1

AN INTRODUCTION TO ROUTING CONGESTION

A traditional standard cell design contains wires that implement the power supply network, clocks, and signal nets. All these wires share the same set of routing resources. With the number of cells in a typical design growing exponentially and the electrical properties of metal wires scaling poorly, the competition for preferred routing resources between the various interconnects that must be routed is becoming more severe. As a consequence, not only is routing congestion increasing, but it is also becoming more damaging to the quality of the designs.

Most conventional design flows synthesize the power supply and clock networks prior to the signal routing stage. The power supply and clock nets do not perform any logical operation, but provide crucial logistical support to the circuits that actually implement the desired logical functionality. The power supply network is designed accounting for several factors such as the current requirements of the design, acceptable bounds on the noise in the supply voltage, and electromigration constraints. This network is designed in the form of a grid which may or may not be regular. Typically, the power supply network is created first and has all the routing resources to choose from. The clock nets are routed next and still have relative freedom, since only the power supply grid has used up some of the routing resources when the clocks are being routed. The clocks, which synchronize the sequential elements in the design, have strict signal integrity and skew requirements. Although they are usually designed as trees in mainstream designs, high-end designs often use more sophisticated clocking schemes such as grids in order to meet their stricter delay and skew requirements (even though such schemes can consume significantly more routing resources). Furthermore, the clock wires are typically shielded or spaced so that the signals on the neighboring wires do not distort the clock waveform; the shielding and spacing also consume some routing resources. The signal nets are routed last and can only use the routing resources that have not been occupied by the power supply and clock wires. Therefore, these are the nets that face the problem of routing congestion most acutely.

In this chapter, we will first introduce the terminology used in the context of routing congestion in Section 1.1, reviewing the basic routing model along the way. Then, we will motivate the need for congestion awareness through a discussion of the harmful effects of congestion in Section 1.2. Next, in Section 1.3, we will try to understand why the problem of routing congestion is getting worse with time. Finally, we will lay out a roadmap for the rest of this book in Sections 1.4 and 1.5 by overviewing the metrics and the optimization schemes, respectively, that are used for congestion.

1.1 The Nature of Congestion

A design is said to exhibit routing congestion when the demand for the routing resources in some region within the design exceeds their supply. However, although this simple intuitive definition suffices to determine whether some design is congested or not, one has to rely on the underlying routing model in order to quantify the congestion and compare its severity in two different implementations of a design.

1.1.1 Basic Routing Model

The routing of standard cell designs follows the placement stage, which fixes the locations of all the cells in rows of uniform height(s) as shown in Fig. 1.1(a). In today's standard cell designs, there is usually no explicit routing space between adjacent rows, since the wires can be routed over the cells because of the availability of multiple metal layers. The entire routing space is tessellated into a grid array as shown in Fig. 1.1(b). The small subregions created by the tessellation of the routing region have variously been referred to as *grid cells*, *global routing cells*, *global routing tiles*, or *bins* in the literature. The dual graph of the tessellation is the *routing graph* $G(V, E)$, an example of which is shown in Fig. 1.2(a). In this graph, each vertex $v \in V$ represents a bin, and the edge $e(u, v) \in E$ represents the boundary between the bins u and v (for $u, v \in V$).

In the routing graph shown in Fig. 1.2(a), the vias and layers are not modeled explicitly. On the other hand, the graph in Fig. 1.2(b) explicitly models bins on two horizontal and two vertical layers, as well as the vias between the different routing layers. The horizontal line segments in this figure represent the boundaries of bins on the same horizontal routing layer, the vertical line segments correspond to vias between adjacent horizontal and vertical routing layers, and the remaining line segments denote the boundaries between the bins on the same vertical routing layer. The process of routing a net on such a graph, therefore, implicitly determines its layer assignment as well.

A routing graph that models each layer explicitly consumes considerably more memory than one that bundles all the layers together. It is possible to