

Engineering a TinyML system for use off-grid

Simon Dobson

March 9, 2026

This talk contains no new science – but a lot of new engineering.

Migrating machine learning inference onto small devices – "TinyML" – is potentially transformative edge devices and especially for sensor systems. Instead of logging data, possibly returning raw data wirelessly to base, we can perform some or all of the analysis on-device. This has enormous advantages in terms of reducing storage volume and communications bandwidth, and thereby reducing power consumption for systems that need to live off-grid.

TinyML has already been deployed widely: it is the basis of the "wake word" detection used on Alexa and similar voice assistants. Our application is different: we want to use a machine learning model to identify bird songs to determine species abundance from sensors placed in the field. The standard model for this is considerably larger than those for wake words (4M parameters rather than around 100K), and so requires both more working memory and significantly more local computation. The devices must also work without established network infrastructure.

Can we run these models on sensor-class devices, and report results back wirelessly? Answering this question has involved exploring the engineering design space for edge devices to trade-off power against available memory, compute, and communications. This turns out to be more complicated than it first appeared. Modern devices are highly optimised towards specific use cases, and there is often very little slack to fitting an application to a device. Specifically, common sensors do not have either the memory or computational power to run a sizeable machine learning model, whereas common single-board computers draw too much power.

In this talk we walk through the design process and the experiments we conducted in building prototype sensor nodes ready for deployment. We also discuss how the process illustrates gaps in the research landscape, both in terms of the structure of models intended for TinyML and in terms of devices for running those models.