Field Programmable Gate Arrays

Field programmable gate arrays (FPGAs) are a type of "field programmable" ASIC. Standard FPGAs can often be reconfigured via hardware description languages such as Verilog to implement new ASIC designs dynamically. While FPGAs are generally less efficient than custom ASICs, they can offer significant speed improvements over CPU implementations. Microsoft in particular has used FPGAs to perform deep learning inference and claims to have achieved significant speedups with their deployment. However, the approach has not yet caught on widely outside Microsoft.

Neuromorphic Chips

The "neurons" in deep networks mathematically model the 1940s understanding of neuronal biology. Needless to say, biological understanding of neuronal behavior has progressed dramatically since then. For one, it's now known that the nonlinear activations used in deep networks aren't accurate models of neuronal nonlinearity. The "spike trains" is a better model (see Figure 9-4), where neurons activate in short-lived bursts (spikes) but fall to background most of the time.
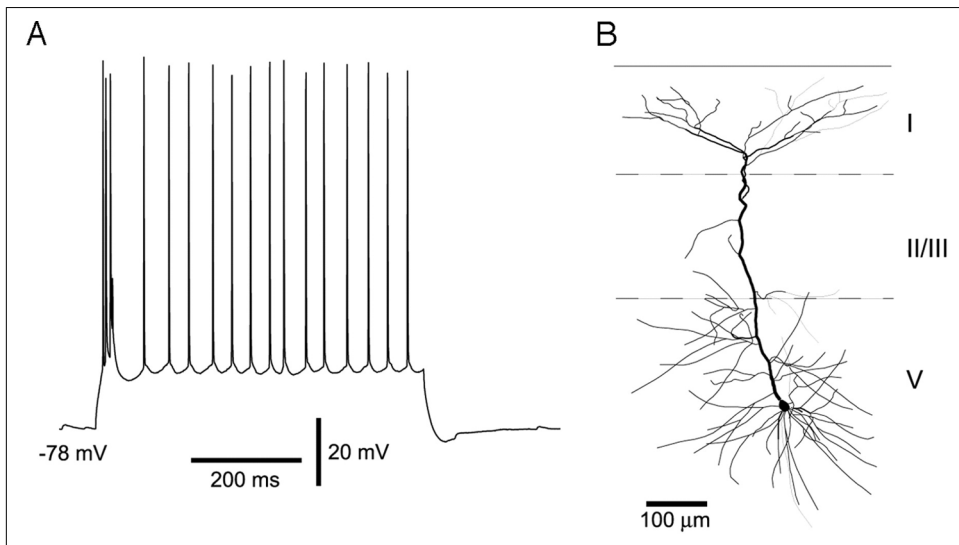

Figure 9-4. Neurons often activate in short-lived bursts called spike trains (A). Neuromorphic chips attempt to model spiking behavior in computing hardware. Biological neurons are complex entities (B), so these models are still only approximate.

Hardware engineers have spent significant effort exploring whether it's possible to create chip designs based on spike trains rather than on existing circuit technologies (CPUs, GPUs, ASICs). These designers argue that today's chip designs suffer from fundamental power limitations; the brain consumes many orders of magnitude less