Chapter 9
Intelligent Power Networks On-Chip

To facilitate the integration of diverse functionality, architectural, an integrated family of circuit, device, and material level power delivery solutions are required. Per core dynamic voltage and frequency scaling is a primary concern for efficiently managing a power budget, and requires the on-chip integration of compact controllers within hundreds of power domains and thousands of cores, further increasing the design complexity of these power delivery systems. While in-package and on-chip power integration has recently became a primary concern [185, 186], focus remains on developing compact and efficient power supplies. A methodology to design and manage in-package and on-chip power has not been a topic of emphasis. Thus, power delivery in modern ICs is currently dominated by ad hoc approaches. With the increasing number of power domains, greater granularity of the on-chip supply voltages, and domain adaptive power requirements, the design of the power delivery process has greatly increased in complexity, and is impractical without a systematic methodology. The primary objective of this chapter is to describe a systematic methodology for distributed on-chip power delivery and management.

A power network on-chip is a vehicle for distributed on-chip power management. The analogy between a network-on-chip (NoC) and PNoC is illustrated in Fig. 9.1 with simplified NoC and PNoC models. Similar to a NoC, a PNoC decreases the design complexity of a power delivery system, while enhancing the control of the quality of power (QoP) and DVS, and providing a scalable platform for efficient power management.

The rest of the chapter is organized as follows. The principles of the PNoC design methodology are described in Sect. 9.1. In Sect. 26.4, the performance of the PNoC architecture is compared with existing approaches based on the evaluation of several test cases. Design and performance issues of the PNoC architecture are also discussed. Some concluding remarks are offered in Sect. 9.3.

I.P. Vaisband et al., *On-Chip Power Delivery and Management*,
DOI 10.1007/978-3-319-29395-0_9

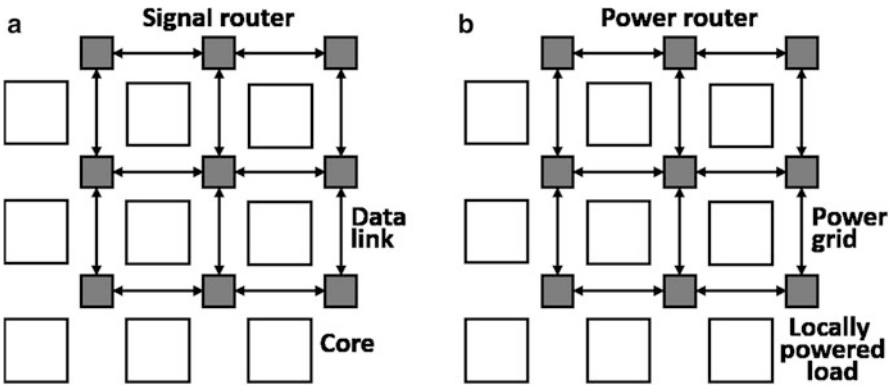

Fig. 9.1 On-chip networks based on the approach of separation of functionality, (**a**) network-on-chip, and (**b**) power network-on-chip

9.1 Power Network-On-Chip Architecture

The key concept in systemizing the design of power delivery is to convert the power off-chip, in-package, and/or on-chip with multiple power efficient but large switching power supplies, deliver the power to on-chip voltage clusters, and regulate the power with hundreds of linear low dropout regulators at the point-of-load [187]. A power network-on-chip is a systematic solution to on-chip power delivery that leverages distributed point-of-load power delivery within a fine grained power management framework. The PNoC architecture is a mesh of power routers and locally powered loads, as depicted in Fig. 9.1. The power routers are connected through power switches, distributing current to those local loads with similar voltage requirements. An example PNoC is illustrated in Fig. 9.2 for a single voltage cluster with nine locally powered loads and three different supply voltages, $V_{DD,1}$, $V_{DD,2}$, and $V_{DD,3}$. The power network configuration is shown at two different times, t_1 and t_2.

A power network-on-chip virtually manages the power in SoCs through specialized power routers, switches, and programmable control logic, while supporting scalable power delivery in heterogeneous ICs. A PNoC is comprised of physical links and routers that provide both virtual and physical power routing. This system senses the voltages and currents throughout the system, and manages the POL regulators through power switches. Based on the sensed voltages and currents, a programmable unit makes real-time decisions to apply a new set of configurations to the routers per time slot, dynamically managing the on-chip power delivery process. Novel algorithms are required to dynamically customize the power delivery policies through a specialized microcontroller that routes the power. These algorithms satisfy real-time power and performance requirements.

A PNoC composed of power routers connected to global power grids and locally powered loads is illustrated in Fig. 9.3. Global power from the converters is managed

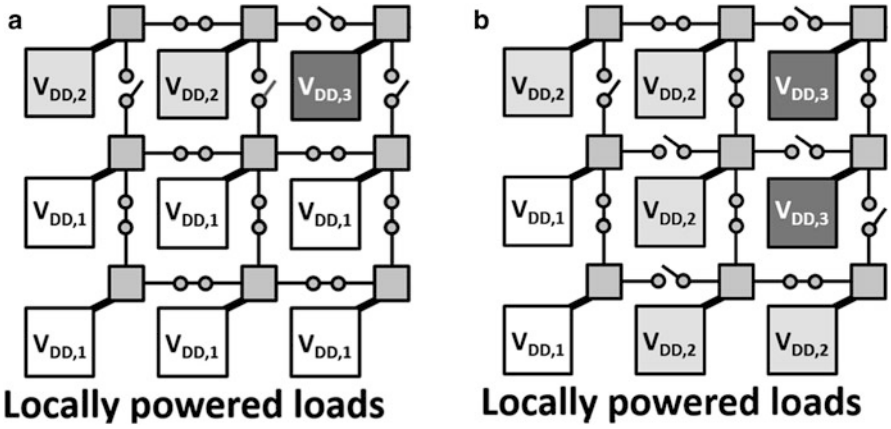

Fig. 9.2 Example of on-chip power network with multiple locally powered loads and three supply voltage levels (**a**) PNoC configuration at time t_1, and (**b**) PNoC configuration at time t_2

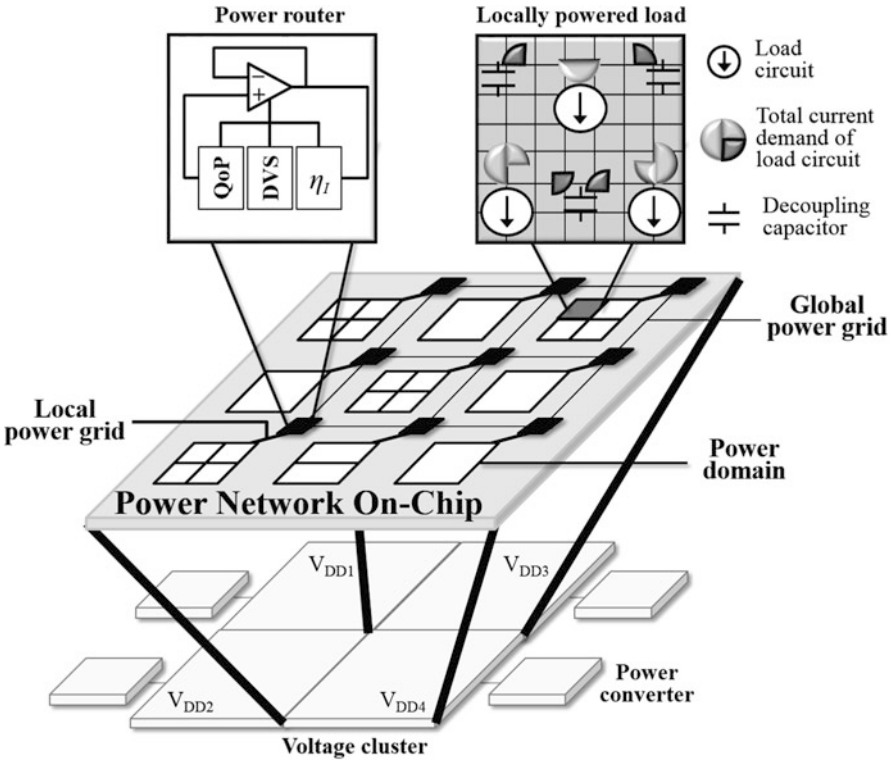

Fig. 9.3 On-chip power network with routers distributing the current over the power grid to the local loads

by the power routers, delivered to individual power domains, and regulated within the locally powered loads. These locally powered loads combine all of the current loads located within a specific on-chip power domain with the decoupling capacitors that supply the local current demand within that region. To support DVS, the power switches within the PNoC are dynamically controlled, (dis)connecting power routers within individual voltage clusters in real time. The current loads are powered at similar voltage levels, and therefore draw current from all of the connected power routers, lessening temporary current variations. Similar to a mesh based clock distribution network [188, 189], the shared power supply lessens the effect of the on-chip parasitic impedances, enhancing voltage regulation and quality of power.

The power routers, local current loads, and power grids are described in the following subsections. Different PNoC topologies and specific design objectives are also considered.

9.1.1 Power Routers

The efficient management of the energy budget is dynamically maintained by the power routers. Each power domain is controlled by a single power router. A router topology ranges from a simple linear voltage regulator, shown in Fig. 9.4a, to a complex power delivery system, as depicted in Fig. 9.4b, with sensors, dynamically adaptable power supplies, switches, and a microcontroller. These structures feature real-time voltage/frequency scaling, adaptable energy allocation, and precise control over the on-chip QoP. With the PNoC routers, the power is managed locally based on specific local current and voltage demands, decreasing the dependence on remotely located loads and power supplies. The scalability of the power delivery process is therefore enhanced with the PNoC approach.

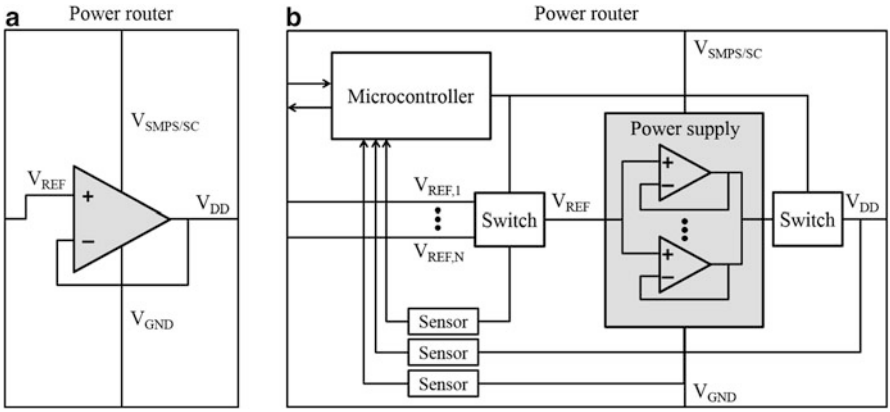

Fig. 9.4 Power routers for PNoC (**a**) Simple topology with linear voltage regulator. (**b**) Advanced topology with dynamically adaptable voltage regulator and microcontroller

9.1.2 Locally Powered Loads

Locally powered loads with different current demands and power budgets can be efficiently managed with a PNoC. The local power grids provide a specific voltage to the nearby load circuits. The highly complex interactions among the multiple power supplies, decoupling capacitors, and load circuits need to be considered, where the interactions among the nearby components are typically more significant. The effective region for a point-of-load power supply is the overlap of the effective regions of the surrounding decoupling capacitors [190, 191]. Loads within the same effective region are combined into a single equivalent locally powered load regulated by a dedicated LDO. All of the LDO regulators within a power domain are controlled by a single power router. A model and closed-form expressions of the interactions among the power supplies, decoupling capacitors, and current loads are required to efficiently partition an IC with billions of loads into power domains and locally powered loads.

9.1.3 Power Grid

Different configurations of local power grids are considered for distributing power from LDO regulators to locally powered loads. A shared power grid with multiple parallel connected LDO sources is illustrated on the left in Fig. 9.5, delivering power to all of the loads within a power domain. A shared power grid with multiple LDO sources is prone to stability issues due to current sharing, and process, voltage, and temperature variations. Specialized adaptive mechanisms are included within the power routers to stabilize a power delivery system that includes a

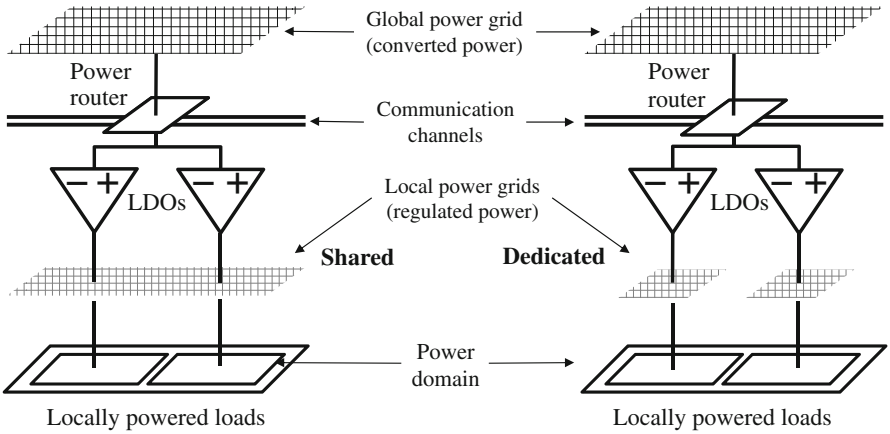

Fig. 9.5 A PNoC power router with two locally powered loads, shared local power grid (*on the left*), and dedicated local power grids (*on the right*)

multi-source shared power grid. Alternatively, to minimize interactions between parallel connected LDO regulators, dedicated power grids each driven by a single LDO should be considered. A topology with dedicated local power grids is illustrated on the right in Fig. 9.5. The dedicated power grids require fine grain distribution of the local power current.

9.2 Case Study

To evaluate the performance of the power router, a PNoC with four power routers is considered, supplying power to four power domains. IBM power grid benchmark circuits [192] model the behavior of the individual power domains. To simulate a dynamic power supply in PNoC, the original IBM voltage profiles are scaled to generate the target power supply voltages between 0.5 and 0.8 V. Target voltage profiles with four voltage levels (0.8, 0.75, 0.7, and 0.65 V) within a PNoC are illustrated in Fig. 9.6. The number of power domains with each of the four supply voltages changes dynamically based on the transient power requirements of the power domains.

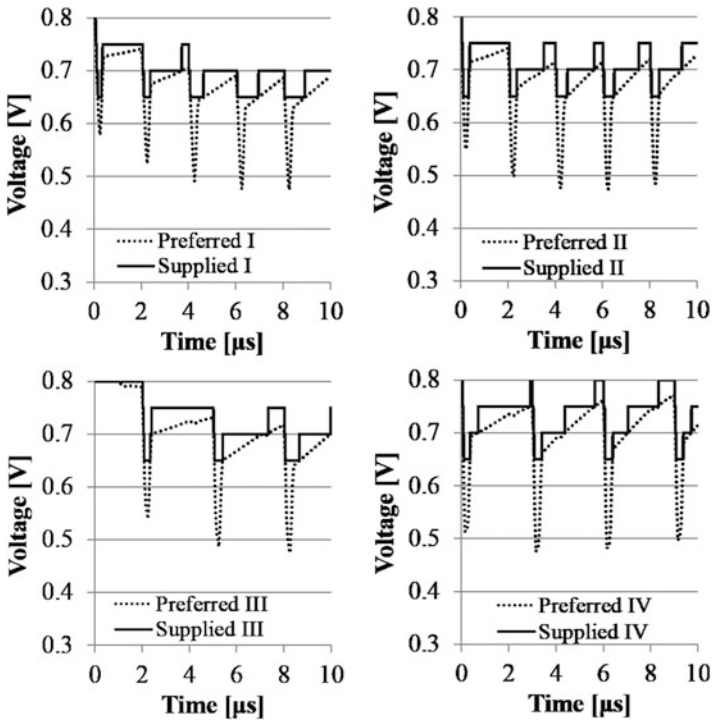

Fig. 9.6 Preferred and supplied voltage levels in PNoC with four power domains

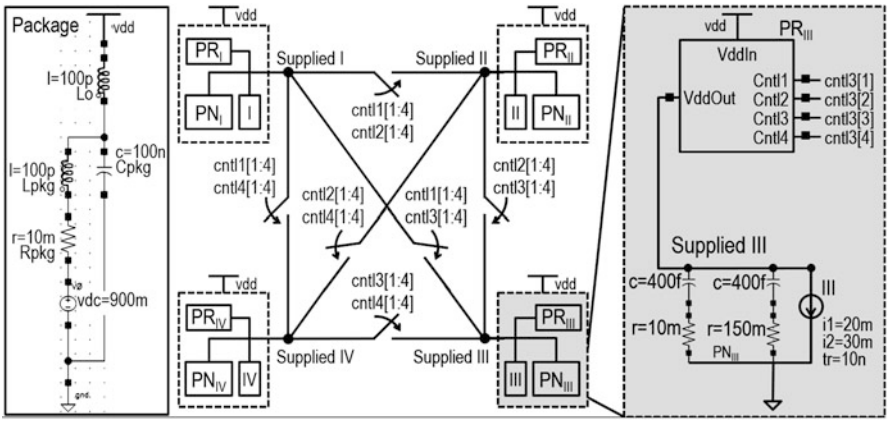

Fig. 9.7 PNoC with four power domains and four power routers connected with control switches

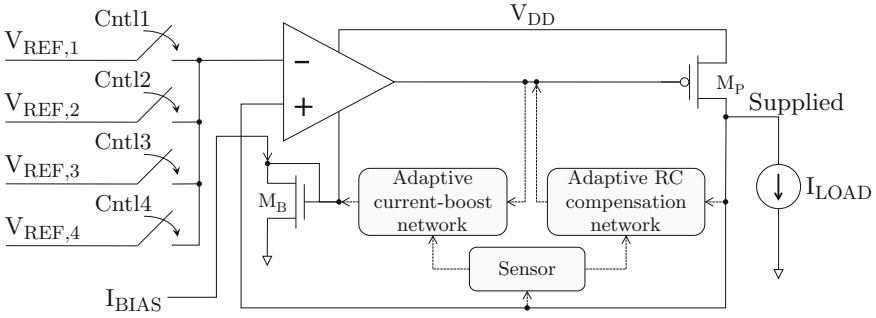

Fig. 9.8 Power router with voltage regulator, load sensor, and adaptive networks

A schematic of a PNoC with four power domains (I, II, III, and IV) and four power routers (PR_I, PR_{II}, PR_{III}, and PR_{IV}) is shown in Fig. 9.7. Each of the power routers is composed of an LDO with four switch controlled reference voltages to support dynamic voltage scaling. In addition, the power routers feature adaptive RC compensation and current boost networks controlled by load sensors to provide quality of power control and optimization, as shown in Fig. 9.8. The adaptive RC compensation network is comprised of a capacitive block connected in series with two resistive blocks, all digitally controlled. These RC impedances are digitally configured to stabilize the LDO within the power routers under a wide range of process variations. The current boost circuit is composed of a sensor block that follows the output voltage at the drain of transistor M_P (see Fig. 9.8), and a current boost block that controls the current through the differential pair within the LDO. When a high slew rate transition at the output of the LDO occurs, the boost mode is activated, raising the tail current of the LDO differential pair. Alternatively, during regular mode, no additional current flows into the differential pair, enhancing the power efficiency of the LDO. A description of the load sensor and adaptive networks are provided in Chap. 18.

Table 9.1 Output load in a PNoC with four power domains

Domain	I	II	III	IV
Minimum output current (mA)	10	75	20	20
Maximum output current (mA)	50	10	30	20
Output transition time (ns)	50	50	10	10

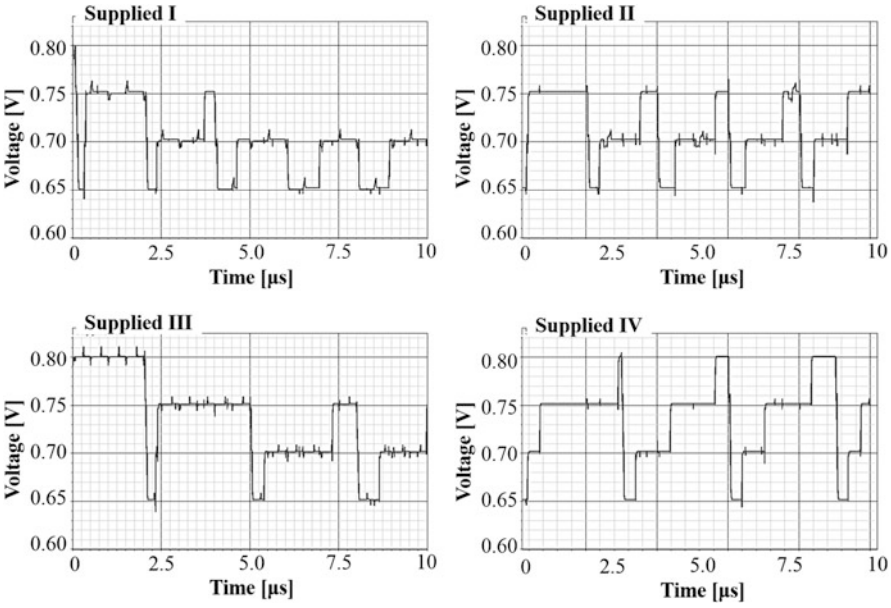

Fig. 9.9 Voltage levels in PNoC with four power domains

The power routers are connected with controlled switches to mitigate load transitions in domains with similar supply voltages. To model the RLC parasitic impedances of the package and power network, non-ideal LDO input and output impedances (PN_I, PN_{II}, PN_{III}, and PN_{IV}) are considered, as shown in Fig. 9.7. The load current characteristics are listed in Table 9.1 for each of the four domains. PNoC SPICE simulation results are shown in Fig. 9.9, exhibiting a maximum error of 0.35 %, 2.0 %, and 2.7 % for, respectively, the steady state dropout voltage, load regulation due to the output current switching, and load regulation due to dynamic PNoC reconfigurations. Good correlation with the required power supply in Fig. 9.6 is demonstrated. The power savings in each of the power domains range between 21.0 % to 31.6 % as compared to a system without dynamic voltage scaling. These average power savings show that the PNoC architecture can control the power supply voltages in real-time, optimizing the power efficiency of the overall power delivery system [193–195].

9.3 Summary

To address the issues of power delivery complexity and quality of power, a power delivery system should provide a scalable modular architecture that supports integration of additional functional blocks and power features (e.g., DVS, adaptive RC compensation, and efficiency optimization with adaptive current boost) without requiring the re-design of the power delivery system. The architecture should also support heterogeneous circuits and technologies.

- The concept and architecture of an on-chip power network PNoC is described in this chapter
- The on-chip network is exploited for systematic power delivery in SoCs to reduce design complexity while increasing scalability
- A methodology that separates power conversion and regulation is provided for efficiently enhancing the quality of power
- The application of local power routing is enabled through a specialized micro-controller for on-chip power management
- Small area power supplies are utilized as point-of-load voltage regulators