Vcc_{ST}$ | — | 0.2 | mA | - |
| Ileak075 | leakage current @ 0.75* $Vcc_{ST}$ | — | 0.13 | mA | - |
| Ileak100 | leakage current @ $Vcc_{ST}$ | — | 0.10 | mA | - |

***Notes:***
1. $Vcc_{ST}$ supplies the PECI interface. PECI behavior does not affect $Vcc_{ST}$ min/max specifications.
2. The leakage specification applies to powered devices on the PECI bus.
3. The PECI buffer internal pull up resistance measured at 0.75* $Vcc_{ST}$.

### Input Device Hysteresis

The input buffers in both client and host models should use a Schmitt-triggered input design for improved noise immunity. Use the following figure as a guide for input buffer design.

**Figure 7-1. Input Device Hysteresis**




§ §

# 8 Package Mechanical Specifications

## 8.1 Package Mechanical Attributes

The X-Processor uses a Flip Chip technology available in Land Grid Array (LGA). The following table provides an overview of the mechanical attributes of the package.

**Table 8-1. Package Mechanical Attributes**

| Package | Parameter | X-Processor Line Quad Core GT0 |
|---|---|---|
| Package Technology | Package Type | Flip Chip Land Grid Array |
| | Interconnect | Land Grid Array (LGA) |
| | Lead Free | N/A |
| | Halogenated Flame Retardant Free | Yes |
| Package Configuration | Solder Ball Composition | N/A |
| | Ball/Pin Count | 2066 |
| | Grid Array Pattern | Grid Array |
| | Land Side Capacitors | Yes |
| | Die Side Capacitors | Yes |
| | Die Configuration | 1 Die<br>Single-Chip Package with IHS |
| Package Dimensions | Nominal Package Size | 58.5x51 mm |
| | Min Ball/Pin pitch | 1.016 mm |

## 8.2 Package Storage Specifications

**Table 8-2. Package Storage Specifications (Sheet 1 of 2)**

| Parameter | Description | Min | Max | Notes |
|---|---|---|---|---|
| $T_{ABSOLUTE\ STORAGE}$ | The non-operating device storage temperature. Damage (latent or otherwise) may occur when subjected to this temperature for any length of time in Intel Original sealed moisture barrier bag. | -25 °C | 125 °C | 1, 2, 3 |
| $T_{SUSTAINED\ STORAGE}$ | The ambient storage temperature limit (in shipping media) for the sustained period of time as specified below in Intel Original sealed moisture barrier bag. | -5 °C | 40 °C | 1, 2, 3 |
| $RH_{SUSTAINED\ STORAGE}$ | The maximum device storage relative humidity for the sustained period of time as specified below in Intel Original sealed moisture barrier bag. | 60% @ 24 °C | | 1, 2, 3 |
| $TIME_{SUSTAINED\ STORAGE}$ | A prolonged or extended period of time: associated with customer shelf life in Intel Original sealed moisture barrier bag. | 0 months | 6 months | 1, 2, 3 |

## Table 8-2. Package Storage Specifications (Sheet 2 of 2)

| Parameter | Description | Min | Max | Notes |
|---|---|---|---|---|
| ***Notes:***<br>1. $T_{ABSOLUTE\ STORAGE}$ applies to the un-assembled component only and does not apply to the shipping media, moisture barrier bags or desiccant. Refers to a component device that is not assembled in a board or socket that is not to be electrically connected to a voltage reference or I/O signals.<br>2. Specified temperatures are based on data collected. Exceptions for surface mount re-flow are specified by applicable JEDEC J-STD-020 and MAS documents. The JEDEC, J-STD-020 moisture level rating and associated handling practices apply to all moisture sensitive devices removed from the moisture barrier bag.<br>3. Post board attach storage temperature limits are not specified. Consult your board manufacturer for storage specifications. | | | | |

§ §