(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
30 June 2022 (30.06.2022)

WIPO | PCT

(10) International Publication Number
WO 2022/133643 A1

(51) International Patent Classification:
G06F 11/20 (2006.01) *G06F 1/30* (2006.01)

(21) International Application Number:
PCT/CN2020/137955

(22) International Filing Date:
21 December 2020 (21.12.2020)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant: **INTEL CORPORATION** [US/US]; 2200 Mission College Blvd., Santa Clara, CA 95054 (US).

(72) Inventors; and

(71) Applicants *(for BZ only)*: **MATSUMURA, Naoki** [JP/US]; 5286 Gatewood Lane, San Jose, CA 95118 (US). **LIANG, Xiaoguo** [CN/CN]; No. 880 Zixing Road, Zizhu Science Park, Shanghai 200241 (CN). **SONG, Chuan** [CN/CN]; Room 1302, #7 Building, #650 Laifang Road, Songjiang, Sahnghai 201615 (CN). **JIANG, Feng** [CN/CN]; No. 880 Zixing Road, Minghang, Shanghai 200241 (CN). **AHUJA, Nishi** [US/US]; 10117 NW Langworthy Terrace, Portland, Oregon 97229 (US). **YAN, Jie** [CN/CN]; No. 880 Zixing Road, Minghang District, Shanghai 210000 (CN). **ZHOU, Alex** [CN/CN]; Room 201, No. 57, Leng 425, Lianhua Road, Shanghai 200100 (CN).

(74) Agent: **NTD PATENT AND TRADEMARK AGENCY LTD.**; 10th Floor, Tower C, Beijing Global Trade Center, 36 North Third Ring Road East, Dongcheng District, Beijing 100013 (CN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO,

(54) Title: PEAK POWER SHAVING WITH MULTIPLE BATTERIES, AND CONTROL MECHANISM FOR HYBRID ENERGY SOURCES

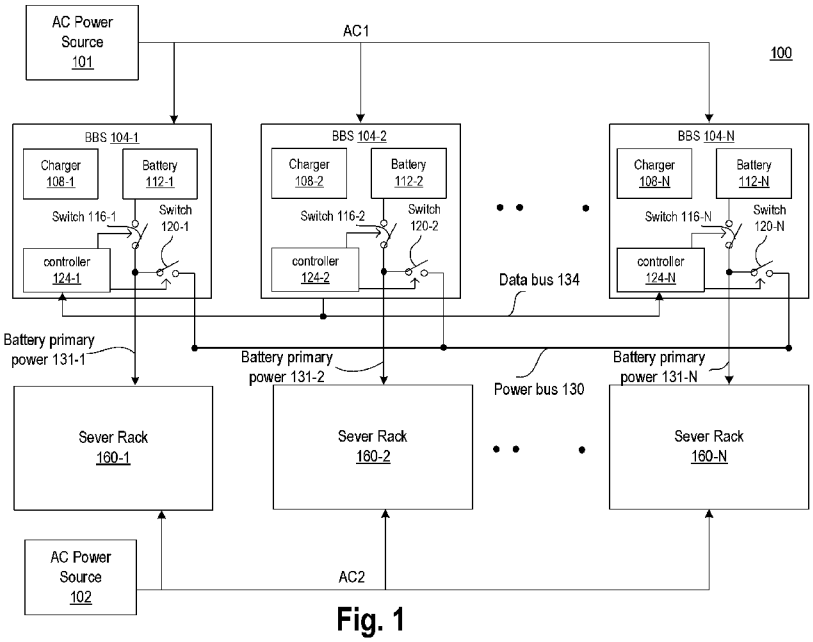


Fig. 1

(57) Abstract: A battery system is provided that during normal operation if for any reason AC power cannot fulfill the power demand of a power sink (e.g., servers in a rack), the battery backup units (BUs) may supplement extra peak power of the rack. As such, more sink devices (e.g., servers) can be added in a rack because BUs can step in to satisfy any peak power demand. The batteries or storage devices of the BUs include hybrid technologies. For example, a mix of high energy density charge (e.g., Li-Ion batteries) carrying device (s) and low energy density charge carrying devices (e.g., supercapacitor) are used as batteries or storage devices. A controller (e.g., hardware, software, or a combination of them) is used analyze peak power patterns and minimizes the impact to overall lifespan of high energy density charge carrying device-based battery cells while maximizing service time for peak power shaving feature.

WO 2022/133643 A1

[Continued on next page]

DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

PEAK POWER SHAVING WITH MULTIPLE BATTERIES, AND CONTROL MECHANISM FOR HYBRID ENERGY SOURCES

BACKGROUND

[0001] Data centers are one example of a use case where a high priority is placed on the reliable delivery of power among multiple servers, racks, clusters and/or other power sink devices. In such use cases, backup systems typically facilitate at least temporary delivery of battery power when a primary power supply is interrupted. As server farms and other such network resources continue to proliferate in number, size, and capability, there is expected to be an increasing demand for always-on operation of such resources.

BRIEF DESCRIPTION OF THE DRAWINGS

[0002] The embodiments of the disclosure will be understood more fully from the detailed description given below and from the accompanying drawings of various embodiments of the disclosure, which, however, should not be taken to limit the disclosure to the specific embodiments, but are for explanation and understanding only.

[0003] **Fig. 1** illustrates an apparatus comprising distributed backup battery system (BBS) with peak power sharing capability, in accordance with some embodiments.

[0004] **Fig. 2** illustrates a system comprising plurality of server racks, each having a BBS with peak power sharing capability, in accordance with some embodiments.

[0005] **Fig. 3** illustrates a flowchart of a method for peak power sharing using distributed BBS, in accordance with some embodiments.

[0006] **Fig. 4** illustrates an apparatus comprising a controller for managing hybrid energy sources to supply peak power demand, in accordance with some embodiments.

[0007] **Fig. 5** illustrates a plot showing peak power scenarios.

[0008] **Fig. 6** illustrates a flowchart of a method for managing hybrid energy sources to supply peak power demand, in accordance with some embodiments.

[0009] **Fig. 7** illustrates a plot showing one case of managing hybrid energy source, in accordance with some embodiments.

[0010] **Fig. 8** illustrates a smart device or a computer system or a SoC (System-on-Chip) with a BBS with peak power sharing capability and/or hybrid energy sources to supply peak power demand, in accordance with some embodiments.

DETAILED DESCRIPTION

[0011] Embodiments described herein variously provide redundancy between batteries which are to provide backup power for multiple power sinks. In some embodiments, a device comprises multiple batteries, which (for example) supply DC power in case of an AC power failure. For example, a first battery is coupled to supply power to a first one or more devices, a second battery is coupled to supply power to a second one or more devices, a third battery is coupled to supply power to a third one or more devices, and so on. The first one or more devices may be, merely as an example, first one or more servers installed in a first server rack of a data center; the second one or more devices may be, merely as an example, second one or more servers installed in a second server rack of the data center, and so on, although the teachings of this disclosure are not limited by such examples.

[0012] In some embodiments, backup units (BUs) of such a device each include a respective one of the batteries – e.g., wherein control circuitry of the BUs communicate, for example, over a communication bus. For example, a given BU broadcasts information regarding its operational state over the communication bus. A BU may periodically (or intermittently) broadcast that it is operating as intended. In an example, a lack of such a broadcast may indicate that the BU has failed. In another example, the broadcast may provide indication of a failed state of the BU. Thus, the BUs in the BU cluster are aware of any BU within the cluster that is not operating as intended.

[0013] In various embodiments, the BUs are coupled to a power bus through corresponding switches. Normally, when the AC power is out and the BUs of the cluster are operating as intended, the power bus is not powered. In such a situation, the first BU supplies power to the first one or more devices, the second BU supplies power to the second one or more devices, and the third BU supplies power to the third one or more devices.

[0014] However, there may arise a situation when the AC power has failed and one of the BUs (e.g., the first BU) of the cluster is also out of service. In such a situation, the second BU continues supplying power to the second one or more devices, and the third BU continues supplying power to the third one or more devices. Additionally, one or both the second or the third BU (or one or more other BUs of the cluster) supply power to the power bus. The first one or more devices now receives power from the power bus. Thus, addition of the power bus enables redundancy to the BU cluster, where in case of a BU failure, other operational BU(s) can take over the operation of the failed BU. The system is readily scalable, as any number of BUs can be coupled to the power bus.

[0015] In some embodiments, an apparatus is provided which comprises a plurality of battery systems (e.g., backup battery units) including a first battery system and a second battery system. In some embodiments, each of the first or second battery systems include: a battery comprising one or more battery cells; a charger to charge the one or more battery cells; a first switch to couple the battery to the first load; a second switch to couple the power bus to the first load; and a controller to control the first and second switches. In some embodiments, a plurality of loads is coupled to the plurality of battery systems, wherein the plurality of loads includes a first load and a second load. The apparatus further includes a power source to power the plurality of loads. Another power source may provide charging power to the plurality of batteries.

[0016] In some embodiments, the apparatus includes a power bus switchably coupled to the plurality of battery systems, wherein the first battery system provides backup power to the first load. The second battery system provides backup power to the second load. In various embodiments, the first and/or second battery systems supply power via the power bus to the first load during power demand by the first load. As such, battery systems are able to supplement power for a load that is demanding more power for workloads. The same battery systems are configured to provide backup power in the absence of outage of AC power to the loads.

[0017] In some embodiments, first and/or second battery system are discharged to a low threshold (e.g., 20% of full capacity or even cut off level) to supply power during the power demand. In some embodiments, the apparatus includes a data bus coupled to the plurality of battery systems. The data bus is used to carry telemetry information from various battery systems. The data bus is also used to send control signals that can enable or disable switches of the battery systems. In some embodiments, the controller monitors power consumption of the first and second loads, and monitors real-time capacity of the first and second battery systems. In some embodiments, the controller provides higher charging priority to one of the first or second battery systems based on workload priority of the first or second loads. In some embodiments, the battery cells include a high energy density storage bank (e.g., Li-ion battery cells) and a low energy density storage bank (e.g., super capacitor(s)).

[0018] In some embodiments, the controller enables the low energy density storage bank to provide power to the first load when the first load demands a sudden power (e.g., power spikes from milliseconds to tens of seconds). In some embodiments, the controller

enables the high energy density storage bank to provide power to the first load when the first load demands a peak power (e.g., power that sustained from tens of seconds to minutes). In some embodiments, the controller charges the low energy density storage bank before the high energy density storage bank is charged. This allows the battery system to be ready to supply power for any sudden demand in power by the load. In some embodiments, the controller enables the low energy density storage bank as a primary battery source, and enables the high energy density storage bank as a secondary battery source, to provide power to the first load when the first load demands a sudden power. In some embodiments, the first and second loads include first and second racks of servers, respectively. In some embodiments, the power demand is a peak power shaving demand.

[0019] In some embodiments, during normal operation if for any reason AC power cannot fulfill the power demand, the BUs may supplement extra peak power of the of a sink device (e.g., a rack). As such, more sink devices (e.g., servers) can be added in a rack because BUs can step in to satisfy peak power demand. In some embodiments, the batteries or storage devices of the BUs include hybrid technologies. For example, a mix of high energy density charge carrying device(s) and low energy density charge carrying devices are used as batteries or storage devices. The high energy density charge carrying devices (e.g., Li-Ion batteries) charge slowly but carry high charge per unit area, while low density charge carrying devices (e.g., supercapacitor) charge relatively faster but carry less charge. As such, performance of BUs is improved by using the best characteristics of each type of storage technology. In some embodiments, a controller (e.g., hardware, software, or a combination of them) is used to analyze peak power patterns and minimizes the impact to overall lifespan of high energy density charge carrying device-based battery cells while maximizing service time for peak power shaving feature. In this case, peak power shaving service is maximized even when the high energy density charge carrying device-based battery cells are still undercharging.

[0020] In some embodiments, one source for discharging or suppling power is prioritized from among high-density and low-density energy sources according to their respective life span statistics. For example, low-density energy source (e.g., super capacitor) is used to supply power first. After real operation for some time, the life span data is monitored and analyzed and the dynamic change process for the supplies is performed to balance life-cycle of both energy sources.

[0021] The hybrid technologies for storage devices significantly reduces the impact to lifespan of an energy storage system under frequent charging-discharging scenarios. Other technical effects will be evident from the various embodiments and figures.

[0022] In the following description, numerous details are discussed to provide a more thorough explanation of embodiments of the present disclosure. It will be apparent, however, to one skilled in the art, that embodiments of the present disclosure may be practiced without these specific details. In other instances, well-known structures and devices are shown in block diagram form, rather than in detail, in order to avoid obscuring embodiments of the present disclosure.

[0023] Note that in the corresponding drawings of the embodiments, signals are represented with lines. Some lines may be thicker, to indicate more constituent signal paths, and/or have arrows at one or more ends, to indicate primary information flow direction. Such indications are not intended to be limiting. Rather, the lines are used in connection with one or more exemplary embodiments to facilitate easier understanding of a circuit or a logical unit. Any represented signal, as dictated by design needs or preferences, may actually comprise one or more signals that may travel in either direction and may be implemented with any suitable type of signal scheme.

[0024] **Fig. 1** illustrates apparatus 100 comprising distributed backup battery system (BBS) with peak power sharing capability, in accordance with some embodiments. Here, backup battery system is also referred to as backup battery unit (BU). Apparatus 100 comprises AC Power Source 101, AC Power Source 102, BUs 104-1, 104-2, through 104-N (where N is a number), sink devices such as server racks 160-1, 160-2, through 160-N, Power bus 130, Data bus 134, and nodes or rails 131-1, 131-2 through 131-N. In some embodiments, each BBS (also referred to as 104 where referred to generally) comprises charger 108, battery 112, switches 116 and 120, and controller 124. Elements referred to herein with a common reference label followed by a particular number or alphabet may be collectively referred to by the reference label alone. For example, BUs 104-1, 104-2, 104-N may be collectively and generally referred to as BUs 104 in plural, and BU 104 in singular.

[0025] In the example embodiment shown, BU 104-1 comprises a battery 112-1, switches 116-1, 120-1, and a node 131- coupled therebetween. Similarly, BU 104-2 comprises a battery 112-2, switches 116-2, 120-2, and a node 131-2 coupled therebetween – e.g., where BU 104-N comprises a battery 112-N, switches 116-N, 120-N, and a node 131-N

coupled therebetween. BUs 104-1, 104-2,...104-N further comprise respective controllers 124-1, 124-2,...124-N which are variously coupled to communicate via a data bus 134.

[0026] A power bus 130 is coupled to switches 120, and devices 160-1, 160-2,..., 160-N are coupled to BUs 104 via nodes 131-1, 131-2,..., 131-N, respectively. BUs 104 are variously coupled each to receive power from an AC power source 101 – e.g., where devices 160 are variously coupled each to receive power from another AC power source 102. AC power source 101 and AC power source 102 are different power sources, for example, or in some embodiments, the same power source.

[0027] In various embodiments, BU 104 provides battery back-up to one or more power sink devices 160 which are to couple to (or, in some embodiments, are part of) apparatus 100. For example, apparatus 100 comprises power sink 160-1, power sink 160-2, power sink 160-N, and so on. Any one of power sinks 160 represents a single device, or a collection of devices. Individual sinks 160 may represent any appropriate computing devices that may receive Alternating Current (AC) power (e.g., from a power source 102) and/or Direct Current (DC) power from BU 104. Merely as an example, apparatus 100 is implemented in a data center comprising servers or computing devices – e.g., where an individual sink 160 represents a server rack which comprises multiple servers. The teachings of this disclosure are not limited by a type and/or a number of sinks 160.

[0028] In some embodiments, power bus 130 is available to variously deliver power between a given one of BU 104 and a given one of sinks 160. As will be discussed in further details herein, power bus 130 may supply DC power to one or more of sinks 160, e.g., when AC power source 102 supplying AC power to the one or more sinks 160 fails and/or when one or more of BUs 104 is non-working.

[0029] In some embodiments, BUs 104 are coupled, directly or indirectly, to receive power from power source 101, where BUs 104 are made available to provide DC voltage to sinks 160 via battery primary power node 131 and/or power bus 130. For example, in some embodiments, some or all of BUs 104 are coupled to AC power source 101 via charging circuitry 108 which facilitates battery (re)charging using power from AC power sources 101. AC power sources 101 and 102 are different AC power sources, for example, or in some embodiments, the same AC power source.

[0030] In the example embodiment shown, BU 104 includes a switching circuitry which comprises a first switch circuit (including, for example, the illustrative switch 116) to couple battery 112 to node 131. In some embodiments, the switching circuitry includes a

second switch circuit (including, for example, the illustrative switch 120) to couple battery 112 to power bus 130. In some embodiments, BU 104 includes controller 124 (e.g., including hardware and/or executing software) which provides functionality to monitor one or more characteristics of power delivery by apparatus 100 and, based on such monitoring, to selectively configure a particular state of the switch circuitry. By way of illustration and not limitation, controller 124 may selectively configure one of multiple possible states of the switch circuitry– e.g., including a "battery backup" state wherein batteries BUs 104-1 through 104-N are electrically coupled to sinks 160-1 through 160-N, respectively. Such a battery backup state is configured, for example, in response to an indication that power delivery by AC power source 102 to sinks 160 is (or is expected to be) deficient, according to some predetermined performance criteria.

[0031] In one such embodiment, the multiple possible states further comprise a "backup redundancy" state wherein one of BUs 104 is electrically coupled to one of sinks 160 via bus 130 – e.g., while another of BUs 104 is electrically decoupled from sinks 160. Such a backup redundancy state is configured, for example, in response to an indication that – according to the same or some other predetermined performance criteria – power delivery by that particular one of BUs 104 is (or is expected to be) deficient – e.g., during the above-described deficiency of power delivery by AC power source 102 to sinks 160.

[0032] Additionally or alternatively, such multiple possible states further comprise, in some embodiments, a "disabled backup" state wherein each of BU 104 is electrically decoupled from sinks 160. Such a disabled backup state is configured, for example, in response to an indication that power delivery by AC power source 102 to sinks 160 is (or is expected to be) sufficient to meet a corresponding predetermined performance criteria.

[0033] In some embodiments, controller 124 is coupled to monitor one or more characteristics of actual or potential power delivery to sinks 160. For example, controller 124 is coupled to receive one or more signals on data bus 134 which describe or otherwise indicate one or more parameters of power delivery performance and/or capability – e.g., where performance monitoring by controller 124 includes one or more operations adapted from conventional power monitoring techniques.

[0034] During operation of apparatus 100, controller 124 may determine, based on the one or more signals on data bus 134 and one or more performance criteria, that some power delivery – e.g., with AC power source 102 and/or with one of BUs 104 – is (or is expected to be) deficient in one or more respects.

[0035] In one example scenario, signal on data bus 134 indicates a deficiency of power delivery with a first one of BUs 104. Based on this indication by signal on data bus 134, controller 124 selects a second one of BUs 104, and configures backup redundancy state via the switch circuitry, wherein the second one of BUs 104 is to function as an alternative to – that is, a substitute for – the first one of BUs 104. In one illustrative embodiment, in response to signal on data bus 134 indicating a deficiency of power delivery with BU 104-1, controller 124 (e.g., one of 124-1 through 124-N) selects another of BUs 104 to serve as an alternative to BU 104-1– e.g., where battery BU 104-2 is selected over battery BU 104-3. Based on this selection, controller 124 provides a control signal to transition switches 116, 120 from a first state to a second state.

[0036] The first state is, for example, a battery backup state which (by respective on states of switches 116-1 through 116-N) electrically couples batteries 112-1 through 112-N to nodes 131-1 through 131-N, respectively, while power bus 130 is electrically decoupled (by respective off states of switches 120-1 through 120-N) from each of nodes 131-1 through 131-N. The second state is a backup redundancy state which, in this example, electrically couples battery 112-2 (by respective on states of switches 116 and 120) to node 131-1 via bus 130, while battery 112-1 is electrically decoupled (by an off state of switch 120-1) from node 131-1. In one such embodiment, signal on data bus 134 subsequently indicates that power delivery with battery 112-1 is again sufficient to satisfy performance criteria. In response to such an indication, controller 124 transitions the switch circuitry from the second state back to the first state.

[0037] In some embodiments, controller 124 is operable to transition the switch circuitry from one backup redundancy state to another backup redundancy state. For example, while the switch circuitry is in the above-described second state, signal on data bus 134 may indicate to controller 124 a deficiency of power delivery with battery 112-2 (which is currently being provided as the alternative to battery 112-1). Based on such indicating by signal on data bus 134, controller 124 performs a selection of battery 112-N (for example) as an alternative to battery 112-2. Based on such selection, control signal on data bus 134 is communicated to transition the switch circuitry from the second state to a third state – i.e., a different backup redundancy state – wherein battery 112-N is electrically coupled to node 131-1 via bus 130, while battery 112-1 is electrically decoupled from node 131-1. During the third state, battery 112-2 is electrically decoupled from bus 130 and/or sink 160-2, in some embodiments.

[0038] In some embodiments, controller 124 additionally or alternatively operates to transition the switch circuitry between a disabled backup state and a battery backup state. For example, control signal on data bus 134 may configure a battery backup state where signal on the data bus indicates a deficiency of power delivery by AC power source 102 to sinks 160. Additionally or alternatively, control signal on data bus 134 may configure a disabled backup state where signal on the data bus 134 indicates that power delivery by AC power source 102 to sinks 160 is sufficient to satisfy some predetermined performance criteria.

[0039] In some embodiments, even when all BBS 104 are normal (i.e., fully charged and/or non-defective), it is possible that some BBS 104 may discharge for some time and the reserved capacity is not enough to continue to sustain peak power shaving demand. Here, peak power shaving demand refers to peak power sustaining time which is longer than one or some BBS's supplying capability. Under this situation, controller 124 may cause other BBSs (associated or paired with other sinks 160) whose capacity is still above a threshold (e.g., a programmable threshold) to provide supply power for that sink. As such, all sinks 160 get more chance to operate at peak power mode, which improves performance and provides for a higher density rack implementation.

[0040] In various embodiments, the distributed BBSs 104 associated with each sink 160 can support other sinks 160. For example, BBSs 104 can be used for not only supplying power to sink 160 it is associated with, but can also address peak power shaving demand by other sinks 160. As such, BBS 104 provides cross-sink power supply capability, in accordance with some embodiments. In various embodiments, power bus 130 connects multiple BBS 104 to share power. In some embodiments, data bus 134 is used to report battery charge status and availability of burst power supply from any one of BBSs 104.

[0041] For example, when charge level of battery 112-1 of BBS 104-1 is below a threshold (e.g., charge level is below 20%) but server rack 160-1 needs to operate in burst mode (with higher power demand than normal usage), controller 124-1 causes switch 120-1 to close and supplemental power is provided from one or more of BBS 104-2 through BBS 104-N to server rack 160-1. In some embodiments, any one of controllers 124 can assume a role of a master or supervisor controller to manage control of switches of any of the BBSs 104. In some embodiments, the supervisor controller monitors the power consumption of each rack 160 and battery level of each BBS 104 in real-time, and sends command(s) over data bus 134 to supervisee or slave controllers 124 to control switches so that any of the

server racks 160 can operate in burst modes despite some batteries of BBS 104 having low charge.

[0042] Continuing with the example, in some embodiments, while supplemental battery power from BBS 104-2 through 104-N are used for servers in rack 160-1, grid power to rack 160-1 may supply power to both servers in rack 160-1 and BBS 104-1. Here, grid power is power provided from AC power source 101 and/or power source 102. In some embodiments, one or few selected racks 160 can act as controller to monitor telemetry information and send out commands over data bus 134 to control switches 116 and/ 120 to move power to any rack 160 that demands it.

[0043] **Fig. 2** illustrates system 200 showing plurality of server racks, each having a BBS with peak power sharing capability, in accordance with some embodiments. System 200 shows another embodiment of apparatus 100. System 200 comprises server racks 260-1 through 260-M. In some embodiments, each server rack includes servers and a corresponding BBS. For example, server rack 260-1 includes servers 261-10, 261-11 through 261-1N. Server rack 260-2 includes servers 261-20, 261-21 through 261-2N. Server rack 260-M includes servers 261-N0, 261-N1 through 261-NN. Here, BBS 104 are relabeled as BBS 204 indicating that BBS are inside the rack. Each server 261 may include an SoC as described with reference to **Fig. 7**.

[0044] Referring back to **Fig. 2**. in some embodiments, BBS 204 in each rack 260 is used to support peak power. For example, when rack 260-1 needs peak power but the capacity of BBS 204-1 in rack 260-1 is not sufficient (e.g., charge is below a threshold such as 20%), other BBSs with enough reserved capacity (e.g., any, some, or all of BBS 204-2 through BBS 204-M in racks 260-2 through 260-M) give power to rack 260-1 if any, some, or all of BBS 204-2 through BBS 204-M are available for peak power shaving. In some embodiments, each BBS in the rack is not allowed to discharge below a lower threshold (e.g., 20%), which is a minimum required for BBS as a backup or uninterrupted power supply (UPS). Such condition is enforced to provide power to a rack during power outage. In some embodiments, each BBS is allowed to discharge to a low charge level (e.g., the lower threshold) or even zero to supply peak power demand to any of the racks. In this case, when AC grid power outage occurs, some of the BBS 204 still get enough energy to sustain a rack for a reasonable time.

[0045] In some embodiments, one of the BBSs 204 is given a role of an agent (master or supervisor) to monitor power consumption of all racks 260 and real-time capacity of all

BBSs 204. The agent BBS 204 calculates how many BBSs may discharge to the low-level threshold (e.g., below a threshold required for UPS). The calculation is dynamically updated and control command is sent by the agent BBS to selected BBS via data bus 134. Those selected BBS can then provide the supplemental power during peak power demand, in accordance with various embodiments.

[0046] In some embodiments, battery 112 of BBS 204 is charged when rack power demand is below a power budget set for grid power supply. In some embodiments, BBS 204 is charged dynamically or adaptively. For example, a controller or the agent BBS 204 can prioritize grid power budget to charge those BBS 204 in the racks 260 with higher workload priority. In some embodiments, a controller or the agent BBS 204 (e.g., 204-1) can use other reserved charge capacity of other BBSs (e.g., 204-2 through 204-M) to sustain more power for the selected rack with high workload priority. As such, remaining grid power budget can be used to charge BBS 204-1. The idea described herein can be extended to BBS of several high priority racks.

[0047] **Fig. 3** illustrates flowchart 300 of a method for peak power sharing using distributed BBS, in accordance with some embodiments. While the blocks are described in a particular order, the order can be modified. For example, some blocks can be performed before others while some blocks can be performed in parallel. The process of flowchart 300 can be performed by hardware, software, or a combination of them. In some embodiments, flowchart 300 is performed by an agent or controller 124 of a BBS, a controller or logic outside of BBS, a power management controller of a sink (e.g., rack 160), a power management unit (p-unit) of a processor in a server of rack 160.

[0048] At block 301, an agent BBS 204-1 or controller 124 of the agent BBS determines whether the charge level of the battery 112-1 of BBS 204-1 in rack 260-1 is too low (e.g., below a 20% threshold relative to a battery full charge) to support bust power mode for rack 260-1. If the agent determines that the charge level of the battery 112-1 of BBS 204-1 in rack 260-1 is too low, the process proceeds to block 302. If the agent determines that the charge level of battery 112-1 of BBS 204-1 in rack 260-1 is above the threshold, the process proceeds to block 305. At block 305, the system continues to perform normally.

[0049] At block 302, agent BBS 204-1 or a controller 124 of the agent BBS (or a power management unit of rack 260-1 or of any other rack) determines whether it needs a burst of power based on the workload (e.g., anticipated workload). If it is determined that rack 260-1 is not expecting workload that requires burst power, the process proceeds to block

305. Otherwise, the process proceeds to block 303. At block 303, agent BBS 204-1 or a controller 124 of the agent BBS (or a power management unit of rack 260-1 or any other rack) determines whether BBS 204-2 through BBS 104-M in racks 260-2 through 260-M are available for peak power shaving. If it is determined that BBS 204-2 through BBS 104-M in racks 260-2 through 260-M are available for peak power shaving, the process proceeds to block 304. Otherwise, the process proceeds to block 305. At block 304, agent BBS 204-1 or a controller 124 of the agent BBS (or a power management unit of rack 260-1 or of any other rack) causes BBS 204-2 through BBS 204-M to supply power to rack 260-1 for bust power mode in rack 260-1.

[0050] In various embodiments, power bus 130 can be switchably coupled to a power grid. The power grid can be node 131 which is organized as a power distribution network to provide power to one or more racks. In one such embodiment, power supply on the power bus 130 and/or power grid can be applied bi-directionally for supporting burst mode for any rack and/or for charging the battery of the BBS.

[0051] Described herein is a control mechanism for a hybrid energy source to supply peak power demand. Extended use of battery 112 (e.g., Li-ion based batter) in BBS 204 for peak power shaving negatively impacts life cycles of BBS 204. Further, to support burst power needs, battery 112 is charged regularly and quickly. For example, for BBS 204 to support peak power shaving, the battery of the BBS is charged above a threshold.

[0052] **Fig. 4** illustrates apparatus 400 comprising a controller for managing hybrid energy sources to supply peak power demand, in accordance with some embodiments. Apparatus 400 illustrates a hybrid energy system 401 and a power sink 402. In this example, power sink 402 is a group of servers 461-1 through 461-N organized in a rack. However, the embodiments are applicable to any power sink. In some embodiments, apparatus 400 includes a global control agent 403 that receives telemetry from power sink 402 and determines an optimal scheme to provide power including peak power to power sink 402. In some embodiments, hybrid energy system 401 comprises a local control agent 404, charging regulator 405, high energy density charge bank 406, low energy density charge bank 407, and discharging regulator 408.

[0053] In some embodiments, global control agent 403 and local control agent 404 takes advantage of characteristics of high energy density charge bank 406 and low energy density charge bank 407 by analyzing peak power patterns of power sink 402. Global control agent 403 manages a number of power sinks including power sink 402. Local control agent

404 monitor the charging and discharging cycles of banks 406 and 407 to determine how much charge should be provided by high energy density charge bank 406 to power sink 402 and how much power should be provided by low energy density charge bank 407. As such, hybrid energy system 401 minimizes the impact to overall lifespan of battery cells of high energy density charge bank 406 while to maximize service time for peak power shaving feature.

[0054] In some embodiments, the roles of global control agent 403 and local control agent 404 can be combined in one controller. In some embodiments, global control agent 403 and local control agent 404 are implemented as hardware, software, or a combination of them. In some embodiments, global control agent 403 and local control agent 404 are part of controller 124.

[0055] Various embodiments herein describe high energy density charge bank 406 as comprising Li-ion battery cells, and low energy density charge bank 407 as comprising super capacitors. However, other types of battery cells and/or charge banks can be used for high energy density charge bank 406 and low energy density charge bank 407. Table 1 provides a comparison between characteristics of high energy density charge bank 406 and low energy density charge bank 407.

Table 1

Characteristics	**High energy density charge bank 406 (e.g., Li-ion or Li-ion phosphate, dry battery, acid battery cells)**	**Low energy density charge bank 407 (e.g., super capacitor)**
Charging Time	Minutes (e.g., 10-60 minutes)	Seconds (e.g., 1-20 seconds)
Cycle life	1000 to 4000 times	Million times (unlimited)
Energy density	High (e.g., 100 wh/Kg to 200 wh/Kg	Low, 5-10 wh/Kg
Power density	Low (e.g., 1k-3k w/kg)	High, 10k w/kg
Service life	5-10 years	10-15 years

[0056] Low energy density charge bank 407 (e.g., super capacitor) has a large of charging-discharging cycles, and it is suitable to supply energy demand in short periods of time but with frequent occurrences. High energy density charge bank 406 (e.g., Li-ion battery bank) has higher energy density and is more suitable for energy demand of long period of time but with lower frequency of occurrences (considering its charge time is relative longer). In various embodiments, low energy density charge bank 407 is added as a

first tier to supply peak power demand (by power sink 402) over power budget. Following definitions are used herein.

[0057] G_Agent 403: Global control agent to manage a bunch of servers;

[0058] L_Agent 404: Local control agent in hybrid energy system;

[0059] A_DoD_s: The allowable depth of discharging of super capacitor bank;

[0060] A_DoD_b: The allowable depth of discharging (DoD) of Li-ion/LFP battery bank;

[0061] SoC_s: The current capacity (state of charge) of low energy density charge bank 407, herein super capacitor bank;

[0062] SoC_b: The current capacity (state of charge) of high energy density charge bank 406, herein Li-ion/LFP battery bank;

[0063] P_{budget}: The power budget for the configured system;

[0064] P_{demand}: The power demand for the configured system;

[0065] P_{in}: The input power for the configured system.

[0066] According to peak power demand pattern of workloads, the discharge control method performed by discharging regulator 408 is decomposed into three level of control policies.

[0067] Level 1 (L1) – Using low energy density charge bank 407 (e.g., super capacitor) to supply the energy demand for short spikes from millions of seconds to tens of seconds, and the allowable depth of discharging is defined as A_DoD_s. If low energy density charge bank 407 may not have view of whether the coming peak power is within above short period, it still serves as primary energy source to supply the peak power till its capacity down to certain level.

[0068] Level 2 (L2) – Using high energy density charge bank 406 (e.g., Li-ion battery) to supply the peak demand from tens of seconds to minutes, and the allowable depth of discharging is defined as A_DoD_b.

[0069] Level 3 (L3) – Using rack power control (like power capping) to manage the power consumption back under defined the power budget.

[0070] **Fig. 5** illustrates plot 500 showing peak power scenarios—L1, L2, and L3. Referring back to **Fig. 4**, for discharging policy, low energy density charge bank 407 is prioritized to supply energy demand while the power is over power budget to reduce the number of discharges for high energy density charge bank 406 (e.g., Li-ion battery bank).

[0071] Charging policy is managed by charging regulator 405. For charging policy, low energy density charge bank 407 is prioritized when there is extra power energy (from total budget) to use.

[0072] In some embodiments, global control agent 403 and/or local control agent 404 adjust ratio of the depth of discharging (DoD) between low energy density charge bank 407 (e.g., super capacitor bank) and high energy density charge bank 406 (e.g., battery bank) based on runtime health status indicator. This is to balance and maximize life cycles of both energy sources 406 and 407.

[0073] The state-of-health (SOH) is one feature to evaluate the lifespan of high energy density charge bank 406 and low energy density charge bank 407. Normally, the SOH ranges from 0 to 100%, and SOH below 80% is considered as the hit of maximum cycles of the battery system.

[0074] Herein using L_{h} is used to represent a default check point of SOH for high energy density charge bank 406 (e.g., Li-ion battery) and low energy density charge bank 407 (e.g., super capacitor). L_{h} is longer than its limit. Here, limit hit refers to a need to replace the super capacitor or Li-ion battery. In some embodiments, local control agent 404 periodically poll SOH of super capacitor bank 407 (SOH_s) and battery bank 406 (SOH_b) respectively, and calculates the SOH drop rate for super capacitor bank 407 (SOH_D_s) and battery bank 406 (SOH_D_b). With the average SOH drop rate calculated, local control agent 404 is able to estimate the remaining lifespan of the super capacitor bank 407 (L_s) and battery bank 406 (L_b). If $L_s > L_h$ and $L_b < L_h$, L_Agent 404 increases A_DoD_s and decreases A_DoD_b. If $L_s < L_h$ and $L_b > L_h$, L_Agent 404 decreases A_DoD_s and increases A_DoD_b. If $L_s < L_h$ and $L_b < L_h$, L_Agent 404 decreases A_DoD_s and decreases A_DoD_b.

[0075] **Fig. 6** illustrates flowchart 600 of a method for managing hybrid energy sources to supply peak power demand, in accordance with some embodiments. While various blocks are illustrated in a particular order, the order can be modified. For example, some blocks can be performed before others while some blocks are performed in parallel. The method of flowchart 600 is performed by global control agent 403 and/or local control agent 404. In some embodiments, method of flowchart 600 is performed by controller 124. In some embodiments, hybrid energy system 401 and/or global control agent 403 is implemented in a power control unit (PCU or p-unit) for any one of servers 461-1 through 461-N. In one such embodiment, method of flowchart 600 is performed by the PCU.

[0076] At block 601, power demand is compared with power budget. If the power demand is greater than the power budget, the process proceeds to block 602. At block 602, state-of-charge (SOC) of the low energy density storage bank 407 (e.g., super capacitor) is compared with A_DoD_s, the allowable depth of discharging of super capacitor bank. If the SOH_s is greater than (1- A_DoD_s), the process proceeds to block 603. At block 603, local agent 404 applies L1 discharging policy. Under this policy, low energy density charge bank 407 (e.g., super capacitor) supplies the energy demand for short spikes from millions of seconds to tens of seconds, and the allowable depth of discharging is defined as A_DoD_s. If low energy density charge bank 407 may not have view of whether the coming peak power is within above short period, it still serves as primary energy source to supply the peak power till its capacity is down to certain level.

[0077] If the SOC_s is less than or equal to (1- A_DoD_s), the process proceeds to block 604. At block 604, state-of-charge (SOC) of the high energy density storage bank 406 (e.g., Li-ion battery) is compared with A_DoD_b, the allowable depth of discharging (DoD) of Li-ion/LFP battery bank. If the SOC_b is greater than (1- A_DoD_b), the process proceeds to block 605. At block 605, L2 discharging policy is applied. In L2, high energy density charge bank 406 (e.g., Li-ion battery) is used to supply the peak demand from tens of seconds to minutes, and the allowable depth of discharging is defined as A_DoD_b. If the SOC_b is less than or equal to (1- A_DoD_b), the process proceeds to block 606. At block 606, L3 control policy is applied. In L3, rack power control (like power capping) is used to manage the power consumption back under defined the power budget. The process then proceeds to block 607.

[0078] At block 607, SOC_s is compared with (1- A_DoD_s), and SOC_b is compared with (1- A_DoD_b). If either, SOC_s is greater than (1- A_DoD_s) or SOC_b is greater that (1- A_DoD_b), then the process proceeds back to block 601. If both SOC_s is less than or equal to (1- A_DoD_s) and SOH_b is less than or equal to (1- A_DoD_b), the process reverts both to 606.

[0079] Referring back to block 601, if power demand is less than or equal to the power budget, the process proceeds to block 608. In this branch of flowchart 600, low energy density bank 407 is charged first followed by charging of high energy density bank 406. At block 608, the state of charge of the super capacitor is compared with its full charge capacity. If SOC_s is full or substantially full, the process proceeds to block 610. At block 610, the state of charge of the Li-ion is compared with its full charge capacity. If SOC_b is full or substantially full, there is not much to charge and the process proceeds to block 601. In some embodiments, both super capacitor and Li-ion storage banks are charged to 100%. In

some embodiments, Li-ion storage bank is charged to up a level slightly less than 100% (e.g., 90% or 95%) to extend life of the Li-ion storage bank.

[0080] If SOC_s is not full, the process proceeds to block 609 where low energy density capacitive bank 407 (super capacitor) is charged using part or completely available energy. In some embodiments, charge from other BBS 104 can be used to charge the battery 112 (here hybrid battery). The process then proceeds to block 608. If SOC_b is not full, the process proceeds to block 611 where high energy density capacitive bank 406 (e.g., Li-ion battery) is charged using part or completely available energy. In some embodiments, charge from other BBS 104 can be used to charge the battery 112 (here hybrid battery). The process then proceeds to block 610.

[0081] **Fig. 7** illustrates plot 700 showing one case of managing hybrid energy source, in accordance with some embodiments. Here, SOH_b is 701 and SOH_s is 702. If $L_s > L_h$ and $L_b < L_h$, L_Agent 404 increases A_DoD_s and decreases A_DoD_b. If $L_s < L_h$ and $L_b > L_h$, L_Agent 404 decreases A_DoD_s and increases A_DoD_b. If $L_s < L_h$ and $L_b < L_h$, L_Agent 404 decreases A_DoD_s and decreases A_DoD_b.

[0082] Elements of embodiments (e.g., flowchart with reference to **Figs. 3 and 6**) are also provided as a machine-readable medium (e.g., memory) for storing the computer-executable instructions (e.g., instructions to implement any other processes discussed herein). In some embodiments, computing platform comprises memory, processor, machine-readable storage media (also referred to as tangible machine-readable medium), communication interface (e.g., wireless or wired interface), and network bus coupled together.

[0083] In some embodiments, processor is a Digital Signal Processor (DSP), an Application Specific Integrated Circuit (ASIC), a general-purpose Central Processing Unit (CPU), or a low power logic implementing a simple finite state machine to perform the method with reference to **Figs. 3 and 6** and/or various embodiments, etc.

[0084] In some embodiments, the various logic blocks of system are coupled together via a Network Bus. Any suitable protocol may be used to implement the network bus. In some embodiments, machine-readable storage medium includes Instructions (also referred to as the program software code/instructions) for calculating or measuring distance and relative orientation of a device with reference to another device as described with reference to various embodiments and flowchart.

[0085] Program software code/instructions associated with flowcharts with reference to **Figs. 3 and 6** (and/or various embodiments) and executed to implement embodiments of

the disclosed subject matter may be implemented as part of an operating system or a specific application, component, program, object, module, routine, or other sequence of instructions or organization of sequences of instructions referred to as "program software code/instructions," "operating system program software code/instructions," "application program software code/instructions," or simply "software" or firmware embedded in processor. In some embodiments, the program software code/instructions associated with flowcharts with reference to **Figs. 3 and 6** (and/or various embodiments) are executed by system.

[0086] In some embodiments, the program software code/instructions associated with reference to **Figs. 3 and 6** (and/or various embodiments) are stored in a computer executable storage medium and executed by the processor. Here, computer executable storage medium is a tangible machine-readable medium that can be used to store program software code/instructions and data that, when executed by a computing device, causes one or more processors to perform a method(s) as may be recited in one or more accompanying claims directed to the disclosed subject matter.

[0087] The tangible machine-readable medium may include storage of the executable software program code/instructions and data in various tangible locations, including for example ROM, volatile RAM, non-volatile memory and/or cache and/or other tangible memory as referenced in the present application. Portions of this program software code/instructions and/or data may be stored in any one of these storage and memory devices. Further, the program software code/instructions can be obtained from other storage, including, e.g., through centralized servers or peer to peer networks and the like, including the Internet. Different portions of the software program code/instructions and data can be obtained at different times and in different communication sessions or in the same communication session.

[0088] The software program code/instructions (associated with reference to **Figs. 3 and 6** and other embodiments) and data can be obtained in their entirety prior to the execution of a respective software program or application by the computing device. Alternatively, portions of the software program code/instructions and data can be obtained dynamically, e.g., just in time, when needed for execution. Alternatively, some combination of these ways of obtaining the software program code/instructions and data may occur, e.g., for different applications, components, programs, objects, modules, routines or other sequences of instructions or organization of sequences of instructions, by way of example.

Thus, it is not required that the data and instructions be on a tangible machine readable medium in entirety at a particular instance of time.

[0089] Examples of tangible computer-readable media include but are not limited to recordable and non-recordable type media such as volatile and non-volatile memory devices, read only memory (ROM), random access memory (RAM), flash memory devices, floppy and other removable disks, magnetic storage media, optical storage media (e.g., Compact Disk Read-Only Memory (CD ROMS), Digital Versatile Disks (DVDs), etc.), among others. The software program code/instructions may be temporarily stored in digital tangible communication links while implementing electrical, optical, acoustical or other forms of propagating signals, such as carrier waves, infrared signals, digital signals, etc. through such tangible communication links.

[0090] In general, tangible machine readable medium includes any tangible mechanism that provides (i.e., stores and/or transmits in digital form, e.g., data packets) information in a form accessible by a machine (i.e., a computing device), which may be included, e.g., in a communication device, a computing device, a network device, a personal digital assistant, a manufacturing tool, a mobile communication device, whether or not able to download and run applications and subsidized applications from the communication network, such as the Internet, e.g., an iPhone®, Galaxy®, or the like, or any other device including a computing device. In one embodiment, processor-based system is in a form of or included within a PDA (personal digital assistant), a cellular phone, a notebook computer, a tablet, a game console, a set top box, an embedded system, a TV (television), a personal desktop computer, etc. Alternatively, the traditional communication applications and subsidized application(s) may be used in some embodiments of the disclosed subject matter.

[0091] **Fig. 8** illustrates a smart device or a computer system or a SoC (System-on-Chip) with a BBS with peak power sharing capability and/or hybrid energy sources to supply peak power demand, in accordance with some embodiments. It is pointed out that those elements of **Fig. 8** having the same reference numbers (or names) as the elements of any other figure may operate or function in any manner similar to that described, but are not limited to such. Any block in this smart device can have the apparatus for dynamically optimizing battery charging voltage.

[0092] In some embodiments, device 5500 represents an appropriate computing device, such as a computing tablet, a mobile phone or smart-phone, a laptop, a desktop, an Internet-of-Things (IOT) device, a server, a wearable device, a set-top box, a wireless-

enabled e-reader, or the like. It will be understood that certain components are shown generally, and not all components of such a device are shown in device 5500.

[0093] In an example, the device 5500 comprises an SoC (System-on-Chip) 5501. An example boundary of the SoC 5501 is illustrated using dotted lines in **Fig. 8**, with some example components being illustrated to be included within SoC 5501 – however, SoC 5501 may include any appropriate components of device 5500.

[0094] In some embodiments, device 5500 includes processor 5504. Processor 5504 can include one or more physical devices, such as microprocessors, application processors, microcontrollers, programmable logic devices, processing cores, or other processing implementations such as disaggregated combinations of multiple compute, graphics, accelerator, I/O and/or other processing chips. The processing operations performed by processor 5504 include the execution of an operating platform or operating system on which applications and/or device functions are executed. The processing operations include operations related to I/O (input/output) with a human user or with other devices, operations related to power management, operations related to connecting computing device 5500 to another device, and/or the like. The processing operations may also include operations related to audio I/O and/or display I/O.

[0095] In some embodiments, processor 5504 includes multiple processing cores (also referred to as cores) 5508a, 5508b, 5508c. Although merely three cores 5508a, 5508b, 5508c are illustrated in **Fig. 8**, processor 5504 may include any other appropriate number of processing cores, e.g., tens, or even hundreds of processing cores. Processor cores 5508a, 5508b, 5508c may be implemented on a single integrated circuit (IC) chip. Moreover, the chip may include one or more shared and/or private caches, buses or interconnections, graphics and/or memory controllers, or other components.

[0096] In some embodiments, processor 5504 includes cache 5506. In an example, sections of cache 5506 may be dedicated to individual cores 5508 (e.g., a first section of cache 5506 dedicated to core 5508a, a second section of cache 5506 dedicated to core 5508b, and so on). In an example, one or more sections of cache 5506 may be shared among two or more of cores 5508. Cache 5506 may be split in different levels, e.g., level 1 (L1) cache, level 2 (L2) cache, level 3 (L3) cache, etc.

[0097] In some embodiments, processor core 5504 may include a fetch unit to fetch instructions (including instructions with conditional branches) for execution by the core 5504. The instructions may be fetched from any storage devices such as the memory 5530.

Processor core 5504 may also include a decode unit to decode the fetched instruction. For example, the decode unit may decode the fetched instruction into a plurality of micro-operations. Processor core 5504 may include a schedule unit to perform various operations associated with storing decoded instructions. For example, the schedule unit may hold data from the decode unit until the instructions are ready for dispatch, e.g., until all source values of a decoded instruction become available. In one embodiment, the schedule unit may schedule and/or issue (or dispatch) decoded instructions to an execution unit for execution.

[0098] The execution unit may execute the dispatched instructions after they are decoded (e.g., by the decode unit) and dispatched (e.g., by the schedule unit). In an embodiment, the execution unit may include more than one execution unit (such as an imaging computational unit, a graphics computational unit, a general-purpose computational unit, etc.). The execution unit may also perform various arithmetic operations such as addition, subtraction, multiplication, and/or division, and may include one or more an arithmetic logic units (ALUs). In an embodiment, a co-processor (not shown) may perform various arithmetic operations in conjunction with the execution unit.

[0099] Further, execution unit may execute instructions out-of-order. Hence, processor core 5504 may be an out-of-order processor core in one embodiment. Processor core 5504 may also include a retirement unit. The retirement unit may retire executed instructions after they are committed. In an embodiment, retirement of the executed instructions may result in processor state being committed from the execution of the instructions, physical registers used by the instructions being de-allocated, etc. Processor core 5504 may also include a bus unit to enable communication between components of processor core 5504 and other components via one or more buses. Processor core 5504 may also include one or more registers to store data accessed by various components of the core 5504 (such as values related to assigned app priorities and/or sub-system states (modes) association.

[00100] In some embodiments, device 5500 comprises connectivity circuitries 5531. For example, connectivity circuitries 5531 includes hardware devices (e.g., wireless and/or wired connectors and communication hardware) and/or software components (e.g., drivers, protocol stacks), e.g., to enable device 5500 to communicate with external devices. Device 5500 may be separate from the external devices, such as other computing devices, wireless access points or base stations, etc.

[00101] In an example, connectivity circuitries 5531 may include multiple different types of connectivity. To generalize, the connectivity circuitries 5531 may include cellular connectivity circuitries, wireless connectivity circuitries, etc. Cellular connectivity circuitries of connectivity circuitries 5531 refers generally to cellular network connectivity provided by wireless carriers, such as provided via GSM (global system for mobile communications) or variations or derivatives, CDMA (code division multiple access) or variations or derivatives, TDM (time division multiplexing) or variations or derivatives, 3rd Generation Partnership Project (3GPP) Universal Mobile Telecommunications Systems (UMTS) system or variations or derivatives, 3GPP Long-Term Evolution (LTE) system or variations or derivatives, 3GPP LTE-Advanced (LTE-A) system or variations or derivatives, Fifth Generation (5G) wireless system or variations or derivatives, 5G mobile networks system or variations or derivatives, 5G New Radio (NR) system or variations or derivatives, or other cellular service standards. Wireless connectivity circuitries (or wireless interface) of the connectivity circuitries 5531 refers to wireless connectivity that is not cellular, and can include personal area networks (such as Bluetooth, Near Field, etc.), local area networks (such as Wi-Fi), and/or wide area networks (such as WiMax), and/or other wireless communication. In an example, connectivity circuitries 5531 may include a network interface, such as a wired or wireless interface, e.g., so that a system embodiment may be incorporated into a wireless device, for example, a cell phone or personal digital assistant.

[00102] In some embodiments, device 5500 comprises control hub 5532, which represents hardware devices and/or software components related to interaction with one or more I/O devices. For example, processor 5504 may communicate with one or more of display 5522, one or more peripheral devices 5524, storage devices 5528, one or more other external devices 5529, etc., via control hub 5532. Control hub 5532 may be a chipset, a Platform Control Hub (PCH), and/or the like.

[00103] For example, control hub 5532 illustrates one or more connection points for additional devices that connect to device 5500, e.g., through which a user might interact with the system. For example, devices (e.g., devices 5529) that can be attached to device 5500 include microphone devices, speaker or stereo systems, audio devices, video systems or other display devices, keyboard or keypad devices, or other I/O devices for use with specific applications such as card readers or other devices.

[00104] As mentioned above, control hub 5532 can interact with audio devices, display 5522, etc. For example, input through a microphone or other audio device can provide input

or commands for one or more applications or functions of device 5500. Additionally, audio output can be provided instead of, or in addition to display output. In another example, if display 5522 includes a touch screen, display 5522 also acts as an input device, which can be at least partially managed by control hub 5532. There can also be additional buttons or switches on computing device 5500 to provide I/O functions managed by control hub 5532. In one embodiment, control hub 5532 manages devices such as accelerometers, cameras, light sensors or other environmental sensors, or other hardware that can be included in device 5500. The input can be part of direct user interaction, as well as providing environmental input to the system to influence its operations (such as filtering for noise, adjusting displays for brightness detection, applying a flash for a camera, or other features).

[00105] In some embodiments, control hub 5532 may couple to various devices using any appropriate communication protocol, e.g., PCIe (Peripheral Component Interconnect Express), USB (Universal Serial Bus), Thunderbolt, High Definition Multimedia Interface (HDMI), Firewire, etc.

[00106] In some embodiments, display 5522 represents hardware (e.g., display devices) and software (e.g., drivers) components that provide a visual and/or tactile display for a user to interact with device 5500. Display 5522 may include a display interface, a display screen, and/or hardware device used to provide a display to a user. In some embodiments, display 5522 includes a touch screen (or touch pad) device that provides both output and input to a user. In an example, display 5522 may communicate directly with the processor 5504. Display 5522 can be one or more of an internal display device, as in a mobile electronic device or a laptop device or an external display device attached via a display interface (e.g., DisplayPort, etc.). In one embodiment display 5522 can be a head mounted display (HMD) such as a stereoscopic display device for use in virtual reality (VR) applications or augmented reality (AR) applications.

[00107] In some embodiments, and although not illustrated in the figure, in addition to (or instead of) processor 5504, device 5500 may include Graphics Processing Unit (GPU) comprising one or more graphics processing cores, which may control one or more aspects of displaying contents on display 5522.

[00108] Control hub 5532 (or platform controller hub) may include hardware interfaces and connectors, as well as software components (e.g., drivers, protocol stacks) to make peripheral connections, e.g., to peripheral devices 5524.

[00109] It will be understood that device 5500 could both be a peripheral device to other computing devices, as well as have peripheral devices connected to it. Device 5500 may have a "docking" connector to connect to other computing devices for purposes such as managing (e.g., downloading and/or uploading, changing, synchronizing) content on device 5500. Additionally, a docking connector can allow device 5500 to connect to certain peripherals that allow computing device 5500 to control content output, for example, to audiovisual or other systems.

[00110] In addition to a proprietary docking connector or other proprietary connection hardware, device 5500 can make peripheral connections via common or standards-based connectors. Common types can include a Universal Serial Bus (USB) connector (which can include any of a number of different hardware interfaces), DisplayPort including MiniDisplayPort (MDP), High Definition Multimedia Interface (HDMI), Firewire, or other types.

[00111] In some embodiments, connectivity circuitries 5531 may be coupled to control hub 5532, e.g., in addition to, or instead of, being coupled directly to the processor 5504. In some embodiments, display 5522 may be coupled to control hub 5532, e.g., in addition to, or instead of, being coupled directly to processor 5504.

[00112] In some embodiments, device 5500 comprises memory 5530 coupled to processor 5504 via memory interface 5534. Memory 5530 includes memory devices for storing information in device 5500.

[00113] In some embodiments, memory 5530 includes apparatus to maintain stable clocking as described with reference to various embodiments. Memory can include nonvolatile (state does not change if power to the memory device is interrupted) and/or volatile (state is indeterminate if power to the memory device is interrupted) memory devices. Memory device 5530 can be a dynamic random-access memory (DRAM) device, a static random-access memory (SRAM) device, flash memory device, phase-change memory device, or some other memory device having suitable performance to serve as process memory. In one embodiment, memory 5530 can operate as system memory for device 5500, to store data and instructions for use when the one or more processors 5504 executes an application or process. Memory 5530 can store application data, user data, music, photos, documents, or other data, as well as system data (whether long-term or temporary) related to the execution of the applications and functions of device 5500.

[00114] Elements of various embodiments and examples are also provided as a machine-readable medium (e.g., memory 5530) for storing the computer-executable instructions (e.g., instructions to implement any other processes discussed herein). The machine-readable medium (e.g., memory 5530) may include, but is not limited to, flash memory, optical disks, CD-ROMs, DVD ROMs, RAMs, EPROMs, EEPROMs, magnetic or optical cards, phase change memory (PCM), or other types of machine-readable media suitable for storing electronic or computer-executable instructions. For example, embodiments of the disclosure may be downloaded as a computer program (e.g., BIOS) which may be transferred from a remote computer (e.g., a server) to a requesting computer (e.g., a client) by way of data signals via a communication link (e.g., a modem or network connection).

[00115] In some embodiments, device 5500 comprises temperature measurement circuitries 5540, e.g., for measuring temperature of various components of device 5500. In an example, temperature measurement circuitries 5540 may be embedded, or coupled or attached to various components, whose temperature are to be measured and monitored. For example, temperature measurement circuitries 5540 may measure temperature of (or within) one or more of cores 5508a, 5508b, 5508c, voltage regulator 5514, memory 5530, a mother-board of SoC 5501, and/or any appropriate component of device 5500. In some embodiments, temperature measurement circuitries 5540 include a low power hybrid reverse (LPHR) bandgap reference (BGR) and digital temperature sensor (DTS), which utilizes subthreshold metal oxide semiconductor (MOS) transistor and the PNP parasitic Bi-polar Junction Transistor (BJT) device to form a reverse BGR that serves as the base for configurable BGR or DTS operating modes. The LPHR architecture uses low-cost MOS transistors and the standard parasitic PNP device. Based on a reverse bandgap voltage, the LPHR can work as a configurable BGR. By comparing the configurable BGR with the scaled base-emitter voltage, the circuit can also perform as a DTS with a linear transfer function with single-temperature trim for high accuracy.

[00116] In some embodiments, device 5500 comprises power measurement circuitries 5542, e.g., for measuring power consumed by one or more components of the device 5500. In an example, in addition to, or instead of, measuring power, the power measurement circuitries 5542 may measure voltage and/or current. In an example, the power measurement circuitries 5542 may be embedded, or coupled or attached to various components, whose power, voltage, and/or current consumption are to be measured and monitored. For example,

power measurement circuitries 5542 may measure power, current and/or voltage supplied by one or more voltage regulators 5514, power supplied to SoC 5501, power supplied to device 5500, power consumed by processor 5504 (or any other component) of device 5500, etc.

[00117] In some embodiments, device 5500 comprises one or more voltage regulator circuitries, generally referred to as voltage regulator (VR) 5514. VR 5514 generates signals at appropriate voltage levels, which may be supplied to operate any appropriate components of the device 5500. Merely as an example, VR 5514 is illustrated to be supplying signals to processor 5504 of device 5500. In some embodiments, VR 5514 receives one or more Voltage Identification (VID) signals, and generates the voltage signal at an appropriate level, based on the VID signals. Various type of VRs may be utilized for the VR 5514. For example, VR 5514 may include a "buck" VR, "boost" VR, a combination of buck and boost VRs, low dropout (LDO) regulators, switching DC-DC regulators, constant-on-time controller-based DC-DC regulator, etc. Buck VR is generally used in power delivery applications in which an input voltage needs to be transformed to an output voltage in a ratio that is smaller than unity. Boost VR is generally used in power delivery applications in which an input voltage needs to be transformed to an output voltage in a ratio that is larger than unity. In some embodiments, each processor core has its own VR, which is controlled by PCU 5510a/b and/or PMIC 5512. In some embodiments, each core has a network of distributed LDOs to provide efficient control for power management. The LDOs can be digital, analog, or a combination of digital or analog LDOs. In some embodiments, VR 5514 includes current tracking apparatus to measure current through power supply rail(s).

[00118] In some embodiments, VR 5514 includes a digital control scheme to manage states of a proportional-integral-derivative (PID) filter (also known as a digital Type-III compensator). The digital control scheme controls the integrator of the PID filter to implement non-linear control of saturating the duty cycle during which the proportional and derivative terms of the PID are set to 0 while the integrator and its internal states (previous values or memory) is set to a duty cycle that is the sum of the current nominal duty cycle plus a deltaD. The deltaD is the maximum duty cycle increment that is used to regulate a voltage regulator from ICCmin to ICCmax and is a configuration register that can be set post silicon. A state machine moves from a non-linear all ON state (which brings the output voltage Vout back to a regulation window) to an open loop duty cycle which maintains the output voltage slightly higher than the required reference voltage Vref. After a certain period in this state of open loop at the commanded duty cycle, the state machine then ramps down the open loop

duty cycle value until the output voltage is close to the Vref commanded. As such, output chatter on the output supply from VR 5514 is completely eliminated (or substantially eliminated) and there is merely a single undershoot transition which could lead to a guaranteed Vmin based on a comparator delay and the di/dt of the load with the available output decoupling capacitance.

[00119] In some embodiments, VR 5514 includes a separate self-start controller, which is functional without fuse and/or trim information. The self-start controller protects VR 5514 against large inrush currents and voltage overshoots, while being capable of following a variable VID (voltage identification) reference ramp imposed by the system. In some embodiments, the self-start controller uses a relaxation oscillator built into the controller to set the switching frequency of the buck converter. The oscillator can be initialized using either a clock or current reference to be close to a desired operating frequency. The output of VR 5514 is coupled weakly to the oscillator to set the duty cycle for closed loop operation. The controller is naturally biased such that the output voltage is always slightly higher than the set point, eliminating the need for any process, voltage, and/or temperature (PVT) imposed trims.

[00120] In some embodiments, device 5500 comprises one or more clock generator circuitries, generally referred to as clock generator 5516. Clock generator 5516 generates clock signals at appropriate frequency levels, which may be supplied to any appropriate components of device 5500. Merely as an example, clock generator 5516 is illustrated to be supplying clock signals to processor 5504 of device 5500. In some embodiments, clock generator 5516 receives one or more Frequency Identification (FID) signals, and generates the clock signals at an appropriate frequency, based on the FID signals.

[00121] In some embodiments, device 5500 comprises battery 5518 supplying power to various components of device 5500. Merely as an example, battery 5518 is illustrated to be supplying power to processor 5504. Although not illustrated in the figures, device 5500 may comprise a charging circuitry, e.g., to recharge the battery, based on Alternating Current (AC) power supply received from an AC adapter.

[00122] In some embodiments, battery 5518 periodically checks an actual battery capacity or energy with charge to a preset voltage (e.g., 4.1 V). The battery then decides of the battery capacity or energy. If the capacity or energy is insufficient, then an apparatus in or associated with the battery slightly increases charging voltage to a point where the capacity is sufficient (e.g. from 4.1 V to 4.11 V). The process of periodically checking and slightly

increase charging voltage is performed until charging voltage reaches specification limit (e.g., 4.2 V). The scheme described herein has benefits such as battery longevity can be extended, risk of insufficient energy reserve can be reduced, burst power can be used as long as possible, and/or even higher burst power can be used.

[00123] In some embodiments, the charging circuitry (e.g., 5518) comprises a buck-boost converter. This buck-boost converter comprises DrMOS or DrGaN devices used in place of half-bridges for traditional buck-boost converters. Various embodiments here are described with reference to DrMOS. However, the embodiments are applicable to DrGaN. The DrMOS devices allow for better efficiency in power conversion due to reduced parasitic and optimized MOSFET packaging. Since the dead-time management is internal to the DrMOS, the dead-time management is more accurate than for traditional buck-boost converters leading to higher efficiency in conversion. Higher frequency of operation allows for smaller inductor size, which in turn reduces the z-height of the charger comprising the DrMOS based buck-boost converter. The buck-boost converter of various embodiments comprises dual-folded bootstrap for DrMOS devices. In some embodiments, in addition to the traditional bootstrap capacitors, folded bootstrap capacitors are added that cross-couple inductor nodes to the two sets of DrMOS switches.

[00124] In some embodiments, device 5500 comprises Power Control Unit (PCU) 5510 (also referred to as Power Management Unit (PMU), Power Management Controller (PMC), Power Unit (p-unit), etc.). In an example, some sections of PCU 5510 may be implemented by one or more processing cores 5508, and these sections of PCU 5510 are symbolically illustrated using a dotted box and labelled PCU 5510a. In an example, some other sections of PCU 5510 may be implemented outside the processing cores 5508, and these sections of PCU 5510 are symbolically illustrated using a dotted box and labelled as PCU 5510b. PCU 5510 may implement various power management operations for device 5500. PCU 5510 may include hardware interfaces, hardware circuitries, connectors, registers, etc., as well as software components (e.g., drivers, protocol stacks), to implement various power management operations for device 5500.

[00125] In various embodiments, PCU or PMU 5510 is organized in a hierarchical manner forming a hierarchical power management (HPM). HPM of various embodiments builds a capability and infrastructure that allows for package level management for the platform, while still catering to islands of autonomy that might exist across the constituent die in the package. HPM does not assume a pre-determined mapping of physical partitions to

domains. An HPM domain can be aligned with a function integrated inside a dielet, to a dielet boundary, to one or more dielets, to a companion die, or even a discrete CXL device. HPM addresses integration of multiple instances of the same die, mixed with proprietary functions or 3rd party functions integrated on the same die or separate die, and even accelerators connected via CXL (e.g., Flexbus) that may be inside the package, or in a discrete form factor.

[00126] HPM enables designers to meet the goals of scalability, modularity, and late binding. HPM also allows PMU functions that may already exist on other dice to be leveraged, instead of being disabled in the flat scheme. HPM enables management of any arbitrary collection of functions independent of their level of integration. HPM of various embodiments is scalable, modular, works with symmetric multi-chip processors (MCPs), and works with asymmetric MCPs. For example, HPM does not need a signal PM controller and package infrastructure to grow beyond reasonable scaling limits. HPM enables late addition of a die in a package without the need for change in the base die infrastructure. HPM addresses the need of disaggregated solutions having dies of different process technology nodes coupled in a single package. HPM also addresses the needs of companion die integration solutions—on and off package.

[00127] In various embodiments, each die (or dielet) includes a power management unit (PMU) or p-unit. For example, processor dies can have a supervisor p-unit, supervisee p-unit, or a dual role supervisor/supervisee p-unit. In some embodiments, an I/O die has its own dual role p-unit such as supervisor and/or supervisee p-unit. The p-units in each die can be instances of a generic p-unit. In one such example, all p-units have the same capability and circuits, but are configured (dynamically or statically) to take a role of a supervisor, supervisee, and/or both. In some embodiments, the p-units for compute dies are instances of a compute p-unit while p-units for IO dies are instances of an IO p-unit different from the compute p-unit. Depending on the role, p-unit acquires specific responsibilities to manage power of the multichip module and/or computing platform. While various p-units are described for dies in a multichip module or system-on-chip, a p-unit can also be part of an external device such as I/O device.

[00128] Here, the various p-units do not have to be the same. The HPM architecture can operate very different types of p-units. One common feature for the p-units is that they are expected to receive HPM messages and are expected to be able to comprehend them. In some embodiments, the p-unit of IO dies may be different than the p-unit of the compute dies.

For example, the number of register instances of each class of register in the IO p-unit is different than those in the p-units of the compute dies. An IO die has the capability of being an HPM supervisor for CXL connected devices, but compute die may not need to have that capability. The IO and computes dice also have different firmware flows and possibly different firmware images. These are choices that an implementation can make. An HPM architecture can choose to have one superset firmware image and selectively execute flows that are relevant to the die type the firmware is associated with. Alternatively, there can be a customer firmware for each p-unit type; it can allow for more streamlined sizing of the firmware storage requirements for each p-unit type.

[00129] The p-unit in each die can be configured as a supervisor p-unit, supervisee p-unit or with a dual role of supervisor/supervisee. As such, p-units can perform roles of supervisor or supervisee for various domains. In various embodiments, each instance of p-unit is capable of autonomously managing local dedicated resources and contains structures to aggregate data and communicate between instances to enable shared resource management by the instance configured as the shared resource supervisor. A message and wire-based infrastructure is provided that can be duplicated and configured to facilitate management and flows between multiple p-units.

[00130] In some embodiments, power and thermal thresholds are communicated by a supervisor p-unit to supervisee p-units. For example, a supervisor p-unit learns of the workload (present and future) of each die, power measurements of each die, and other parameters (e.g., platform level power boundaries) and determines new power limits for each die. These power limits are then communicated by supervisor p-units to the supervisee p-units via one or more interconnects and fabrics. In some embodiments, a fabric indicates a group of fabrics and interconnect including a first fabric, a second fabric, and a fast response interconnect. In some embodiments, the first fabric is used for common communication between a supervisor p-unit and a supervisee p-unit. These common communications include change in voltage, frequency, and/or power state of a die which is planned based on a number of factors (e.g., future workload, user behavior, etc.). In some embodiments, the second fabric is used for higher priority communication between supervisor p-unit and supervisee p-unit. Example of higher priority communication include a message to throttle because of a possible thermal runaway condition, reliability issue, etc. In some embodiments, a fast response interconnect is used for communicating fast or hard throttle of all dies. In this case, a supervisor p-unit may send a fast throttle message to all other p-units, for example. In some

embodiments, a fast response interconnect is a legacy interconnect whose function can be performed by the second fabric.

[00131] The HPM architecture of various embodiments enables scalability, modularity, and late binding of symmetric and/or asymmetric dies. Here, symmetric dies are dies of same size, type, and/or function, while asymmetric dies are dies of different size, type, and/or function. Hierarchical approach also allows PMU functions that may already exist on other dice to be leveraged, instead of being disabled in the traditional flat power management scheme. HPM does not assume a pre-determined mapping of physical partitions to domains. An HPM domain can be aligned with a function integrated inside a dielet, to a dielet boundary, to one or more dielets, to a companion die, or even a discrete CXL device. HPM enables management of any arbitrary collection of functions independent of their level of integration. In some embodiments, a p-unit is declared a supervisor p-unit based on one or more factors. These factors include memory size, physical constraints (e.g., number of pin-outs), and locations of sensors (e.g., temperature, power consumption, etc.) to determine physical limits of the processor.

[00132] The HPM architecture of various embodiments, provides a means to scale power management so that a single p-unit instance does not need to be aware of the entire processor. This enables power management at a smaller granularity and improves response times and effectiveness. Hierarchical structure maintains a monolithic view to the user. For example, at an operating system (OS) level, HPM architecture gives the OS a single PMU view even though the PMU is physically distributed in one or more supervisor-supervisee configurations.

[00133] In some embodiments, the HPM architecture is centralized where one supervisor controls all supervisees. In some embodiments, the HPM architecture is decentralized, wherein various p-units in various dies control overall power management by peer-to-peer communication. In some embodiments, the HPM architecture is distributed where there are different supervisors for different domains. One example of a distributed architecture is a tree-like architecture.

[00134] In some embodiments, device 5500 comprises Power Management Integrated Circuit (PMIC) 5512, e.g., to implement various power management operations for device 5500. In some embodiments, PMIC 5512 is a Reconfigurable Power Management ICs (RPMICs) and/or an IMVP (Intel® Mobile Voltage Positioning). In an example, the PMIC is within an IC die separate from processor 5504. The may implement various power

management operations for device 5500. PMIC 5512 may include hardware interfaces, hardware circuitries, connectors, registers, etc., as well as software components (e.g., drivers, protocol stacks), to implement various power management operations for device 5500.

[00135] In an example, device 5500 comprises one or both PCU 5510 or PMIC 5512. In an example, any one of PCU 5510 or PMIC 5512 may be absent in device 5500, and hence, these components are illustrated using dotted lines.

[00136] Various power management operations of device 5500 may be performed by PCU 5510, by PMIC 5512, or by a combination of PCU 5510 and PMIC 5512. For example, PCU 5510 and/or PMIC 5512 may select a power state (e.g., P-state) for various components of device 5500. For example, PCU 5510 and/or PMIC 5512 may select a power state (e.g., in accordance with the ACPI (Advanced Configuration and Power Interface) specification) for various components of device 5500. Merely as an example, PCU 5510 and/or PMIC 5512 may cause various components of the device 5500 to transition to a sleep state, to an active state, to an appropriate C state (e.g., C0 state, or another appropriate C state, in accordance with the ACPI specification), etc. In an example, PCU 5510 and/or PMIC 5512 may control a voltage output by VR 5514 and/or a frequency of a clock signal output by the clock generator, e.g., by outputting the VID signal and/or the FID signal, respectively. In an example, PCU 5510 and/or PMIC 5512 may control battery power usage, charging of battery 5518, and features related to power saving operation.

[00137] The clock generator 5516 can comprise a phase locked loop (PLL), frequency locked loop (FLL), or any suitable clock source. In some embodiments, each core of processor 5504 has its own clock source. As such, each core can operate at a frequency independent of the frequency of operation of the other core. In some embodiments, PCU 5510 and/or PMIC 5512 performs adaptive or dynamic frequency scaling or adjustment. For example, clock frequency of a processor core can be increased if the core is not operating at its maximum power consumption threshold or limit. In some embodiments, PCU 5510 and/or PMIC 5512 determines the operating condition of each core of a processor, and opportunistically adjusts frequency and/or power supply voltage of that core without the core clocking source (e.g., PLL of that core) losing lock when the PCU 5510 and/or PMIC 5512 determines that the core is operating below a target performance level. For example, if a core is drawing current from a power supply rail less than a total current allocated for that core or processor 5504, then PCU 5510 and/or PMIC 5512 can temporality increase the power draw for that core or processor 5504 (e.g., by increasing clock frequency and/or power supply

voltage level) so that the core or processor 5504 can perform at higher performance level. As such, voltage and/or frequency can be increased temporality for processor 5504 without violating product reliability.

[00138] In an example, PCU 5510 and/or PMIC 5512 may perform power management operations, e.g., based at least in part on receiving measurements from power measurement circuitries 5542, temperature measurement circuitries 5540, charge level of battery 5518, and/or any other appropriate information that may be used for power management. To that end, PMIC 5512 is communicatively coupled to one or more sensors to sense/detect various values/variations in one or more factors having an effect on power/thermal behavior of the system/platform. Examples of the one or more factors include electrical current, voltage droop, temperature, operating frequency, operating voltage, power consumption, inter-core communication activity, etc. One or more of these sensors may be provided in physical proximity (and/or thermal contact/coupling) with one or more components or logic/IP blocks of a computing system. Additionally, sensor(s) may be directly coupled to PCU 5510 and/or PMIC 5512 in at least one embodiment to allow PCU 5510 and/or PMIC 5512 to manage processor core energy at least in part based on value(s) detected by one or more of the sensors.

[00139] Also illustrated is an example software stack of device 5500 (although not all elements of the software stack are illustrated). Merely as an example, processors 5504 may execute application programs 5550, Operating System 5552, one or more Power Management (PM) specific application programs (e.g., generically referred to as PM applications 5558), and/or the like. PM applications 5558 may also be executed by the PCU 5510 and/or PMIC 5512. OS 5552 may also include one or more PM applications 5556a, 5556b, 5556c. The OS 5552 may also include various drivers 5554a, 5554b, 5554c, etc., some of which may be specific for power management purposes. In some embodiments, device 5500 may further comprise a Basic Input/output System (BIOS) 5520. BIOS 5520 may communicate with OS 5552 (e.g., via one or more drivers 5554), communicate with processors 5504, etc.

[00140] For example, one or more of PM applications 5558, 5556, drivers 5554, BIOS 5520, etc. may be used to implement power management specific tasks, e.g., to control voltage and/or frequency of various components of device 5500, to control wake-up state, sleep state, and/or any other appropriate power state of various components of device 5500, control battery power usage, charging of the battery 5518, features related to power saving operation, etc.

[00141] In some embodiments, battery 5518 is a Li-metal battery with a pressure chamber to allow uniform pressure on a battery. The pressure chamber is supported by metal plates (such as pressure equalization plate) used to give uniform pressure to the battery. The pressure chamber may include pressured gas, elastic material, spring plate, etc. The outer skin of the pressure chamber is free to bow, restrained at its edges by (metal) skin, but still exerts a uniform pressure on the plate that is compressing the battery cell. The pressure chamber gives uniform pressure to battery, which is used to enable high-energy density battery with, for example, 20% more battery life.

[00142] In some embodiments, battery 5518 includes hybrid technologies. For example, a mix of high energy density charge (e.g., Li-Ion batteries) carrying device(s) and low energy density charge carrying devices (e.g., supercapacitor) are used as batteries or storage devices. In some embodiments, a controller (e.g., hardware, software, or a combination of them) is used analyze peak power patterns and minimizes the impact to overall lifespan of high energy density charge carrying device-based battery cells while maximizing service time for peak power shaving feature. The controller may be part of battery 5518 or part of p-unit 5510b.

[00143] In some embodiments, pCode executing on PCU 5510a/b has a capability to enable extra compute and telemetries resources for the runtime support of the pCode. Here pCode refers to a firmware executed by PCU 5510a/b to manage performance of the SoC 5501. For example, pCode may set frequencies and appropriate voltages for the processor. Part of the pCode are accessible via OS 5552. In various embodiments, mechanisms and methods are provided that dynamically change an Energy Performance Preference (EPP) value based on workloads, user behavior, and/or system conditions. There may be a well-defined interface between OS 5552 and the pCode. The interface may allow or facilitate the software configuration of several parameters and/or may provide hints to the pCode. As an example, an EPP parameter may inform a pCode algorithm as to whether performance or battery life is more important.

[00144] This support may be done as well by the OS 5552 by including machine-learning support as part of OS 5552 and either tuning the EPP value that the OS hints to the hardware (e.g., various components of SoC 5501) by machine-learning prediction, or by delivering the machine-learning prediction to the pCode in a manner similar to that done by a Dynamic Tuning Technology (DTT) driver. In this model, OS 5552 may have visibility to the same set of telemetries as are available to a DTT. As a result of a DTT machine-learning

hint setting, pCode may tune its internal algorithms to achieve optimal power and performance results following the machine-learning prediction of activation type. The pCode as example may increase the responsibility for the processor utilization change to enable fast response for user activity, or may increase the bias for energy saving either by reducing the responsibility for the processor utilization or by saving more power and increasing the performance lost by tuning the energy saving optimization. This approach may facilitate saving more battery life in case the types of activities enabled lose some performance level over what the system can enable. The pCode may include an algorithm for dynamic EPP that may take the two inputs, one from OS 5552 and the other from software such as DTT, and may selectively choose to provide higher performance and/or responsiveness. As part of this method, the pCode may enable in the DTT an option to tune its reaction for the DTT for different types of activity.

[00145] In some embodiments, pCode improves the performance of the SoC in battery mode. In some embodiments, pCode allows drastically higher SoC peak power limit levels (and thus higher Turbo performance) in battery mode. In some embodiments, pCode implements power throttling and is part of Intel's Dynamic Tuning Technology (DTT). In various embodiments, the peak power limit is referred to PL4. However, the embodiments are applicable to other peak power limits. In some embodiments, pCode sets the *Vth* threshold voltage (the voltage level at which the platform will throttle the SoC) in such a way as to prevent the system from unexpected shutdown (or black screening). In some embodiments, pCode calculates the *Psoc,pk* SoC Peak Power Limit (e.g., PL4), according to the threshold voltage (*Vth*). These are two dependent parameters, if one is set, the other can be calculated. pCode is used to optimally set one parameter (*Vth*) based on the system parameters, and the history of the operation. In some embodiments, pCode provides a scheme to dynamically calculate the throttling level (*Psoc,th*) based on the available battery power (which changes slowly) and set the SoC throttling peak power (*Psoc,th*). In some embodiments, pCode decides the frequencies and voltages based on *Psoc,th*. In this case, throttling events have less negative effect on the SoC performance. Various embodiments provide a scheme which allows maximum performance (Pmax) framework to operate.

[00146] In some embodiments, VR 5514 includes a current sensor to sense and/or measure current through a high-side switch of VR 5514. In some embodiments the current sensor uses an amplifier with capacitively coupled inputs in feedback to sense the input offset of the amplifier, which can be compensated for during measurement. In some embodiments,

the amplifier with capacitively coupled inputs in feedback is used to operate the amplifier in a region where the input common-mode specifications are relaxed, so that the feedback loop gain and/or bandwidth is higher. In some embodiments, the amplifier with capacitively coupled inputs in feedback is used to operate the sensor from the converter input voltage by employing high-PSRR (power supply rejection ratio) regulators to create a local, clean supply voltage, causing less disruption to the power grid in the switch area. In some embodiments, a variant of the design can be used to sample the difference between the input voltage and the controller supply, and recreate that between the drain voltages of the power and replica switches. This allows the sensor to not be exposed to the power supply voltage. In some embodiments, the amplifier with capacitively coupled inputs in feedback is used to compensate for power delivery network related (PDN-related) changes in the input voltage during current sensing.

[00147] Some embodiments use three components to adjust the peak power of SoC 5501 based on the states of a USB TYPE-C device 5529. These components include OS Peak Power Manager (part of OS 5552), USB TYPE-C Connector Manager (part of OS 5552), and USB TYPE-C Protocol Device Driver (e.g., one of drivers 5554a, 5554b, 5554c). In some embodiments, the USB TYPE-C Connector Manager sends a synchronous request to the OS Peak Power Manager when a USB TYPE-C power sink device is attached or detached from SoC 5501, and the USB TYPE-C Protocol Device Driver sends a synchronous request to the Peak Power Manager when the power sink transitions device state. In some embodiments, the Peak Power Manager takes power budget from the CPU when the USB TYPE-C connector is attached to a power sink and is active (e.g., high power device state). In some embodiments, the Peak Power Manager gives back the power budget to the CPU for performance when the USB TYPE-C connector is either detached or the attached and power sink device is idle (lowest device state).

[00148] In some embodiments, logic is provided to dynamically pick the best operating processing core for BIOS power-up flows and sleep exit flows (e.g., S3, S4, and/or S5). The selection of the bootstrap processor (BSP) is moved to an early power-up time instead of a fixed hardware selection at any time. For maximum boot performance, the logic selects the fastest capable core as the BSP at an early power-up time. In addition, for maximum power saving, the logic selects the most power efficient core as the BSP. Processor or switching for selecting the BSP happens during the boot-up as well as power-up flows (e.g., S3, S4, and/or S5 flows).

[00149] In some embodiments, the memories herein are organized in multi-level memory architecture and their performance is governed by a decentralized scheme. The decentralized scheme includes p-unit 5510 and memory controllers. In some embodiments, the scheme dynamically balances a number of parameters such as power, thermals, cost, latency and performance for memory levels that are progressively further away from the processor in platform 5500 based on how applications are using memory levels that are further away from processor cores. In some examples, the decision making for the state of the far memory (FM) is decentralized. For example, a processor power management unit (p-unit), near memory controller (NMC), and/or far memory host controller (FMHC) makes decisions about the power and/or performance state of the FM at their respective levels. These decisions are coordinated to provide the most optimum power and/or performance state of the FM for a given time. The power and/or performance state of the memories adaptively change to changing workloads and other parameters even when the processor(s) is in a particular power state.

[00150] In some embodiments, a hardware and software coordinated processor power state policy (e.g., policy for C-state) is implemented that delivers optimal power state selection by taking in to account the performance and/or responsiveness needs of thread expected to be scheduled on the core entering idle, to achieve improved instructions per cycle (IPC) and performance for cores running user critical tasks. The scheme provides the ability to deliver responsiveness gains for important and/or user-critical threads running on a system-on-chip. P-unit 5510 which coupled to the plurality of processing cores, receives a hint from operating system 5552 indicative of a bias towards a power state or performance state for at least one of the processing cores of the plurality of processing cores based on a priority of a thread in context switch.

[00151] Reference in the specification to "an embodiment," "one embodiment," "some embodiments," or "other embodiments" means that a particular feature, structure, or characteristic described in connection with the embodiments is included in at least some embodiments, but not necessarily all embodiments. The various appearances of "an embodiment," "one embodiment," or "some embodiments" are not necessarily all referring to the same embodiments. If the specification states a component, feature, structure, or characteristic "may," "might," or "could" be included, that particular component, feature, structure, or characteristic is not required to be included. If the specification or claim refers to "a" or "an" element, that does not mean there is only one of the elements. If the

specification or claims refer to "an additional" element, that does not preclude there being more than one of the additional elements.

[00152] Throughout the specification, and in the claims, the term "connected" means a direct connection, such as electrical, mechanical, or magnetic connection between the things that are connected, without any intermediary devices.

[00153] The term "coupled" means a direct or indirect connection, such as a direct electrical, mechanical, or magnetic connection between the things that are connected or an indirect connection, through one or more passive or active intermediary devices.

[00154] The term “adjacent” here generally refers to a position of a thing being next to (e.g., immediately next to or close to with one or more things between them) or adjoining another thing (e.g., abutting it).

[00155] The term "circuit" or “module” may refer to one or more passive and/or active components that are arranged to cooperate with one another to provide a desired function.

[00156] The term "signal" may refer to at least one current signal, voltage signal, magnetic signal, or data/clock signal. The meaning of "a," "an," and "the" include plural references. The meaning of "in" includes "in" and "on."

[00157] The term “analog signal” is any continuous signal for which the time varying feature (variable) of the signal is a representation of some other time varying quantity, i.e., analogous to another time varying signal.

[00158] The term “digital signal” is a physical signal that is a representation of a sequence of discrete values (a quantified discrete-time signal), for example of an arbitrary bit stream, or of a digitized (sampled and analog-to-digital converted) analog signal.

[00159] The term “scaling” generally refers to converting a design (schematic and layout) from one process technology to another process technology and may be subsequently being reduced in layout area. In some cases, scaling also refers to upsizing a design from one process technology to another process technology and may be subsequently increasing layout area. The term “scaling” generally also refers to downsizing or upsizing layout and devices within the same technology node. The term “scaling” may also refer to adjusting (e.g., slowing down or speeding up – i.e. scaling down, or scaling up respectively) of a signal frequency relative to another parameter, for example, power supply level.

[00160] The terms “substantially,” “close,” “approximately,” “near,” and “about,” generally refer to being within +/- 10% of a target value.

[00161] Unless otherwise specified the use of the ordinal adjectives "first," "second," and "third," etc., to describe a common object, merely indicate that different instances of like objects are being referred to and are not intended to imply that the objects so described must be in a given sequence, either temporally, spatially, in ranking or in any other manner.

[00162] For the purposes of the present disclosure, phrases "A and/or B" and "A or B" mean (A), (B), or (A and B). For the purposes of the present disclosure, the phrase "A, B, and/or C" means (A), (B), (C), (A and B), (A and C), (B and C), or (A, B and C).

[00163] The terms "left," "right," "front," "back," "top," "bottom," "over," "under," and the like in the description and in the claims, if any, are used for descriptive purposes and not necessarily for describing permanent relative positions.

[00164] It is pointed out that those elements of the figures having the same reference numbers (or names) as the elements of any other figure can operate or function in any manner similar to that described but are not limited to such.

[00165] For purposes of the embodiments, the transistors in various circuits and logic blocks described here are metal oxide semiconductor (MOS) transistors or their derivatives, where the MOS transistors include drain, source, gate, and bulk terminals. The transistors and/or the MOS transistor derivatives also include Tri-Gate and FinFET transistors, Gate All Around Cylindrical Transistors, Tunneling FET (TFET), Square Wire, or Rectangular Ribbon Transistors, ferroelectric FET (FeFETs), or other devices implementing transistor functionality like carbon nanotubes or spintronic devices. MOSFET symmetrical source and drain terminals i.e., are identical terminals and are interchangeably used here. A TFET device, on the other hand, has asymmetric Source and Drain terminals. Those skilled in the art will appreciate that other transistors, for example, Bi-polar junction transistors (BJT PNP/NPN), BiCMOS, CMOS, etc., may be used without departing from the scope of the disclosure.

[00166] Here the term "die" generally refers to a single continuous piece of semiconductor material (e.g. silicon) where transistors or other components making up a processor core may reside. Multi-core processors may have two or more processors on a single die, but alternatively, the two or more processors may be provided on two or more respective dies. Each die has a dedicated power controller or power control unit (p-unit) power controller or power control unit (p-unit) which can be dynamically or statically configured as a supervisor or supervisee. In some examples, dies are of the same size and functionality i.e., symmetric cores. However, dies can also be asymmetric. For example,

some dies have different size and/or function than other dies. Each processor may also be a dielet or chiplet.

[00167] Here the term "dielet" or "chiplet" generally refers to a physically distinct semiconductor die, typically connected to an adjacent die in a way that allows the fabric across a die boundary to function like a single fabric rather than as two distinct fabrics. Thus at least some dies may be dielets. Each dielet may include one or more p-units which can be dynamically or statically configured as a supervisor, supervisee or both.

[00168] Here the term "fabric" generally refers to communication mechanism having a known set of sources, destinations, routing rules, topology and other properties. The sources and destinations may be any type of data handling functional unit such as power management units. Fabrics can be two-dimensional spanning along an x-y plane of a die and/or three-dimensional (3D) spanning along an x-y-z plane of a stack of vertical and horizontally positioned dies. A single fabric may span multiple dies. A fabric can take any topology such as mesh topology, star topology, daisy chain topology. A fabric may be part of a network-on-chip (NoC) with multiple agents. These agents can be any functional unit.

[00169] Here, the term "processor core" generally refers to an independent execution unit that can run one program thread at a time in parallel with other cores. A processor core may include a dedicated power controller or power control unit (p-unit) which can be dynamically or statically configured as a supervisor or supervisee. This dedicated p-unit is also referred to as an autonomous p-unit, in some examples. In some examples, all processor cores are of the same size and functionality i.e., symmetric cores. However, processor cores can also be asymmetric. For example, some processor cores have different size and/or function than other processor cores. A processor core can be a virtual processor core or a physical processor core.

[00170] Here, the term "interconnect" refers to a communication link, or channel, between two or more points or nodes. It may comprise one or more separate conduction paths such as wires, vias, waveguides, passive components, and/or active components. It may also comprise a fabric. In some embodiments, a p-unit is coupled to an OS via an interface.

[00171] Here the term "interface" generally refers to software and/or hardware used to communicate with an interconnect. An interface may include logic and I/O driver/receiver to send and receive data over the interconnect or one or more wires.

[00172] Here the term "domain" generally refers to a logical or physical perimeter that has similar properties (e.g., supply voltage, operating frequency, type of circuits or logic, and/or workload type) and/or is controlled by a particular agent. For example, a domain may be a group of logic units or function units that are controlled by a particular supervisor. A domain may also be referred to an Autonomous Perimeter (AP). A domain can be an entire system-on-chip (SoC) or part of the SoC, and is governed by a p-unit.

[00173] Here the term "supervisor" generally refers to a power controller, or power management, unit (a "p-unit"), which monitors and manages power and performance related parameters for one or more associated power domains, either alone or in cooperation with one or more other p-units. Power/performance related parameters may include but are not limited to domain power, platform power, voltage, voltage domain current, die current, load-line, temperature, device latency, utilization, clock frequency, processing efficiency, current/future workload information, and other parameters. It may determine new power or performance parameters (limits, average operational, etc.) for the one or more domains. These parameters may then be communicated to supervisee p-units, or directly to controlled or monitored entities such as VR or clock throttle control registers, via one or more fabrics and/or interconnects. A supervisor learns of the workload (present and future) of one or more dies, power measurements of the one or more dies, and other parameters (e.g., platform level power boundaries) and determines new power limits for the one or more dies. These power limits are then communicated by supervisor p-units to the supervisee p-units via one or more fabrics and/or interconnect. In examples where a die has one p-unit, a supervisor (Svor) p-unit is also referred to as supervisor die.

[00174] Here the term "supervisee" generally refers to a power controller, or power management, unit (a "p-unit"), which monitors and manages power and performance related parameters for one or more associated power domains, either alone or in cooperation with one or more other p-units and receives instructions from a supervisor to set power and/or performance parameters (e.g., supply voltage, operating frequency, maximum current, throttling threshold, etc.) for its associated power domain. In examples where a die has one p-unit, a supervisee (Svee) p-unit may also be referred to as a supervisee die. Note that a p-unit may serve either as a Svor, a Svee, or both a Svor/Svee p-unit.

[00175] Furthermore, the particular features, structures, functions, or characteristics may be combined in any suitable manner in one or more embodiments. For example, a first embodiment may be combined with a second embodiment anywhere the particular features,

structures, functions, or characteristics associated with the two embodiments are not mutually exclusive.

[00176] While the disclosure has been described in conjunction with specific embodiments thereof, many alternatives, modifications and variations of such embodiments will be apparent to those of ordinary skill in the art in light of the foregoing description. The embodiments of the disclosure are intended to embrace all such alternatives, modifications, and variations as to fall within the broad scope of the appended claims.

[00177] In addition, well-known power/ground connections to integrated circuit (IC) chips and other components may or may not be shown within the presented figures, for simplicity of illustration and discussion, and so as not to obscure the disclosure. Further, arrangements may be shown in block diagram form in order to avoid obscuring the disclosure, and also in view of the fact that specifics with respect to implementation of such block diagram arrangements are highly dependent upon the platform within which the present disclosure is to be implemented (i.e., such specifics should be well within purview of one skilled in the art). Where specific details (e.g., circuits) are set forth in order to describe example embodiments of the disclosure, it should be apparent to one skilled in the art that the disclosure can be practiced without, or with variation of, these specific details. The description is thus to be regarded as illustrative instead of limiting.

[00178] The following examples pertain to further embodiments. Specifics in the examples may be used anywhere in one or more embodiments. All optional features of the apparatus described herein may also be implemented with respect to a method or process. The examples can be combined in any combinations. For example, example 4 can be combined with example 2.

[00179] Example 1: An apparatus comprising: a plurality of battery systems including a first battery system and a second battery system; a plurality of loads including a first load and a second load; a power source to power the plurality of loads; and a power bus switchably coupled to the plurality of battery systems, wherein the first battery system is to provide backup power to the first load, wherein the second battery system is to provide backup power to the second load, wherein first and/or second battery systems is to supply power via the power bus to the first load during power demand by the first load.

[00180] Example 2: The apparatus of example 1, wherein first and/or second battery system is to discharge to a low threshold to supply power during the power demand.

[00181] Example 3: The apparatus of example 1 comprising a data bus coupled to the plurality of battery systems.

[00182] Example 4: The apparatus of example 1, wherein the first battery system comprises: a battery comprising one or more battery cells; a charger to charge the one or more battery cells; a first switch to couple the battery to the first load; a second switch to couple the power bus to the first load; and a controller to control the first and second switches.

[00183] Example 5: The apparatus of example 4, wherein the controller is to monitor power consumption of the first and second loads, and to monitor real-time capacity of the first and second battery systems.

[00184] Example 6: The apparatus of example 4, wherein the controller is to provide higher charging priority to one of the first or second battery systems based on workload priority of the first or second loads.

[00185] Example 7: The apparatus of example 4, wherein the battery cells include a high energy density storage bank and a low energy density storage bank.

[00186] Example 8: The apparatus of example 7, wherein the high energy density storage bank comprises Li-ion battery cells.

[00187] Example 9: The apparatus of example 7, wherein the low energy density storage bank comprises super capacitor.

[00188] Example 10: The apparatus of example 7, wherein the controller is to enable the low energy density storage bank to provide power to the first load when the first load demands a sudden power.

[00189] Example 11: The apparatus of example 7, wherein the controller is to enable the high energy density storage bank to provide power to the first load when the first load demands a peak power.

[00190] Example 12: The apparatus of example 7, wherein the controller is to charge the low energy density storage bank before the high energy density storage bank is charged.

[00191] Example 13: The apparatus of example 7, wherein the controller is to enable the low energy density storage bank as a primary battery source, and is to enable the high energy density storage bank as a secondary battery source, to provide power to the first load when the first load demands a sudden power.

[00192] Example 14: The apparatus of example 1, wherein the first and second loads including first and second racks of servers.

[00193] Example 15: The apparatus of example 1, wherein the power demand is a peak power shaving demand.

[00194] Example 16: A system comprising: a plurality of racks of servers including a first rack and a second rack, wherein the first rack includes a first battery system, and wherein the second rack includes a second battery system; and a power bus coupled to the first and second battery systems, wherein the upon peak power demand by the first rack, the second battery system is supplement power via the power bus to the first rack if the first battery system has insufficient charge.

[00195] Example 17: The system of example 16, wherein one of the first or second battery systems assumes a role of a controller to monitor power consumption of the plurality of racks and to monitor real-time capacity of the first and second battery systems.

[00196] Example 18: The system of example 17, wherein the controller is to prioritize grid power budget to one of the first or second battery systems based on workload priority by servers of the first or second rack.

[00197] Example 19: An apparatus comprising: a load; and a first battery system which is operable to provide backup power to a load when AC power to the load is out, and to route power from a second battery system to the load when the load demands a peak power.

[00198] Example 20: The apparatus of example 19, wherein the battery is a hybrid battery comprising a high energy battery storage bank, and a low energy battery storage bank.

[00199] An abstract is provided that will allow the reader to ascertain the nature and gist of the technical disclosure. The abstract is submitted with the understanding that it will not be used to limit the scope or meaning of the claims. The following claims are hereby incorporated into the detailed description, with each claim standing on its own as a separate embodiment.

CLAIMS

What is claimed is:

1. An apparatus comprising:

a plurality of battery systems including a first battery system and a second battery system;

a plurality of loads including a first load and a second load;

a power source to power the plurality of loads; and

a power bus switchably coupled to the plurality of battery systems, wherein the first battery system is to provide backup power to the first load, wherein the second battery system is to provide backup power to the second load, wherein first and/or second battery systems is to supply power via the power bus to the first load during power demand by the first load.

2. The apparatus of claim 1, wherein first and/or second battery system is to discharge to a low threshold to supply power during the power demand.

3. The apparatus of claim 1 comprising a data bus coupled to the plurality of battery systems.

4. The apparatus of claim 1, wherein the first battery system comprises:

a battery comprising one or more battery cells;

a charger to charge the one or more battery cells;

a first switch to couple the battery to the first load;

a second switch to couple the power bus to the first load; and

a controller to control the first and second switches.

5. The apparatus of claim 4, wherein the controller is to monitor power consumption of the first and second loads, and to monitor real-time capacity of the first and second battery systems.

6. The apparatus of claim 4, wherein the controller is to provide higher charging priority to one of the first or second battery systems based on workload priority of the first or second loads.

7. The apparatus of claim 4, wherein the battery cells include a high energy density storage bank and a low energy density storage bank.

8. The apparatus of claim 7, wherein the high energy density storage bank comprises Li-ion battery cells.

9. The apparatus of claim 7, wherein the low energy density storage bank comprises super capacitor.

10. The apparatus of claim 7, wherein the controller is to enable the low energy density storage bank to provide power to the first load when the first load demands a sudden power.

11. The apparatus of claim 7, wherein the controller is to enable the high energy density storage bank to provide power to the first load when the first load demands a peak power.

12. The apparatus of claim 7, wherein the controller is to charge the low energy density storage bank before the high energy density storage bank is charged.

13. The apparatus of claim 7, wherein the controller is to enable the low energy density storage bank as a primary battery source, and is to enable the high energy density storage bank as a secondary battery source, to provide power to the first load when the first load demands a sudden power.

14. The apparatus of claim 1, wherein the first and second loads including first and second racks of servers.

15. The apparatus of claim 1, wherein the power demand is a peak power shaving demand.

16. A system comprising:

a plurality of racks of servers including a first rack and a second rack, wherein the first rack includes a first battery system, and wherein the second rack includes a second battery system; and

a power bus coupled to the first and second battery systems, wherein the upon peak power demand by the first rack, the second battery system is supplement power via the power bus to the first rack if the first battery system has insufficient charge.

17. The system of claim 16, wherein one of the first or second battery systems assumes a role of a controller to monitor power consumption of the plurality of racks and to monitor real-time capacity of the first and second battery systems.

18. The system of claim 17, wherein the controller is to prioritize grid power budget to one of the first or second battery systems based on workload priority by servers of the first or second rack.

19. An apparatus comprising:

a load; and

a first battery system which is operable to provide backup power to a load when AC power to the load is out, and to route power from a second battery system to the load when the load demands a peak power.

20. The apparatus of claim 19, wherein the battery is a hybrid battery comprising a high energy battery storage bank, and a low energy battery storage bank.

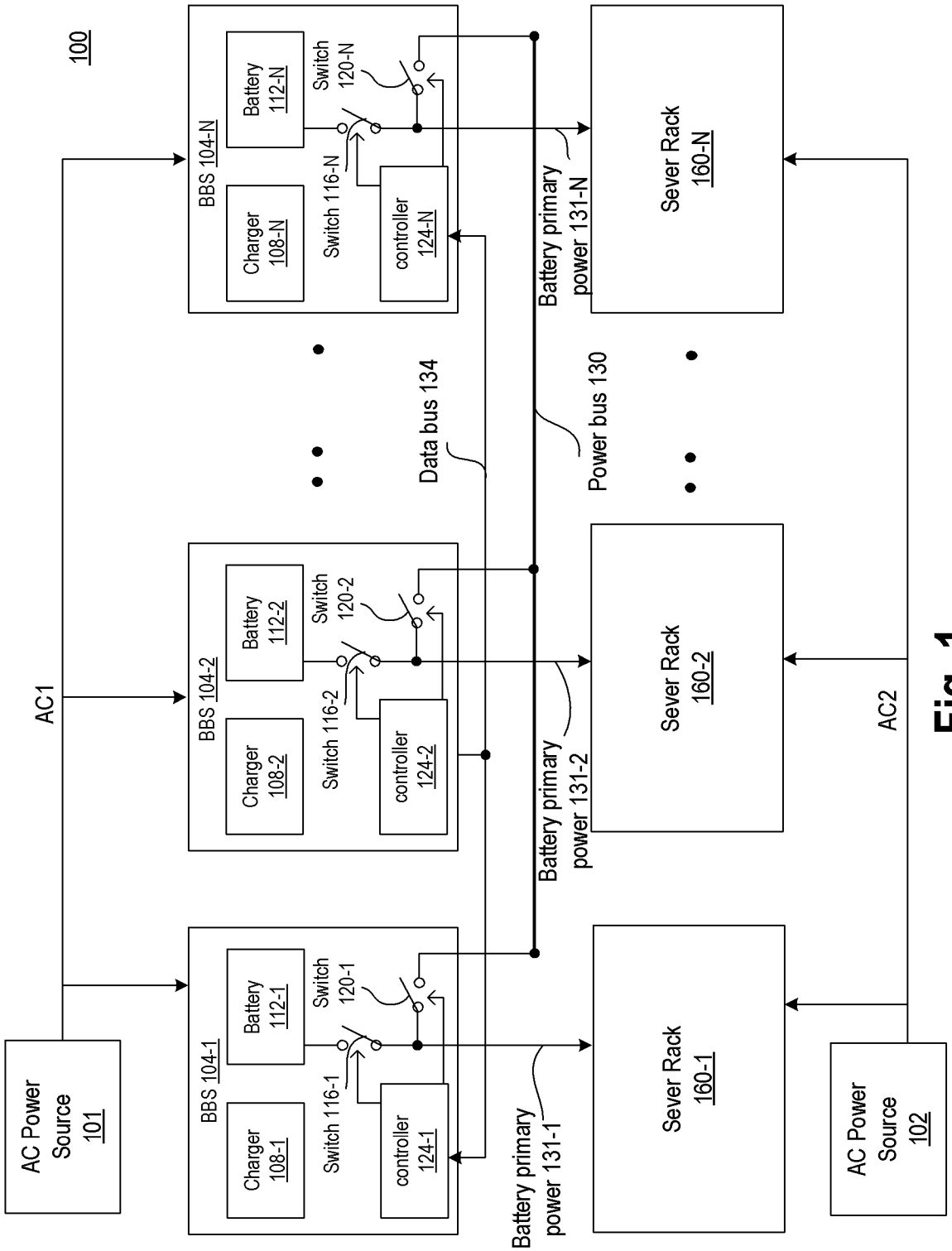

100
AC Power Source 101
AC1
BBS 104-1
Charger 108-1
Battery 112-1
Switch 120-1
Switch 116-1
controller 124-1
BBS 104-2
Charger 108-2
Battery 112-2
Switch 120-2
Switch 116-2
controller 124-2
BBS 104-N
Charger 108-N
Battery 112-N
Switch 120-N
Switch 116-N
controller 124-N
Data bus 134
Power bus 130
Battery primary power 131-1
Battery primary power 131-2
Battery primary power 131-N
Sever Rack 160-1
Sever Rack 160-2
Sever Rack 160-N
AC2
AC Power Source 102

Fig. 1

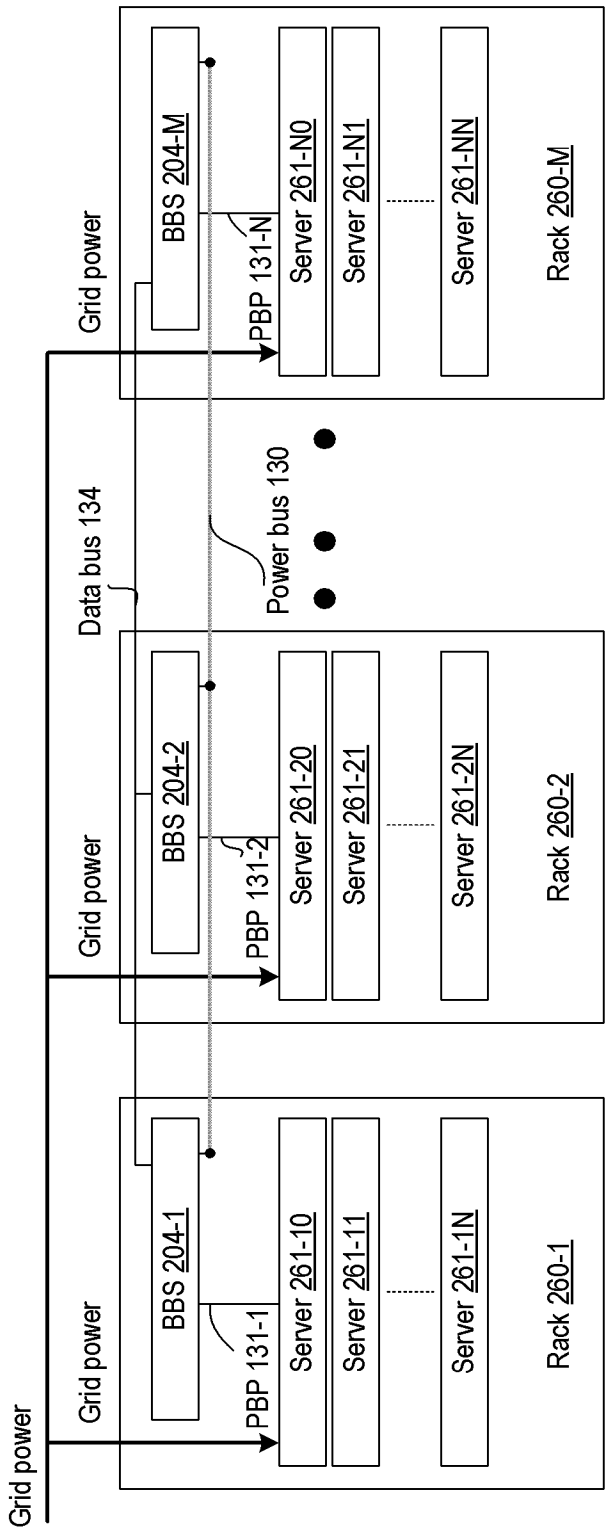

200
Grid power
Grid power
Grid power
Data bus 134
Power bus 130
BBS 204-1
PBP 131-1
Server 261-10
Server 261-11
Server 261-1N
Rack 260-1
BBS 204-2
PBP 131-2
Server 261-20
Server 261-21
Server 261-2N
Rack 260-2
BBS 204-M
PBP 131-N
Server 261-N0
Server 261-N1
Server 261-NN
Rack 260-M

Fig. 2

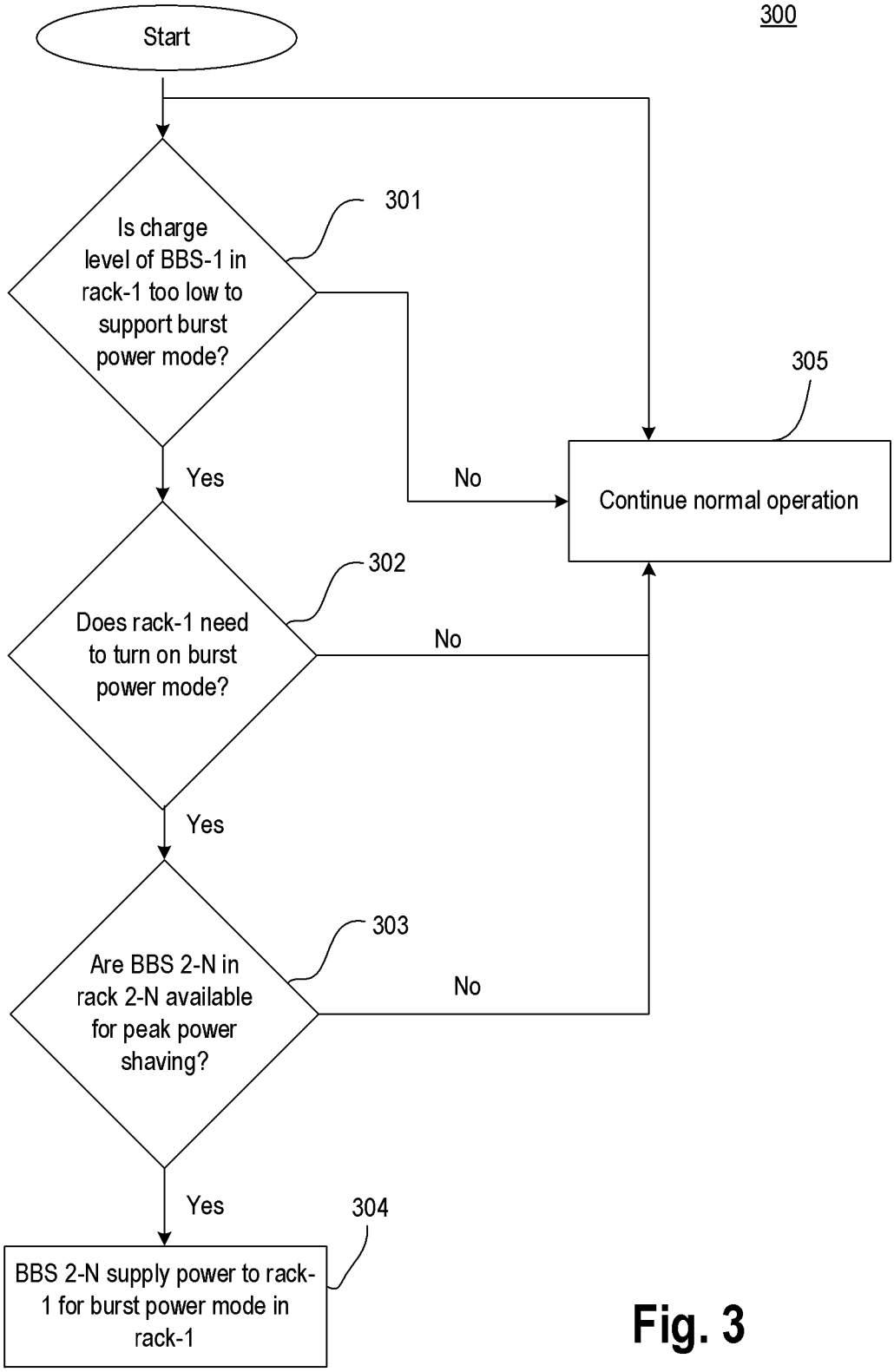

300
Start
301
Is charge level of BBS-1 in rack-1 too low to support burst power mode?
Yes
No
305
Continue normal operation
302
Does rack-1 need to turn on burst power mode?
No
Yes
303
Are BBS 2-N in rack 2-N available for peak power shaving?
No
Yes
304
BBS 2-N supply power to rack-1 for burst power mode in rack-1

Fig. 3

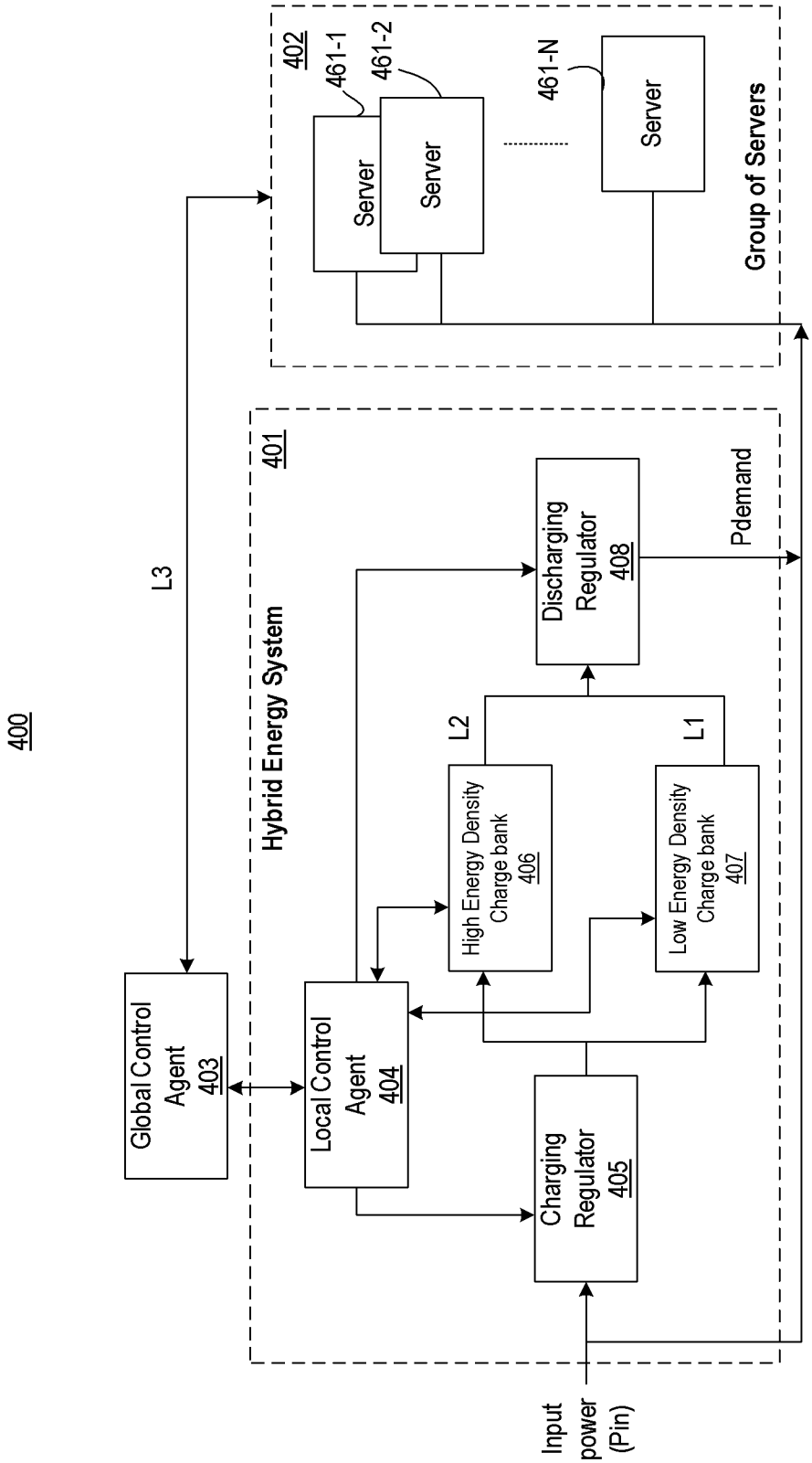

400
402
461-1
461-2
461-N
Server
Server
Server
Group of Servers
L3
401
Hybrid Energy System
Discharging Regulator 408
Pdemand
L2
L1
High Energy Density Charge bank 406
Low Energy Density Charge bank 407
Global Control Agent 403
Local Control Agent 404
Charging Regulator 405
Input power (Pin)

Fig. 4

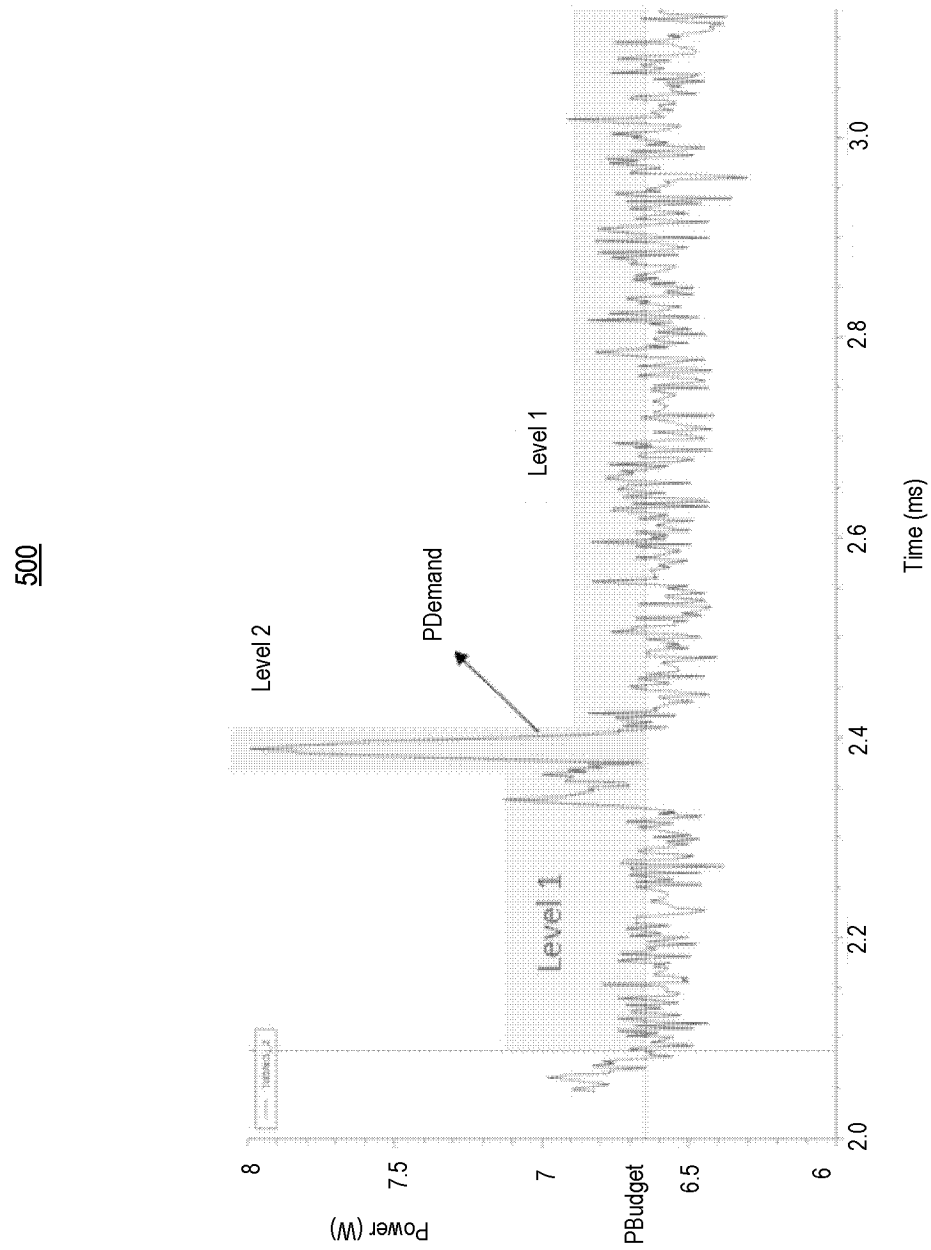

500
Level 2
PDemand
Level 1
Level 1
Power (W)
PBudget
8
7.5
7
6.5
6
2.0
2.2
2.4
2.6
2.8
3.0
Time (ms)

Fig. 5

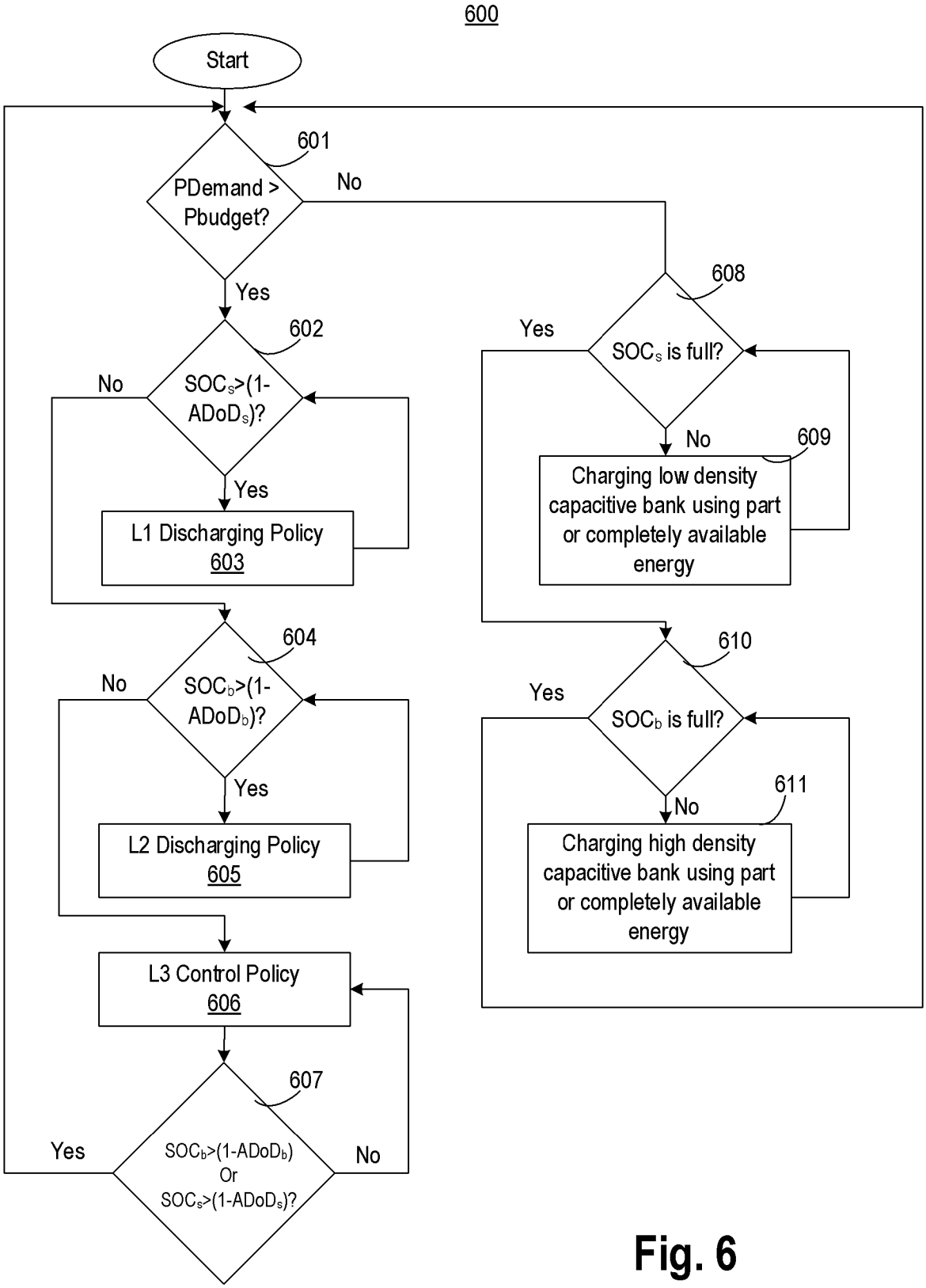

600
Start
601
PDemand > Pbudget?
No
Yes
602
SOCs>(1-ADoDs)?
No
Yes
L1 Discharging Policy 603
604
SOCb>(1-ADoDb)?
No
Yes
L2 Discharging Policy 605
L3 Control Policy 606
607
SOCb>(1-ADoDb) Or SOCs>(1-ADoDs)?
Yes
No
608
SOCs is full?
Yes
No
609
Charging low density capacitive bank using part or completely available energy
610
SOCb is full?
Yes
No
611
Charging high density capacitive bank using part or completely available energy

Fig. 6

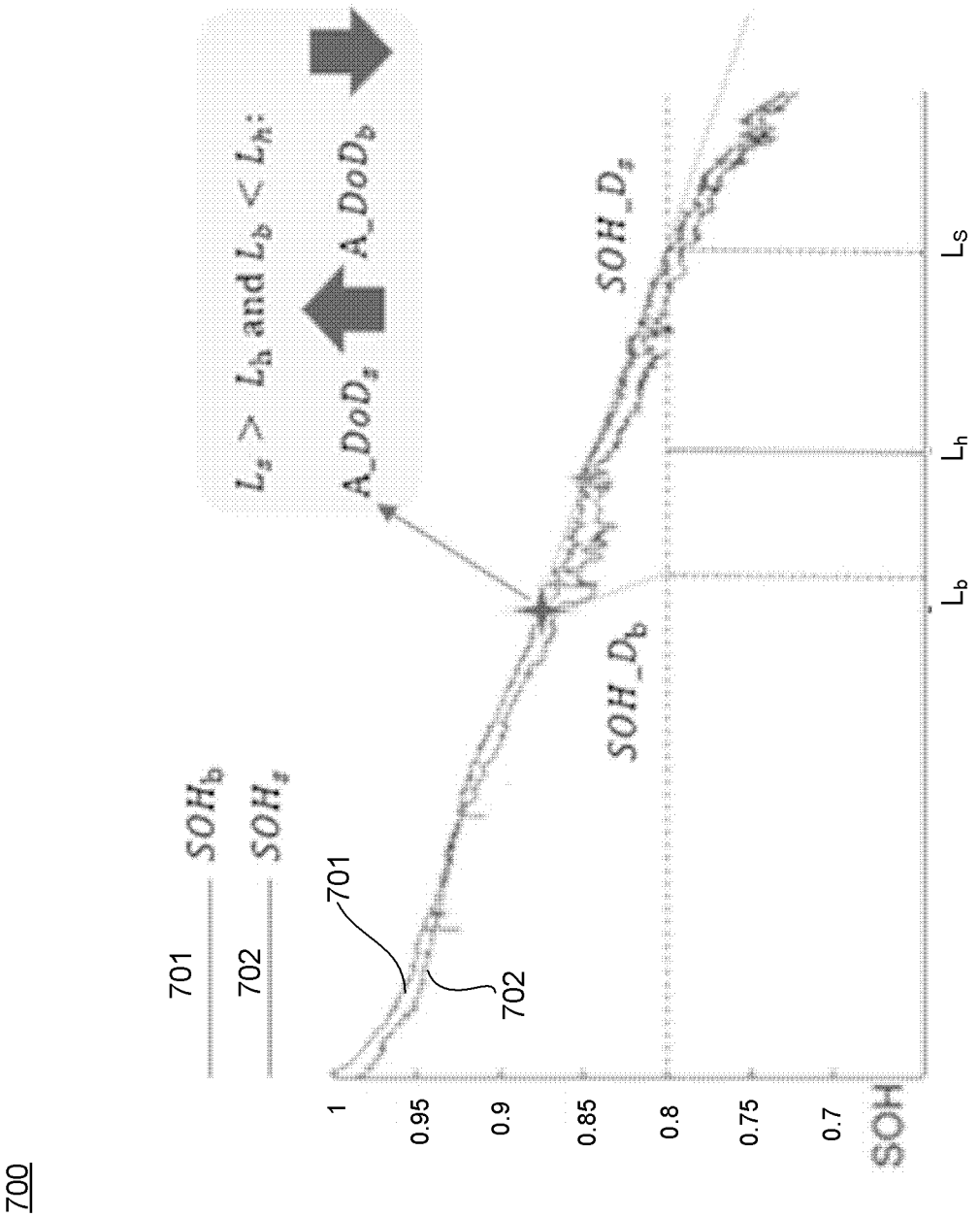

700
701
702
SOH_b
SOH_s
1
0.95
0.9
0.85
0.8
0.75
0.7
SOH
L_b
L_h
L_s
SOH_D_b
SOH_D_s
$L_s > L_h$ and $L_b < L_h$:
A_DoD_s
A_DoD_b

Fig. 7

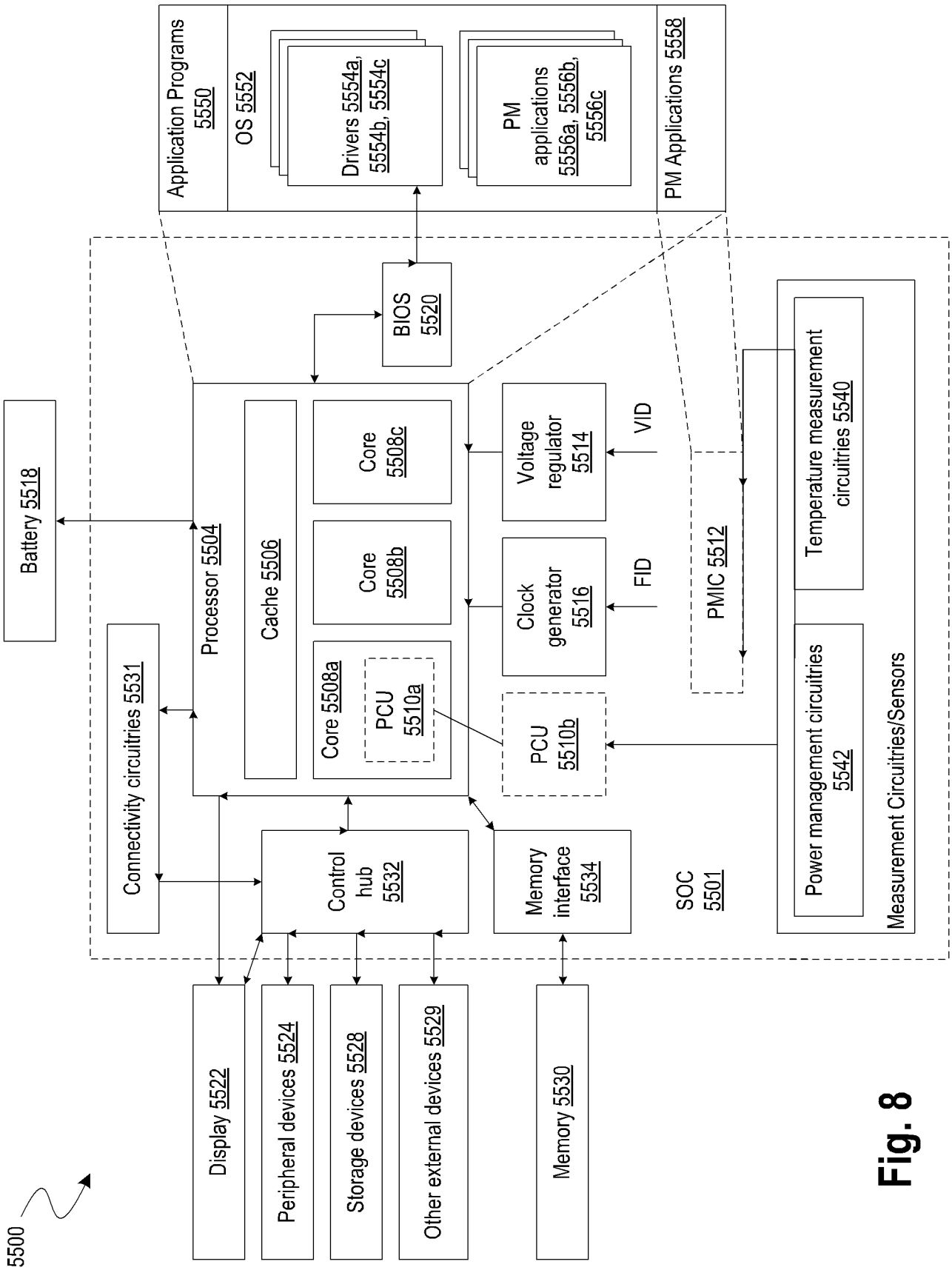

5500
Application Programs 5550
OS 5552
Drivers 5554a, 5554b, 5554c
PM applications 5556a, 5556b, 5556c
PM Applications 5558
BIOS 5520
Battery 5518
Processor 5504
Cache 5506
Core 5508c
Core 5508b
Core 5508a
PCU 5510a
PCU 5510b
Voltage regulator 5514
VID
Clock generator 5516
FID
PMIC 5512
Temperature measurement circuitries 5540
Power management circuitries 5542
Measurement Circuitries/Sensors
Connectivity circuitries 5531
Control hub 5532
Memory interface 5534
SOC 5501
Display 5522
Peripheral devices 5524
Storage devices 5528
Other external devices 5529
Memory 5530

Fig. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2020/137955

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/20(2006.01)i; G06F 1/30(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, WPI, EPODOC: server, rack, load, backup, battery, peak, sudden, power, shaving, bus

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 9336030 B1 (AMAZON TECHNOLOGIES, INC.) 10 May 2016 (2016-05-10) description: column 7 line 64 to column 8 line 57, figure 3B	1-20
Y	US 2018060197 A1 (LITE-ON, INC.) 01 March 2018 (2018-03-01) description: paragraphs [0031]-[0086], figures 1-16	1-20
A	CN 111108460 A (INTEL CORPORATION) 05 May 2020 (2020-05-05) the whole document	1-20
A	US 10725519 B1 (AMAZON TECHNOLOGIES, INC.) 28 July 2020 (2020-07-28) the whole document	1-20
A	US 2018074561 A1 (DELL PRODUCTS, L.P.) 15 March 2018 (2018-03-15) the whole document	1-20

☐ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 September 2021	**24 September 2021**
Name and mailing address of the ISA/CN **National Intellectual Property Administration, PRC** **6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing** **100088** **China** Facsimile No. **(86-10)62019451**	Authorized officer **DING,Xiaoting** Telephone No. **86-(10)-53961261**

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2020/137955

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
US 9336030 B1	10 May 2016	None	
US 2018060197 A1	01 March 2018	US 10565073 B2	18 February 2020
		US 9965365 B2	08 May 2018
		US 2016154716 A1	02 June 2016
CN 111108460 A	05 May 2020	WO 2019019163 A1	31 January 2019
		EP 3659009 A1	03 June 2020
		EP 3659009 A4	03 March 2021
US 10725519 B1	28 July 2020	None	
US 2018074561 A1	15 March 2018	US 10095292 B2	09 October 2018