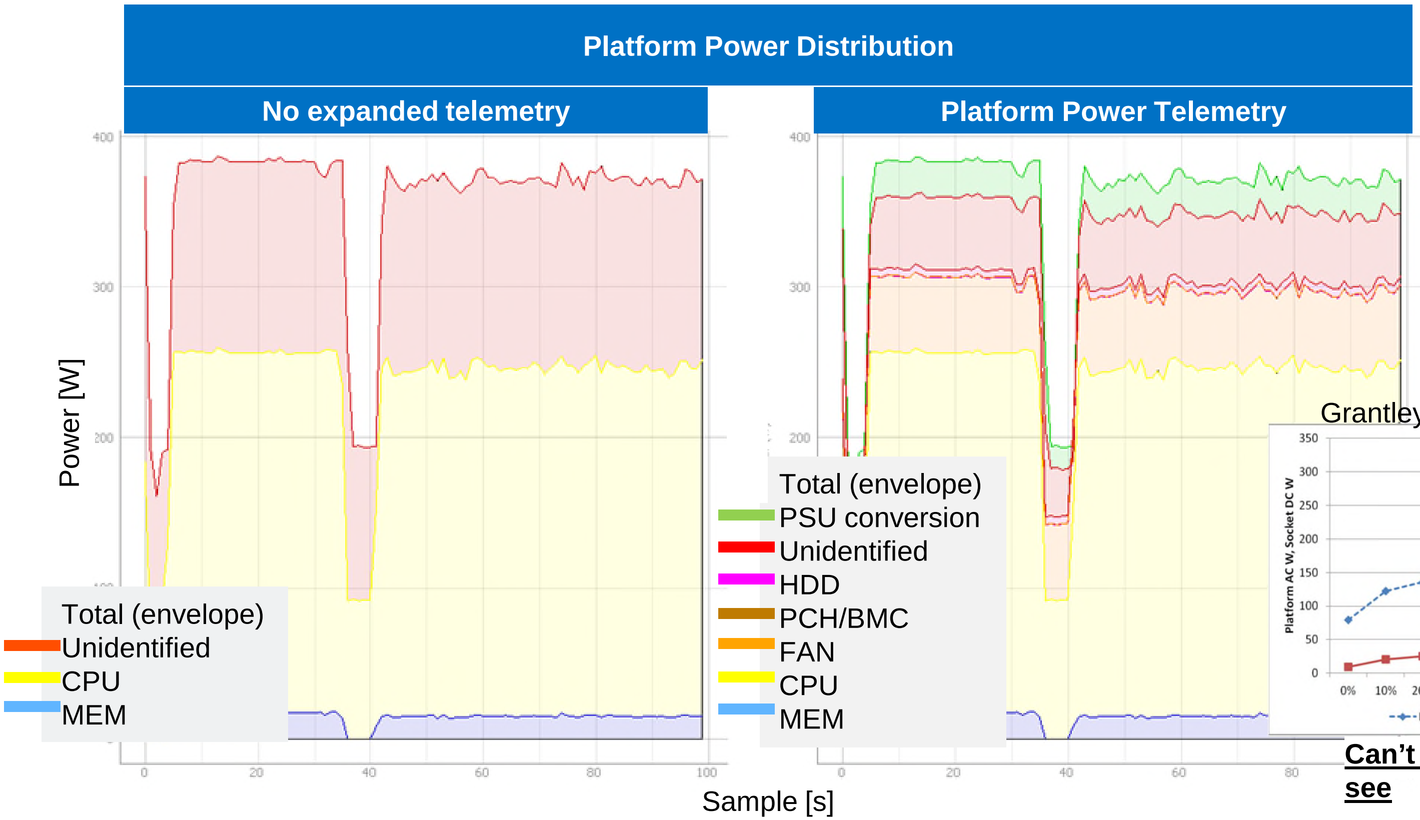
Server Telemetry

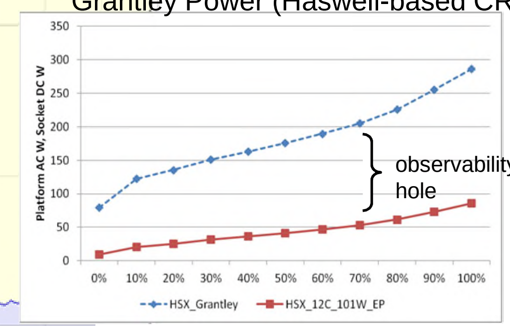
Insight into platform power & performance with Intel® Node Manager and PMBus® standard

March, 2018

So why does it matter? - observability hole



Grantley Power (Haswell-based CRB)



<u>Can't optimize what we can't see</u>

Server Board sensor map

Sensor Map

PSU

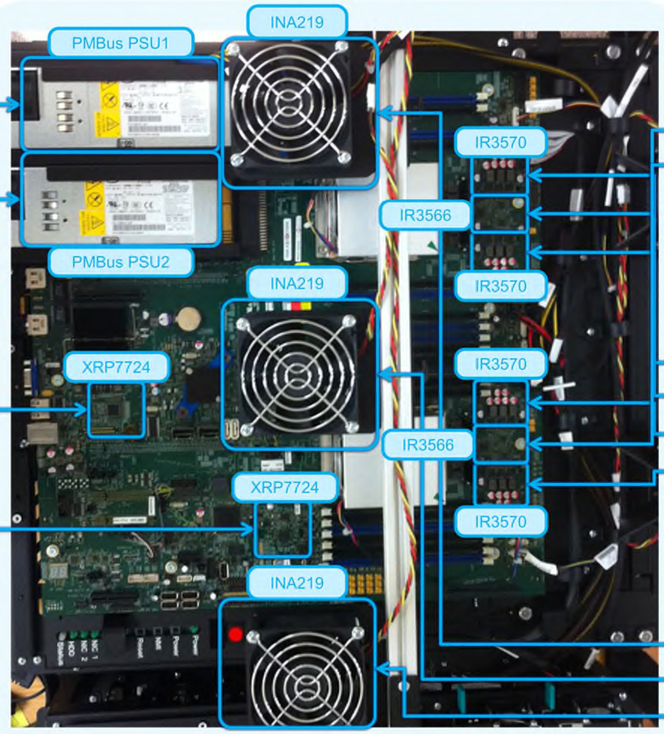
Devices	Address
PMBus PSU1	0xB0
PMBus PSU2	0xB2

LAN/PCH VR

Devices	Sensor	Address
LAN-Aux (4 channels)	XRP7724	0x52
PCH (3 channels), BMC	XRP7724	0x54

Auxiliary

IR3897 (Aux 5V)	APE8901 (Aux 1V)	APE8901 (Aux 1.8V)
IR3843 (I/O 1V)	IR3876 (3.3V)	APE8901 (Aux 1.5V)
IR3876 (5V)	IR3550 (CPU)	IR3537 (CPU)
IR3550 (MEM)	IR3533 (MEM)	LCMX0256 CPLD/FPGA



CPU VR 12.5

Devices	Sensor	Address
CPU #0	IR3566	0x8A
CPU #1	IR3566	0x96

MEM VR 12.0

Devices	Sensor	Address
MEM #1	IR3570	0x8E
MEM #2	IR3570	0x92
MEM #3	IR3570	0x9A
MEM #4	IR3570	0x9E

FAN IMON

Devices	Sensor	Address
Sysfan 1&2	INA219	0x80
Sysfan 3&4	INA219	0x82
Sysfan 5&6	INA219	0x88

Server Telemetry Using Intel® Node Manager

Power Telemetry

- Total platform power
- Individual CPU\, Memory and Xeon Phi power domains

Thermal Telemetry

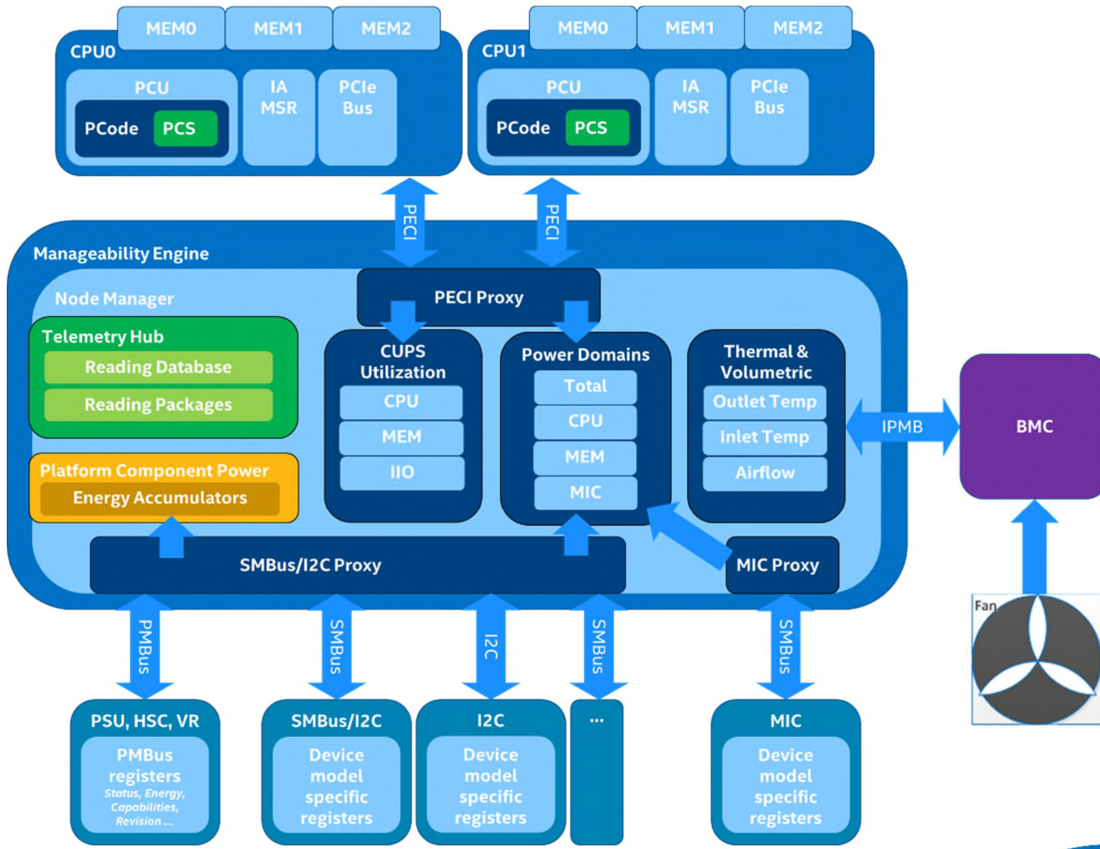
- Inlet & Outlet Airflow temperature
- Volumetric Airflow

Utilization Telemetry

- Aggregate Compute Utilization per sec
- CPU, Memory and I/O Utilization Metrics

Component Telemetry

- On-board telemetry devices monitoring
- Synchronously polled and stored into energy accumulators



PMBus® and i^2c/SMBus

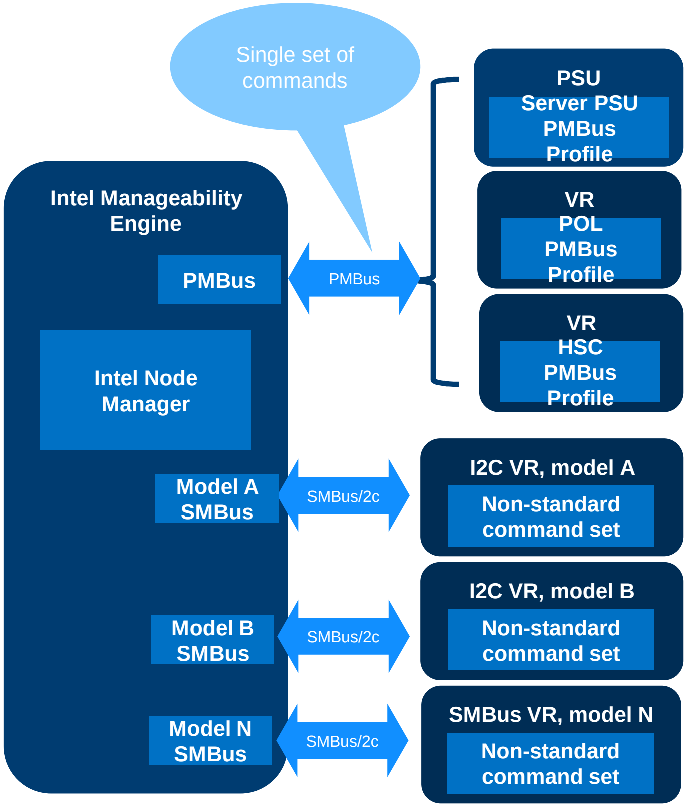
PMBus® forms the basis of several platform power management standards used in the Computing industry.

Various types of sensors and devices from different vendors monitored on server board.

- Each SMBus/i^2c device model has its own, unique set of commands
- Each requires custom, hard-coded support that drives validation and development costs
- Adopting a new device is costly, increases code size and risk

PMBus ® standard:

- **AC/DC Server PSU PMBus® Profile (2008)** that defined PSU behavior and has been used for years (since 2007) and adopted by all major PSU vendors and Intel
- **Hot Swap Controller PMBus® Profile (2010)** used in servers unified how hot-swap controllers are produced and deployed
- Unifies and standardizes the server environment
- Profile support simplifies device capability identification and ensures compliance with set of requirements



PMBus® and i^{2}c/SMBus

Single image of code instantiates on many varieties of platforms using device discovery protocol based on PMBus® Hot-Swap Controller Profile

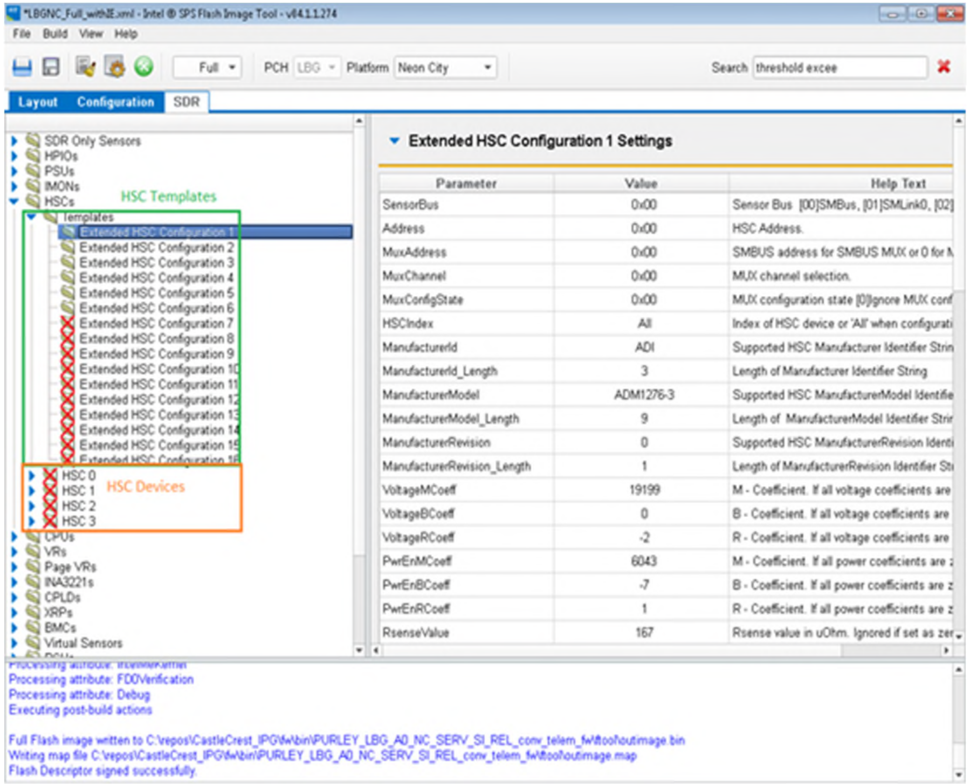
Intel® Node Manager based on superset of potential board configurations dynamically discovers present device topology of the platform

This supports hot-plug, hot-removal of board components

Implements best-match algorithm for PMBus® devices using PMBus® DEVICE, MANUFACTURER, REVISION and a range of address

Configuration supports wild cards, e.g. any manufacturer, any device revision, etc.

Physical device is instantiated dynamically based on matching HSC template and result of discovery



Use cases for Telemetry

Cloud scheduling algorithms:

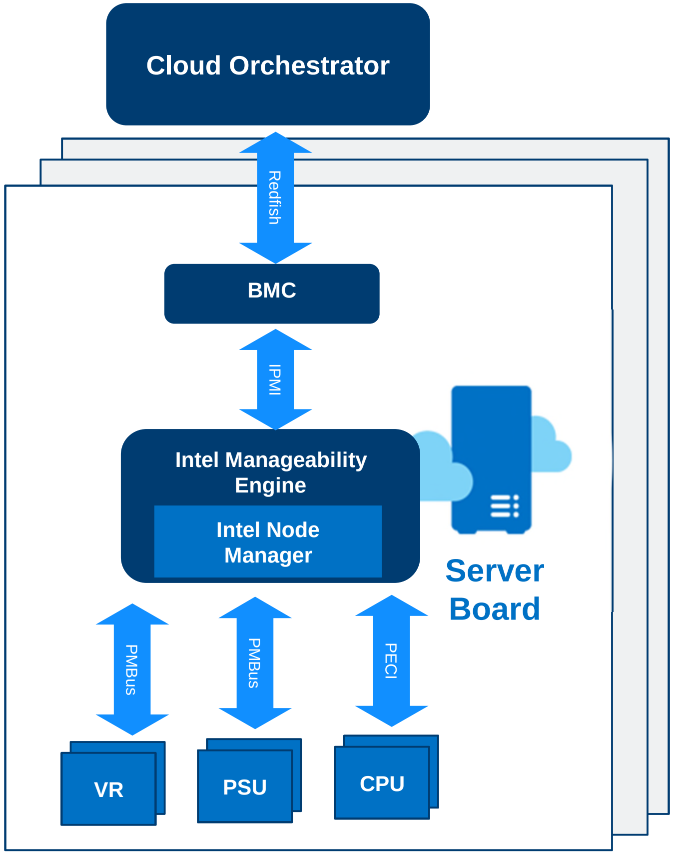
- Efficient workload placement
- Cooling optimization
- Building heatmap of Data Center
- Optimizing PUE

Improve observability of the platform board

- Maximizing rack density
- System sensor calibration and fine-tuning
- Run-time fault monitoring
- Power monitoring for peak, average power, etc.

Future of telemetry in Data Center

- Amount of telemetry points collected and processed is always increasing
- Telemetry is used to optimize and control workloads on Node-level, Rack-level and Data Center-level
- Node-level telemetry is aggregated using Telemetry Aggregators, e.g. Node Manager, BMC, Host Agent
- Orchestrator uses Redfish, RESTful API, to subscribe for telemetry
- Switch from polling to Publish/Subscribe model of delivering telemetry to Cloud Orchestrator
- Telemetry capabilities in external machine-readable schema file in XML, JSON



End Message