An Area Lower Bound for a Class of Fat-Trees*
(Extended Abstract)

Gianfranco Bilardi¹ and Paul Bay²

¹ Dipartimento di Elettronica e Informatica, Università di Padova, Via Gradenigo 6/A, I35131 Padova, Italy
² Thinking Machines Corporation, 245 First Street, Cambridge, MA 02142, USA

Abstract. A graph-theoretic definition is proposed to make precise the sense in which a "fat-tree" is a tree-like interconnection of subnetworks whose bandwidth is adequately described by the capacity of the channels between subnetworks. The definition is shown to encompass a number of known networks such as the concentrator fat-tree, the pruned butterfly, and the tree-of-meshes. In the established framework, a non-trivial $\Omega(N \log^2 N)$ lower bound is derived for the area of a class of fat-trees, with implications for their area-universality.

1 Introduction

A number of networks have been introduced in the literature and referred to as fat-trees with some further specifier such as concentrator fat-tree, pruned-butterfly fat-tree, and sorting fat-tree. Fat-trees have important universality properties in VLSI and form the basis for number of universal routers [11, 6, 12, 2, 7, 1] and universal circuits [3]. The CM-5 parallel supercomputer uses a fat-tree as its interconnection pattern [9].

Loosely speaking, a fat-tree is a tree whose leaves act as input/output terminals, whose internal nodes are subnetworks with switching capability, and whose edges are channels of appropriate capacity. Proposed fat-trees differ in node structure and channel capacities. In spite of its wide use, the term fat-tree has not been defined precisely. Here, we develop a graph-theoretic definition that captures an important class of networks of the fat-tree type, providing a framework for a general investigation of their properties, such as their layout area.

In Section 2 we introduce the notion of tree-structured network to model tree-like interconnections of subnetworks. For such a network, we call reference tree a weighted tree whose nodes represent the subnetworks and whose edges

* The research of the first author was supported in part by the ESPRIT Basic Research Project 9072 GEPPCOM: Foundations of GEneralPurpose Parallel COMPUTing, by the Italian National Research Council, and by the Italian Ministry of University and Research.
represent the channels between them. Edges are weighted by the capacities of these channels.

We then introduce \(\gamma\)-channel-sufficient (tree-structured) networks, where the number of edge-disjoint paths between two sets of terminals is determined by the maximum flow that can be pushed between those two sets in the reference tree. Using network-flow arguments and Menger's theorem on edge-disjoint paths, we characterize channel-sufficient networks as those for which the load factor of a message set can be estimated (in linear time) to within a multiplicative constant by considering only the capacities of the tree channels. Thus, the reference tree provides an adequate description of network bandwidth.

The pruned butterfly [2], the concentrator fat-tree [11], and the tree-of-meshes [5] are shown to be \(\gamma\)-channel-sufficient in Section 3. The proof is not straightforward, indicating that a non-obvious property is being exposed.

We are interested in the layout area of area-universal fat-trees, where, typically, channel capacity is constant at the \(N\) leaves and doubles every other level to become \(\Theta(\sqrt{N})\) at the root (standard capacities). These fat-trees admit layouts of area \(O(N \log^2 N)\). Since bisection [14] or bifurcator [5] techniques yield only trivial \(O(N)\) lower bounds, it is natural to ask whether there exist smaller layouts.

In Section 4, \(\gamma\)-channel-sufficient fat-trees with standard capacities (satisfying a further technical assumption) are shown to require \(\Omega(N \log^2 N)\) wire area [10]. This result provides us with an entire class of graphs which have the maximum area compatible with their bifurcator. The only previously known graphs exhibiting this behavior were the mesh-of-trees [10] and the expander-connected mesh-of-trees [5] (and graphs that can efficiently embed these ones).

Our lower bound applies, in particular, to the concentrator fat-tree (for which we do not know of previous results) and to the pruned-butterfly and the tree-of-meshes (for which the result was known from mesh-of-trees embeddings). All of these graphs admit optimal layouts of area \(\Theta(N \log^2 N)\).

The fact that the bisection width of fat-trees is a factor \(\Theta(\log N)\) smaller than the square root of the area implies a logarithmic slowdown in the simulation of some networks (e.g., a mesh) of the same area. The existence of area-universal networks capable to simulate any other network of the same area with only constant slowdown remains an open question. Our results indicate that such a network is unlikely to be a fat-tree.

2 A Formal Notion of Fat-tree

We begin by introducing the notion of tree-structured network to make precise the requirement that a fat-tree can be viewed as a tree-like connection of sub-networks.

**Definition 1.** Let \(N\) be a power of two. Let \(R = (V, E)\) be an undirected graph with a distinguished subset of \(N\) vertices referred to as the terminals of \(R\). A tree-representation of \(R\) is a partition of \(V\) into sets \(\{V_{ij} : 0 \leq i \leq \log N, 0 \leq j < 2^i\}\), such that: