Taste of Research
Gough Yumu LUI
Engineer's Log Book

Week 10

- Monday 31st January 2011

So, I've now got a pile of data. What to do with it? Already I've found the mean, standard deviation and mode – but there's not really much I can say about it yet. I do notice that the mode seems to be a bit "flat" – several consecutive equal values, and sometimes it differs from the mean quite significantly – but sometimes it seems to be better than the mean because it's less jumpy. Given the large range of standard deviation values – I don't even know what's the best way to filter it – let alone compare it.

What I do think is that we can start with the outdoor values and do linear fits to the data – and derive equations for the gradient of the line. This way, we can look at the gradients to decide which cards are more "similar" in terms of offset and slope, and we can probably do it for the mean and the mode sets of data. Unfortunately, this data isn't filtered yet, which means I might have to do this again after we've decided on a filtering algorithm – or maybe we won't filter it at all.

One thing I forgot to mention is that for some devices, there is an unusual temporal pattern in the data sometimes. This might be something to investigate in the future.

Device	Mean		Mode	
	Gradient	Intercept	Gradient	Intercept
Netgear WG-111U (2.4Ghz)	-20.741	-34.938	-20.919	-33.723
Netgear WG-111U (5Ghz)	-20.203	-37.895	-20.451	-37.484
Netgear WPN111	-2.594	-74.385	-0.847	-80.668
Atheros 5006UG	-17.209	-36.861	-16.944	-36.554
Billion BiPAC3011G Card A	-17.878	-42.544	-19.091	-40.339
Billion BiPAC3011G Card B	-14.067	-49.726	-17.937	-41.797
Billion BiPAC3011G Card C	-18.612	-44.625	-22.791	-36.947
Netgear WG-111v2 Card A	-5.087	-48.510	-5.365	-47.029
Netgear WG-111v2 Card B	-5.283	-49.900	-5.440	-48.915
D-Link DWA-140	-15.930	-38.048	-16.886	-35.124
D-Link DWL-122G	-17.143	-37.556	-17.242	-37.300
Diamond Digital A101 Card A	-14.964	-37.749	-16.599	-34.904
Diamond Digital A101 Card B	-18.066	-34.991	-19.413	-32.469
Netgear MA101	-3.815	-24.912	-0.000	-24.000
Broadcom BCM4312 (2.4Ghz)	-15.414	-37.739	-16.004	-36.811
Broadcom BCM4312 (5Ghz)	-20.647	-43.074	-20.755	-42.960
Belkin Play (2.4Ghz)	-16.638	-37.176	-16.790	-34.451
Belkin Play (5Ghz)	-20.091	-29.180	-20.061	-29.196
Intel Centrino 2200BG	-9.340	-41.138	-6.224	-41.023
Intel Centrino 3945ABG (2.4Ghz)	-17.600	-42.665	-19.564	-38.108
Intel Centrino 3945ABG (5Ghz)	-18.319	-48.370	-16.643	-50.007
Intel Wi-Fi Link 5300N (2.4Ghz)	-18.462	-32.830	-22.142	-22.755
Intel Wi-Fi Link 5300N (5Ghz)	-19.089	-39.236	-17.708	-40.341
HTC Android Phone	-17.461	-40.415	-17.290	-40.404
Nokia N95	-13.338	-42.168	-14.407	-40.615

A quick scan of this table (re-ordered by chipset) shows that it is difficult to prove ANYTHING. Cards of the same make show *significant* differences in their intercept and slope (possibly an artifact of testing). Cards of similar chipsets both show similarities and marked differences depending on the maker. So far it seems that the original hypotheses that I had proposed are difficult to prove altogether, and are in fact partially disproved by the data altogether.

Maybe we should look at the average standard deviation on a per-card basis, although I suspect that human influences would cause more impact than anything on these results:

Device	Average SD Indoors	Average SD Outdoors
Netgear WG-111U (2.4Ghz)	3.067	3.480
Netgear WG-111U (5Ghz)	0.861	0.655
Netgear WPN111	12.339	13.460
Atheros 5006UG	4.586	3.500
Billion BiPAC3011G Card A	9.089	5.501
Billion BiPAC3011G Card B	7.770	9.658
Billion BiPAC3011G Card C	9.343	10.013
Netgear WG-111v2 Card A	1.175	1.751
Netgear WG-111v2 Card B	1.019	1.651
D-Link DWA-140	7.124	8.224
D-Link DWL-122G	1.621	1.321
Diamond Digital A101 Card A	5.239	5.650
Diamond Digital A101 Card B	3.905	4.578
Netgear MA101	5.660	6.285
Broadcom BCM4312 (2.4Ghz)	3.271	2.231
Broadcom BCM4312 (5Ghz)	0.533	0.605
Belkin Play (2.4Ghz)	5.438	8.706
Belkin Play (5Ghz)	0.531	0.143
Intel Centrino 2200BG	6.238	8.105
Intel Centrino 3945ABG (2.4Ghz)	8.256	8.091
Intel Centrino 3945ABG (5Ghz)	2.570	1.892
Intel Wi-Fi Link 5300N (2.4Ghz)	12.978	12.961
Intel Wi-Fi Link 5300N (5Ghz)	2.593	2.417
HTC Android Phone	1.858	1.261
Nokia N95	5.867	3.553
Roving Networks Wi-Fi Tag	4.900	N/A

Unfortunately, again, it seems it's a mixed bag. Cards of the same make but different manufacture displayed differences again – sometimes by a factor of two. Variances for outdoor were, at times, larger than the variance for indoors which seems a little unusual. It can be seen that the variance on the 5Ghz band is consistently better than that of the 2.4Ghz band and that certain cards are better and produce less variance than others (however, it must be kept in mind that different cards have different sensitivity "slopes" and that would impact on the average SD seen for a given amount of signal variation).

I'm not too sure where to go from here in terms of data analysis. If there were any true similarities, I would expect them to show themselves in the above two "tests". I think this will have to wait for the meeting to see how we can proceed from this.

- Tuesday 1st February 2011

Time to look towards doing something for the poster given the workshop is coming tomorrow. I have to prepare some visuals – I thought it would be rather instructive if we had some photos of the devices that were tested – this would be great for the report and for the poster. We will also have some pictures of the indoor test and outdoor test environment and setups. I'll draw a diagram indicating test distances (maybe not), and I'll produce some graphs of all the card signal trends (plot-all graphs would be nice and simple) and maybe one or two graphs of a few cards. Unfortunately, I'm not sure how far I've gone to proving certain things – this is going to take more time than I envisaged it would.

I've managed to take photos of all the devices which I have in my possession. Unfortunately, there are devices which are at the school which I might need to take photos of just to make sure the report can be completed. I have used Photoshop to erase the background so that it is more professional.

There is a draft copy of the poster as a separate PDF available from my homepage. You can take a look at it – I think it's a little messy at the moment but it's alright. Just printed one out to bring to the meeting tomorrow.

I also have to prepare a two minute speech (of about 250 words) – I'll worry about that early tomorrow. I would think that it's predominantly the same as what I've covered in the poster.

> Let's say you wanted to find a certain room at this uni. It would be much easier if you knew where you were now. You might pull out your GPS only to find that it doesn't work well indoors. That's why there has been interest in other methods – one of which uses Wi-Fi access points with a technique called Fingerprinting.
>
> Fingerprinting involves taking measurements at reference points and recording the signal strengths of visible access points into a database. When a user wants a position, they scan for the access points and relay the data to a server which finds the closest match in order to return their position. This, of course, relies on the fact that the user's device receives the signals in the same way as the device which was used to build the database.
>
> The aim of this experiment was to look at how different devices react to the same signal environment. So, we got a bunch of different Wi-Fi devices including laptops, dongles and phones and put them onto a trolley and asked them to measure the signal strength at varying distances from a fixed access point.
>
> We found that different Wi-Fi devices reacted differently to the signal – from the graphs we can see that there are significant differences in the spread of signal levels reported for the same distances. Some devices don't react to the signal in an expected way at all. All of this confirms the need for device calibration to make the fingerprinting system a viable system. The data we've collected could go towards some sort of calibration which could compensate for differences between cards and improve the accuracy.
>
> Another conclusion we've come to is that the 5Ghz band is better than the 2.4Ghz band at producing stable signal readings. Using signals from the 5Ghz band has the potential to improve the fingerprinting position accuracy as well.

Unfortunately, this is still 50+ words over. I need to find a way to trim it. As for the questions posed at least workshop – here are the following answers:

To whom do you need to Acknowledge/thank for support, TOR opportunity?
University of New South Wales for sponsoring and providing the ToR.
Binghao Li for being my supervisor, Thomas Gallagher for assisting me with my experiments.

What have you gained personally from the TOR experience?
An experience in conducting research semi-independently, solving numerous problems as they arise, critical thought in examining and justifying results, ability to apply my skills toward a goal which I feel is worthy.

What are the rationale, aim/s and key outcome of your project?
My research is supporting the area of Wi-Fi Fingerprinting based location systems. It aims to observe the differences between 2.4Ghz and 5Ghz signal propagation and differences in RSSI between different Wi-Fi devices so that we can make some conclusions on how they affect Wi-Fi Fingerprinting systems. The key outcome is that we observed significant differences in the way different Wi-Fi cards report their signal strengths, and that 5Ghz band signals were more stable than 2.4Ghz signals in all our tests.

What are you most proud of in your project? E.g. interesting results, solved a problem during the project?
That testing of this scale has not been done in the past.

I was able to contribute my skills in programming to solve problems in parsing the data, and also contribute my own devices to perform the testing.
Was able to retain my sanity throughout despite the long hours of testing doing repetitive work.

- Wednesday 2nd February 2011

Today was the day for Poster Workshop 2. In this workshop, we were taught how to prepare our speech and practice it to a green piece of paper. An interesting psychological idea is that it is bad to practice a speech sitting down as it is associated by the body to the action of sitting down and you'll essentially forget it if you stand up. We were told to distill our purposes and ideas and write down a draft plan for a speech. I realized I had neglected to thank UNSW at the closing of the speech and reflect upon what I have learnt – that being said, I would like to shorten the speech slightly and include it just to be safe, but it is probably of the least importance. It would be great if I could confidently and consistently deliver this speech as it is and remain within the 2 minute mark. That's why it is important to rehearse.

I was also supposed to meet with Binghao and Thomas at midday, right after the workshop. Unfortunately, I was told that Binghao wasn't in and Thomas was on lunch. I checked my e-mail and Thomas sent me an e-mail asking what time were we meeting tomorrow – unfortunately, it was supposed to be today. I resolved to send them all an e-mail in return to let them know we would meet on Monday instead, and in the interim, to take a look through my draft poster and draft speech and give me some feedback.

Which also reminds me, none of the manufacturers have yet replied to any of my e-mails. I think it is a loss altogether. Unfortunately, it gives us no solid evidence on which to back up our hypothesis about diversity reception.

- Thursday 3rd February 2011

I first received this e-mail from Thomas:

> Hi Gough,
>
> very good work you did there, very detailed and easy to understand. We can talk more about data analysis on Monday.
>
> As for poster and speech feedback, a few points:
>
> 1) I've actually never done a poster before, but of what I see, there is a bit too much text on it, which makes it kinda hard to read. You can delete a few parts, and make the font bigger. For instance about the the testing methodology, no needs to get too much into details (eg. "The data was downloaded and processed using a C program and Matlab to produce the results." not very useful for instance).
> And no need to make big sentences (eg The device was set to log the RSSI of the AP on all available bands simultaneously. would become Log all Wi-Fi channels). You have to make it shorter and clearer, othewise people wont read it.
>
> 2) About the speech, I think it's very good like that, don't change anything
>
> See you on Monday
>
> Thomas

Furthermore, I then received an e-mail from Binghao:

> Well done Gough.
>
> I agreed with Thomas' comments. Further comments:
> 1. In many indoor environments, it is not possible to receive accurate positioning information via satellite based systems such as GPS. One alternative ...

I would simply say GPS has difficulties to work indoors or in urban canyons, WiFi has been widely accept as ...

2. Unfortunately, with the wide
variety of devices available, we cannot guarantee that
all devices will perform equally under the same
environment.

I would emphasize researchers put a lot of effort on algorithms to improve the accuracy, they assume the device perform similarly or only have a offset. You are trying to find the real performance of different cards. The finding will give us a clear understanding of the possible accuracy of WiFi positioning system (limitation).

3. You may want to summarize the factors which makes the reported RSSI different - this is a question has not been well answered by researchers (at least in WiFi positioning area) so far.

4. You may want to explain what are the difference of indoor and outdoor test. If no significant difference, then remove one of the figures. Figure4 and 5 is hard to read, too many polylines and the legend is not very helpful. But I do not know how to make it clearer.

5. Conclusion, you may want to point out several things: limit the user chipsets; calibration if it is possible or decrease the accuracy - but the lower bound is the cell ID like result - or a potential simple algorithm to provide a reasonable accuracy (even the poor performed WiFi receiver).

Cheers

Binghao

I sent the following reply to Thomas and Binghao:

Dear Thomas,

Thanks very much for the feedback. I will work on cutting down the text on the poster, but unfortunately the font size is fixed by the faculty template - I'm told that I'm not allowed to change that. Maybe instead we can make some more room so that we can space it out better or maybe enlarge some of the figures slightly.

- Gough

I agree, there are changes which need to be made to the poster and so I will endeavour to improve it as soon as I can. The poster is due on the 9th Feb 2011 by 4pm to Chloe Fong.

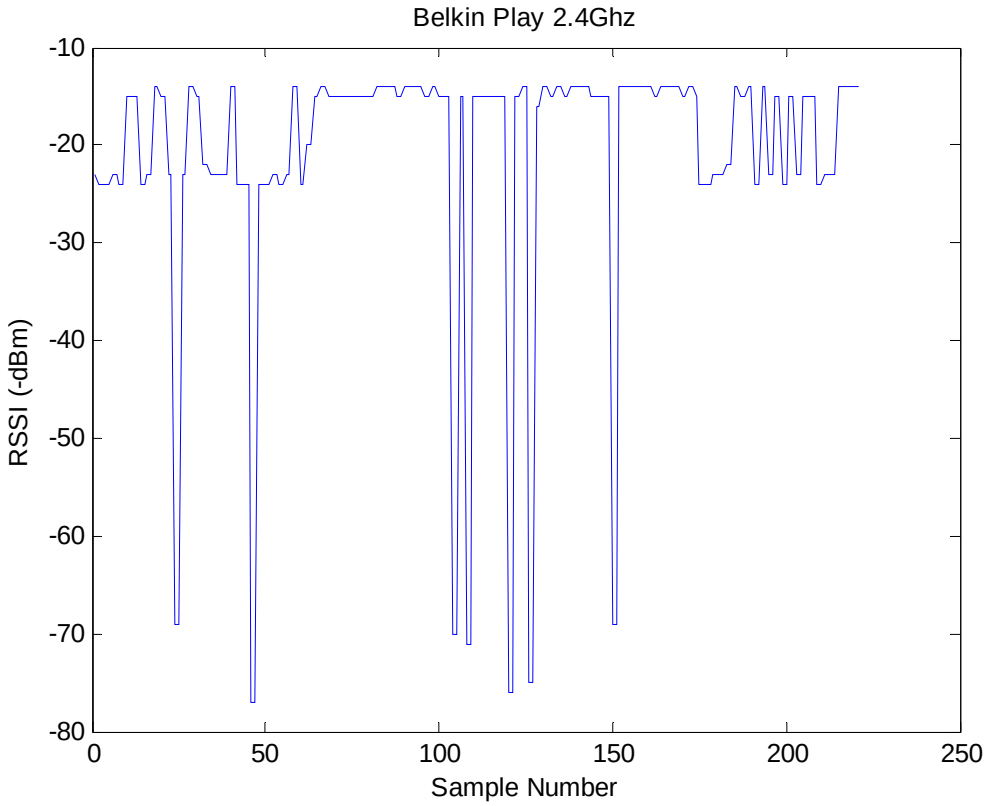
I think the factors that make RSSI different is hard to easily put – as it very much depends on how the RF Frontend of the card is designed, and how the chipset aligns the AGC amp values to an arbritary signal level value which is free for the chipset manufacturer or driver to determine. Not to mention the diversity antenna theory which I had but am still unable to easily prove. Aside from that, you have a lot of environmental influences. In the meantime, I think I should focus more on the data.

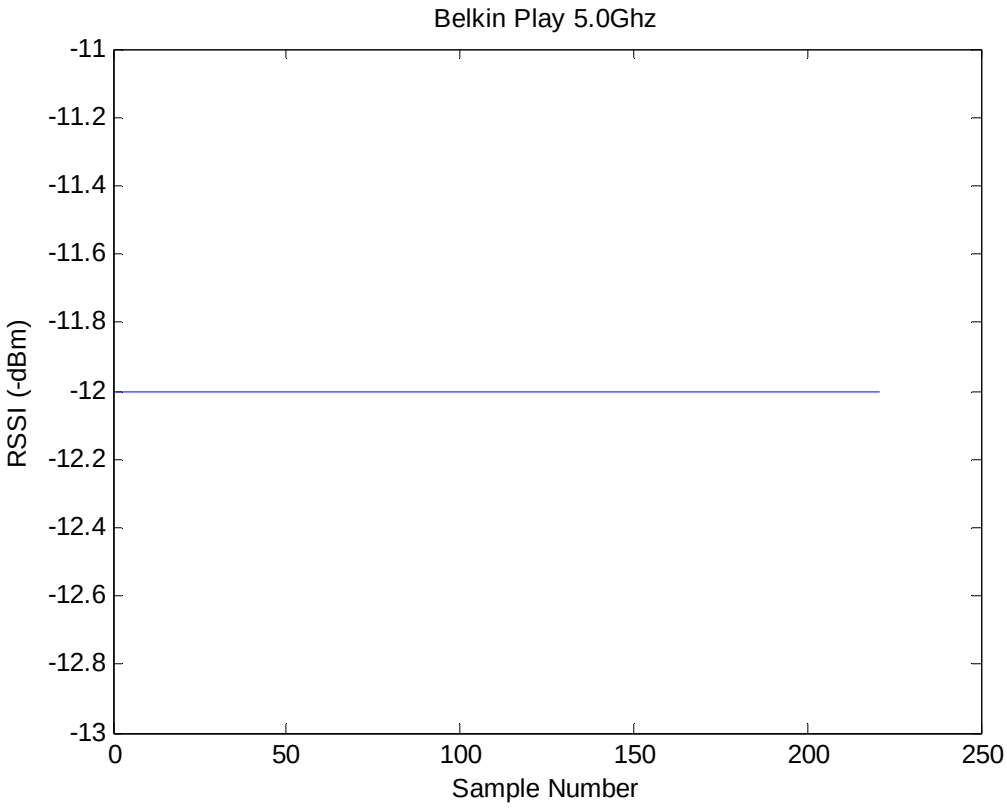
I just realized that the previous linear fits for the outdoor lines did not include the R^2 value which was a bit silly. I should probably modify it to produce the R^2 value – however – I would expect for pretty poor values as the outdoor testing has one particular point which deviates significantly from the expected linear value. Furthermore, I'm not exactly sure how to deal with the indoor values, as I wouldn't expect them to correlate with a line as they're rather affected by multipath. Also, it can be seen that the values really close up and far away have more variance – and worse fits – and there is no consistent set of data points with a nice trend either. I'm not sure how to do it either, I can't seem to make sense of the matlab help for the function regress, and polyfit results in either normr or r value which I don't really know what to do with.

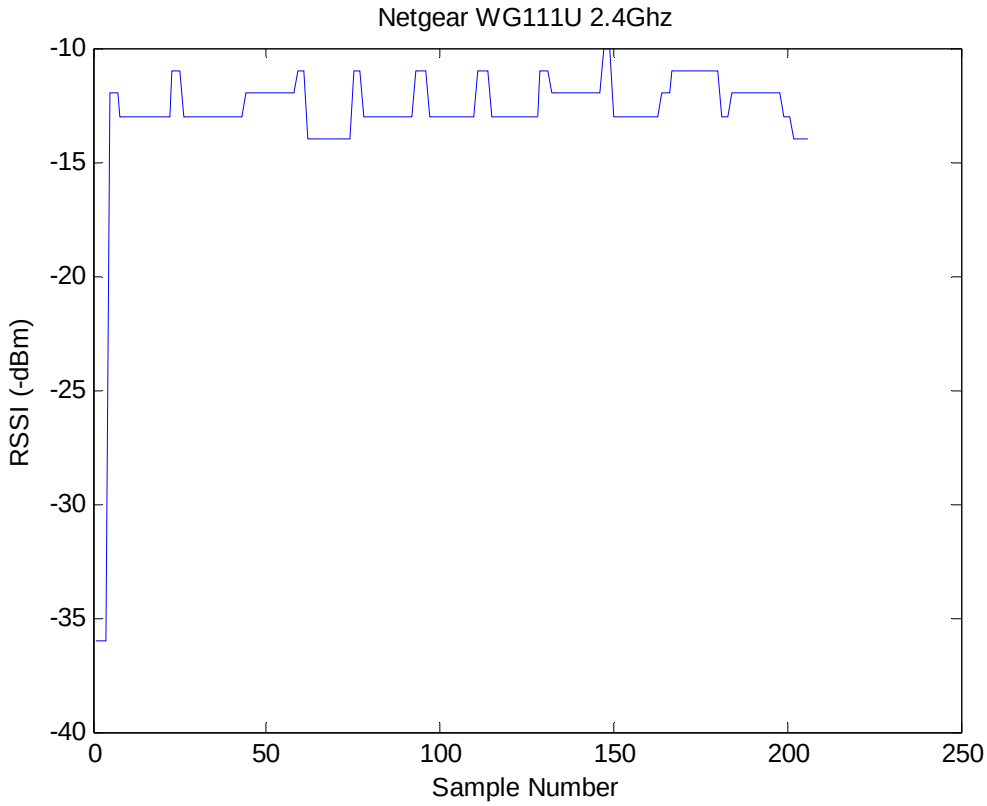
- Friday 4th February 2011

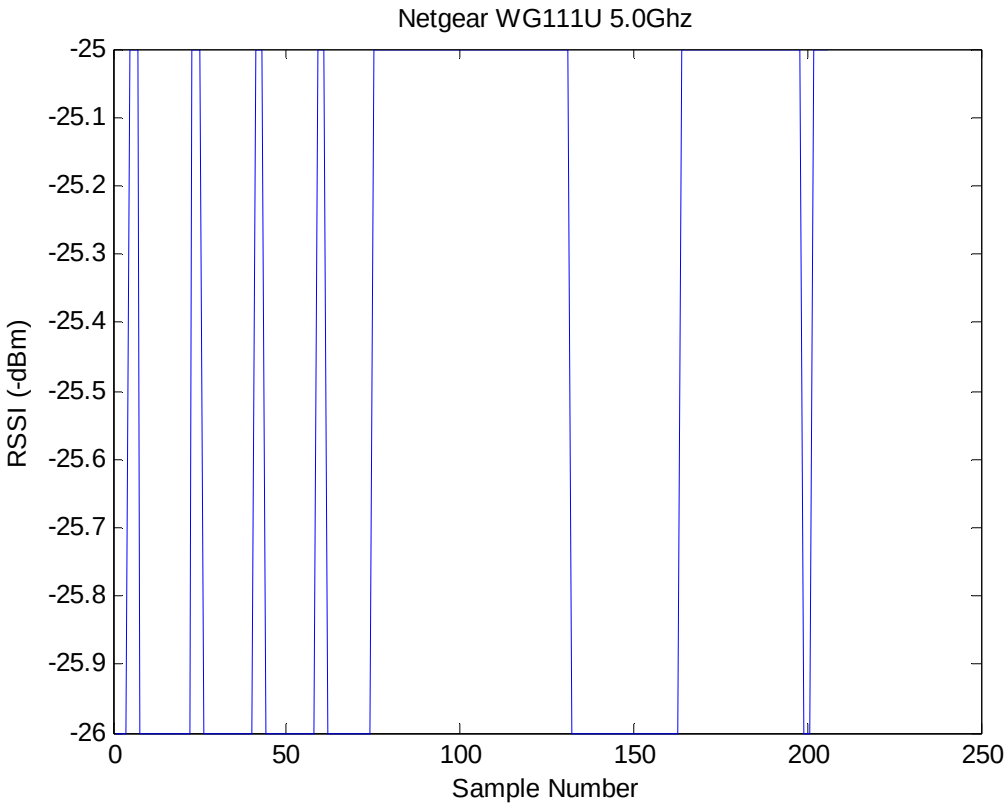
Today, I spent some time altering and cleaning up the poster and reducing the number of words on it. A new draft is available online now which is much more refined.

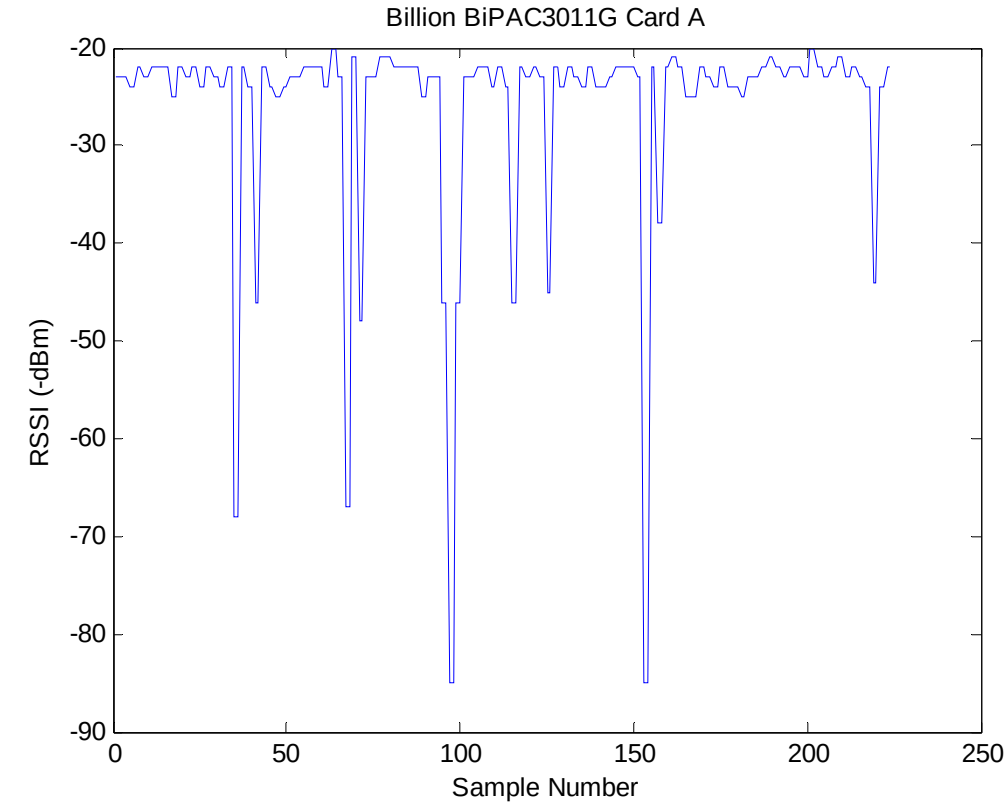
I've also taken some interest in the temporal variations in the signals – here are all the outdoor 30cm results – look at how different cards react to the same signals (note the subtle failure modes that some cards have):

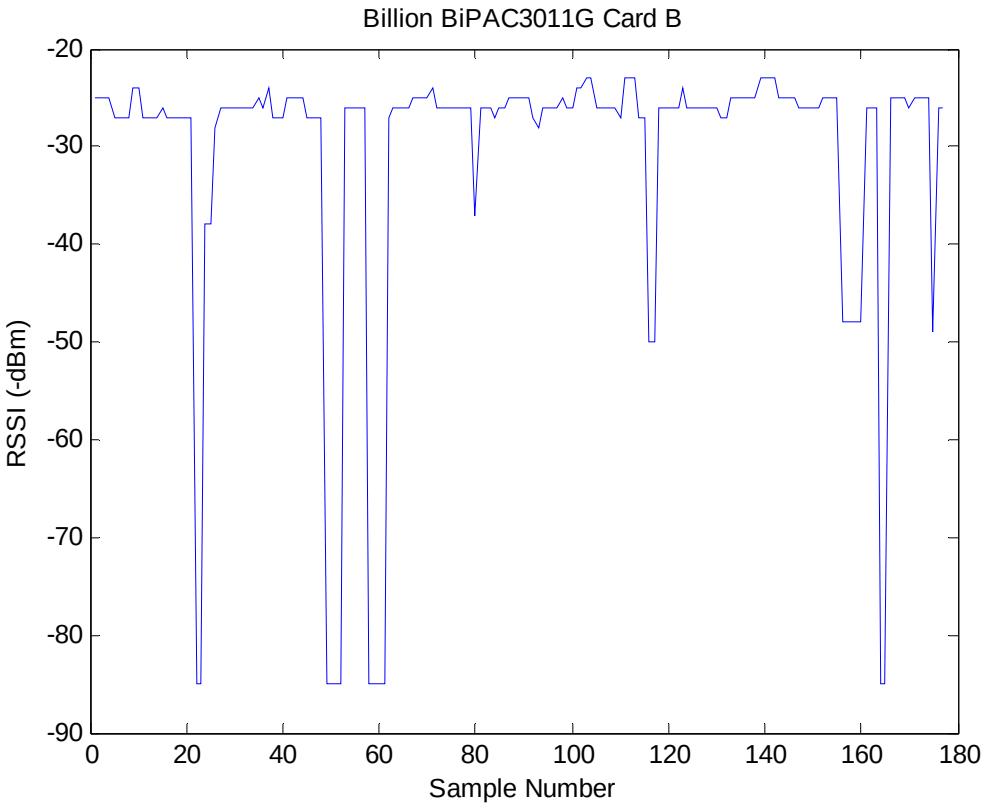


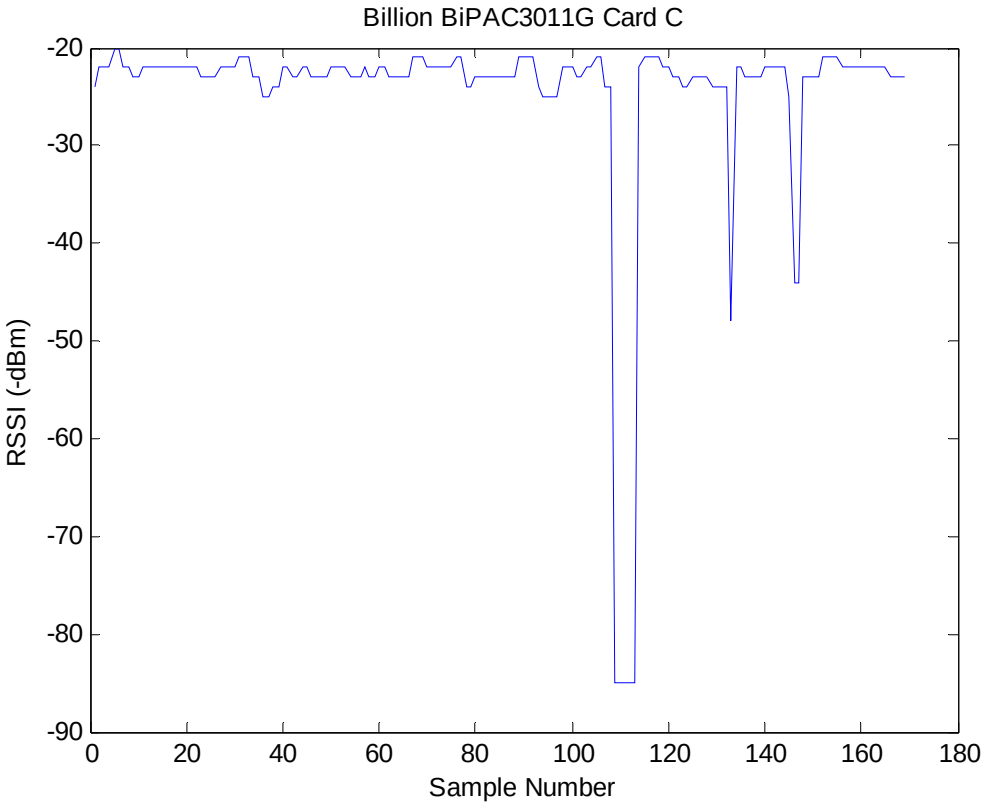

Belkin Play 5.0Ghz
RSSI (-dBm)
-11
-11.2
-11.4
-11.6
-11.8
-12
-12.2
-12.4
-12.6
-12.8
-13
0
50
100
150
200
250
Sample Number

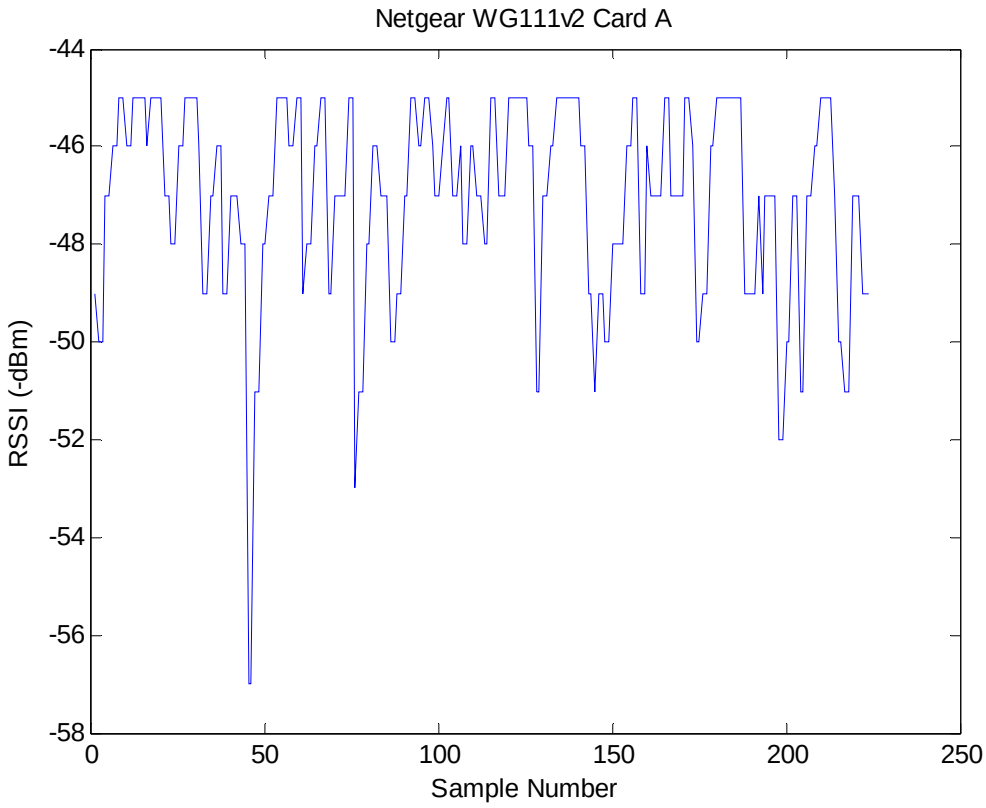

Netgear WG111U 2.4Ghz
RSSI (-dBm)
-10
-15
-20
-25
-30
-35
-40
0
50
100
150
200
250
Sample Number

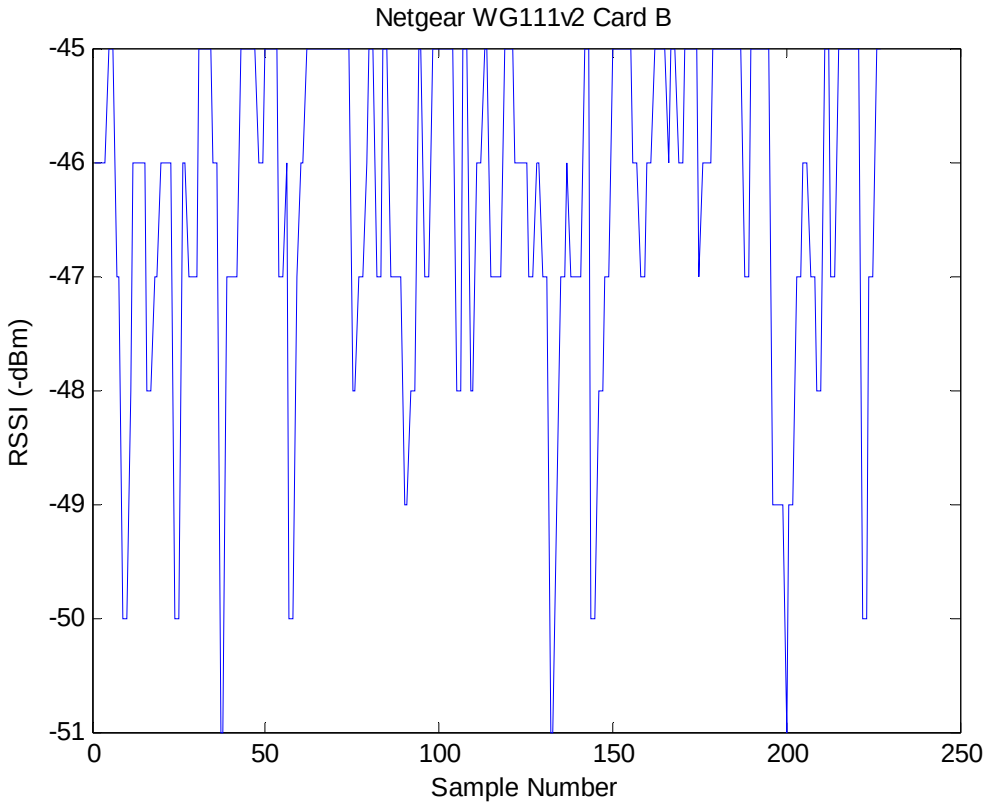

Netgear WG111U 5.0Ghz
RSSI (-dBm)
-25
-25.1
-25.2
-25.3
-25.4
-25.5
-25.6
-25.7
-25.8
-25.9
-26
0
50
100
150
200
250
Sample Number

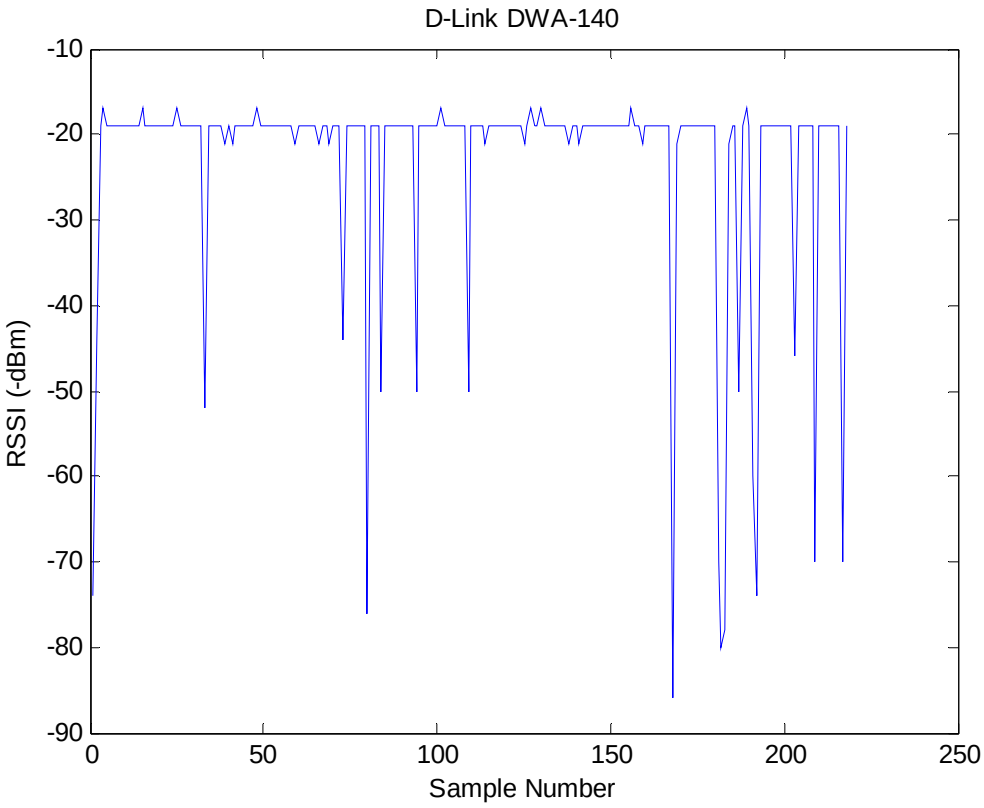

Billion BiPAC3011G Card A
RSSI (-dBm)
-20
-30
-40
-50
-60
-70
-80
-90
0
50
100
150
200
250
Sample Number

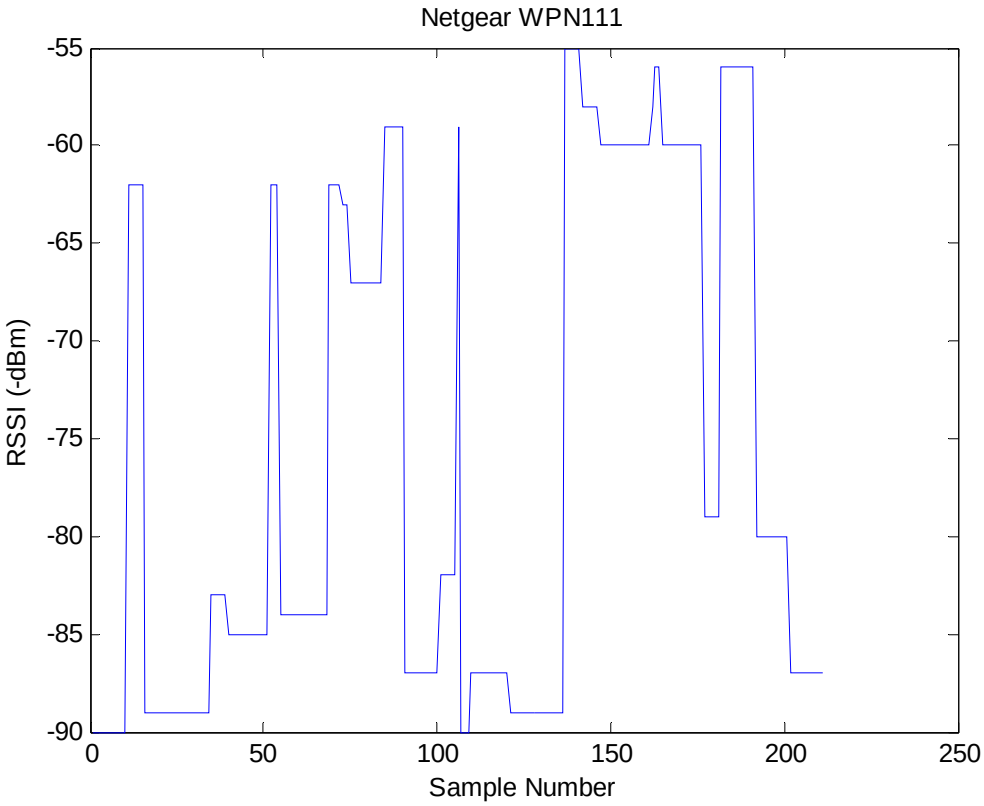

Billion BiPAC3011G Card B
RSSI (-dBm)
-20
-30
-40
-50
-60
-70
-80
-90
0
20
40
60
80
100
120
140
160
180
Sample Number

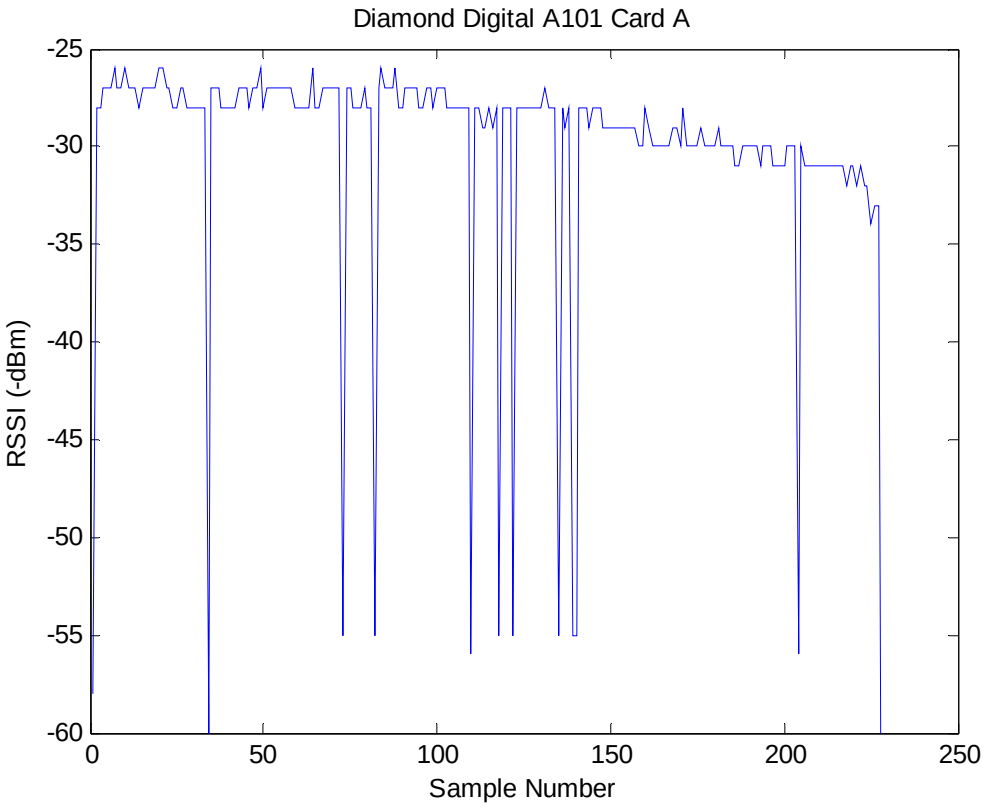

Billion BiPAC3011G Card C
RSSI (-dBm)
-20
-30
-40
-50
-60
-70
-80
-90
0
20
40
60
80
100
120
140
160
180
Sample Number

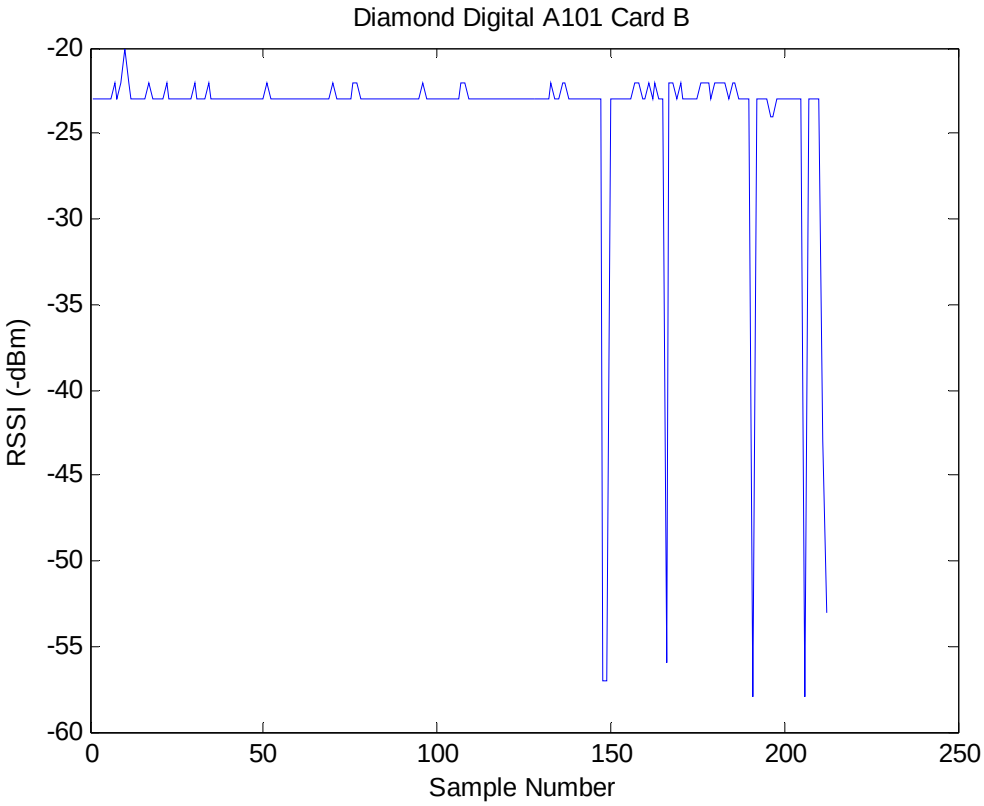

Netgear WG111v2 Card A
-44
-46
-48
-50
-52
-54
-56
-58
RSSI (-dBm)
0
50
100
150
200
250
Sample Number

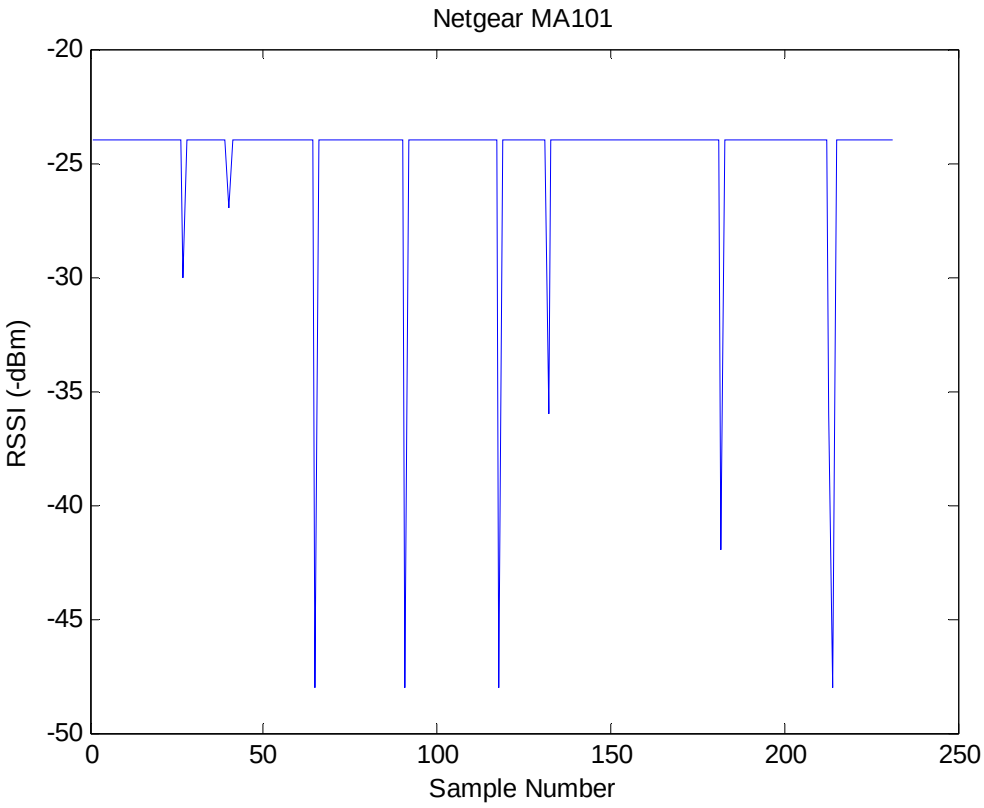

Netgear WG111v2 Card B
-45
-46
-47
-48
-49
-50
-51
RSSI (-dBm)
0
50
100
150
200
250
Sample Number

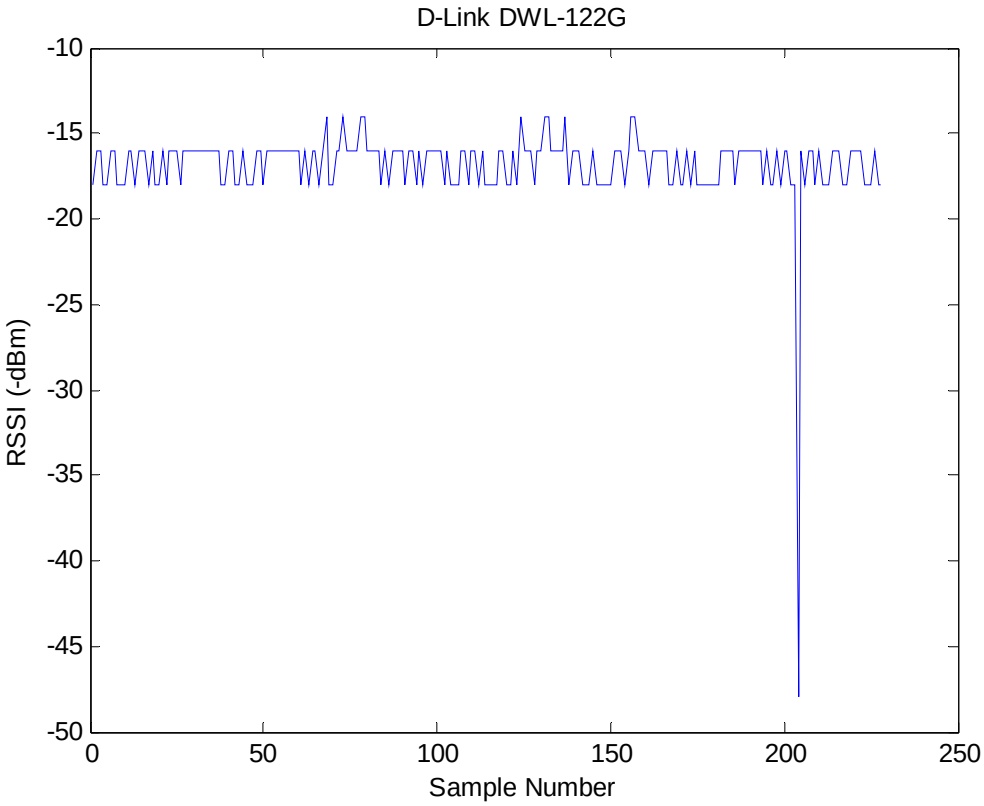

D-Link DWA-140
-10
-20
-30
-40
-50
-60
-70
-80
-90
RSSI (-dBm)
0
50
100
150
200
250
Sample Number

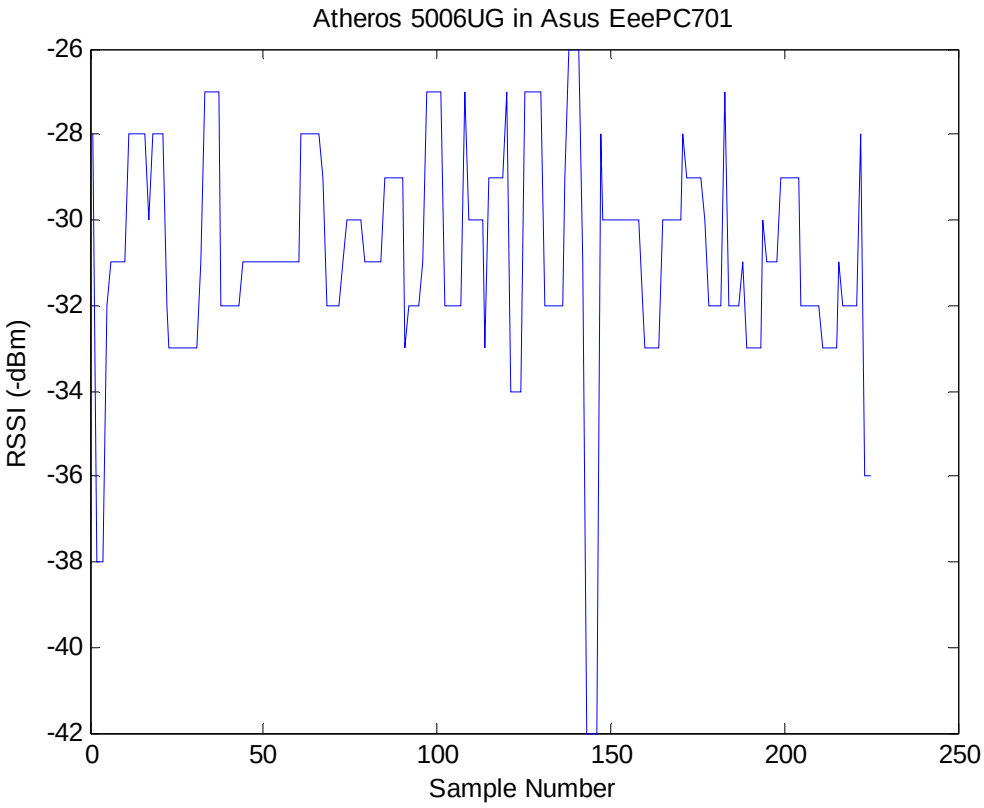

Netgear WPN111
-55
-60
-65
-70
-75
-80
-85
-90
RSSI (-dBm)
0
50
100
150
200
250
Sample Number

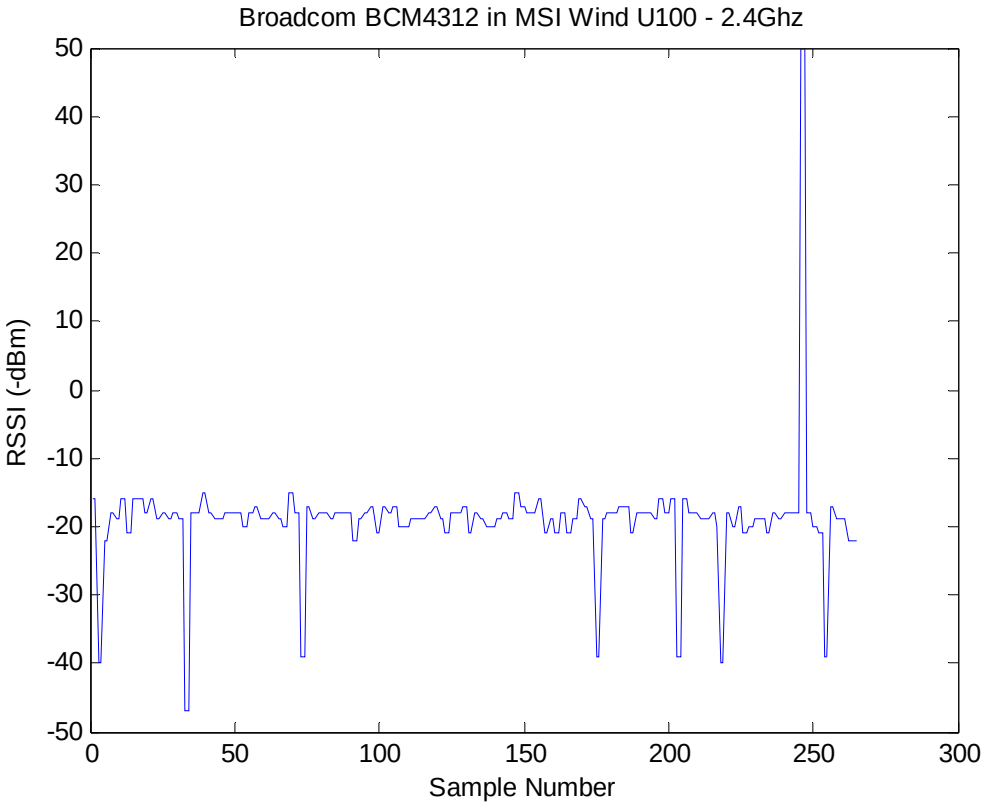

Diamond Digital A101 Card A
RSSI (-dBm)
-25
-30
-35
-40
-45
-50
-55
-60
0
50
100
150
200
250
Sample Number

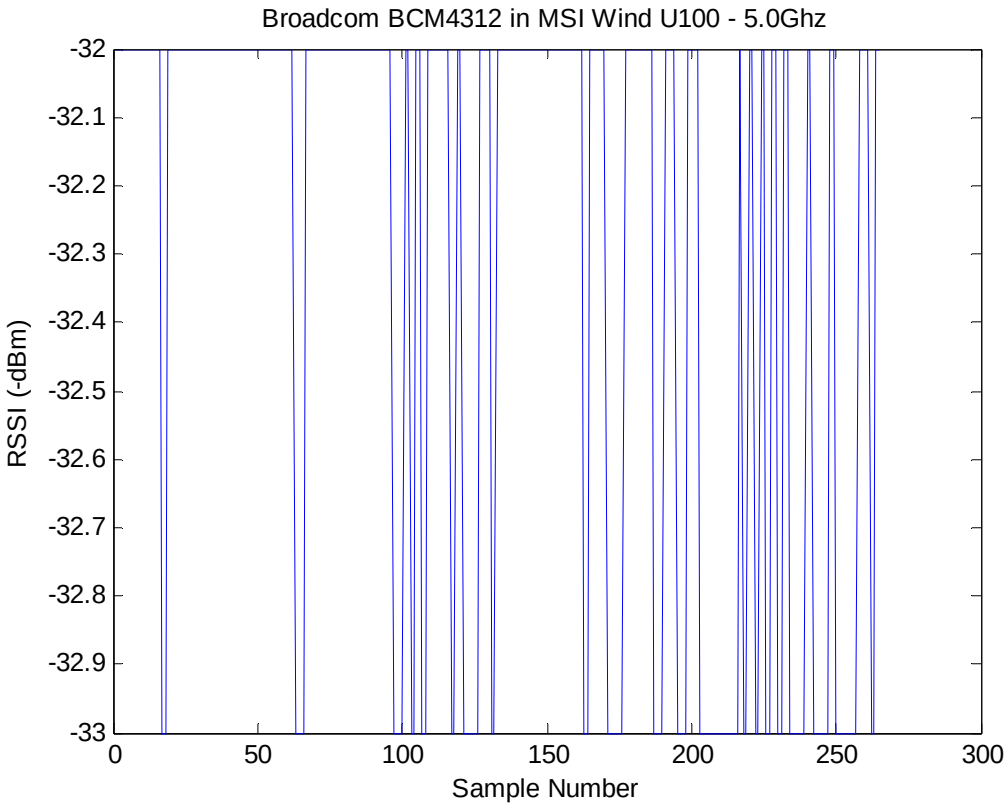

Diamond Digital A101 Card B
RSSI (-dBm)
-20
-25
-30
-35
-40
-45
-50
-55
-60
0
50
100
150
200
250
Sample Number

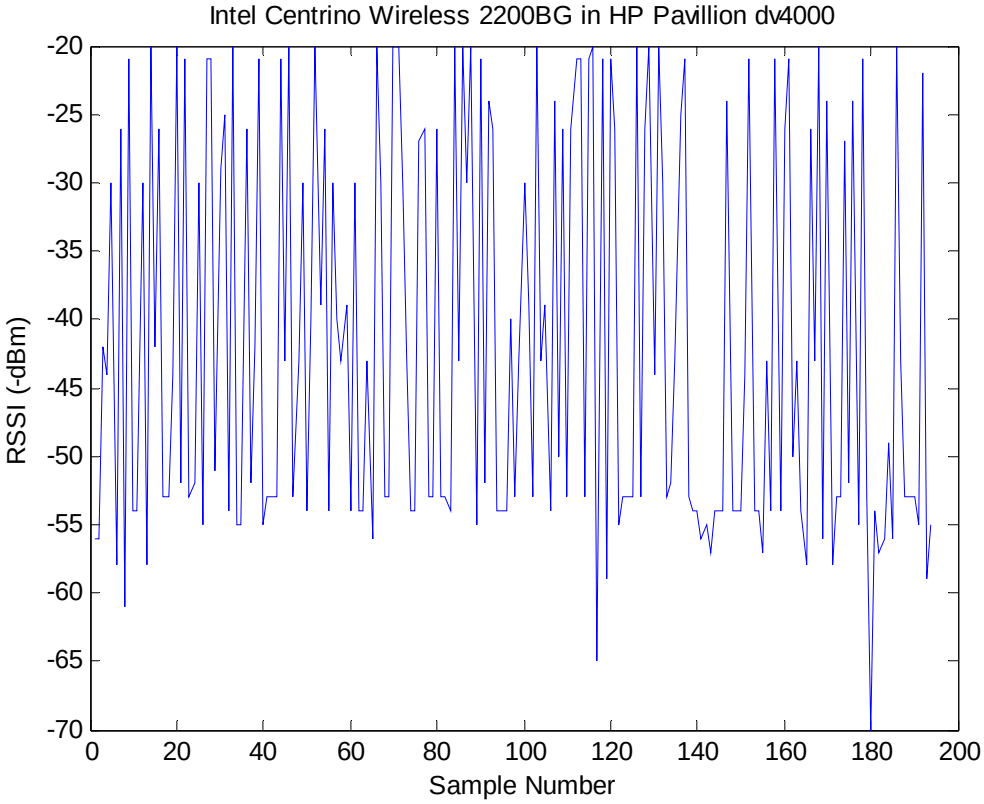

Netgear MA101
-20
-25
-30
-35
-40
-45
-50
RSSI (-dBm)
0
50
100
150
200
250
Sample Number

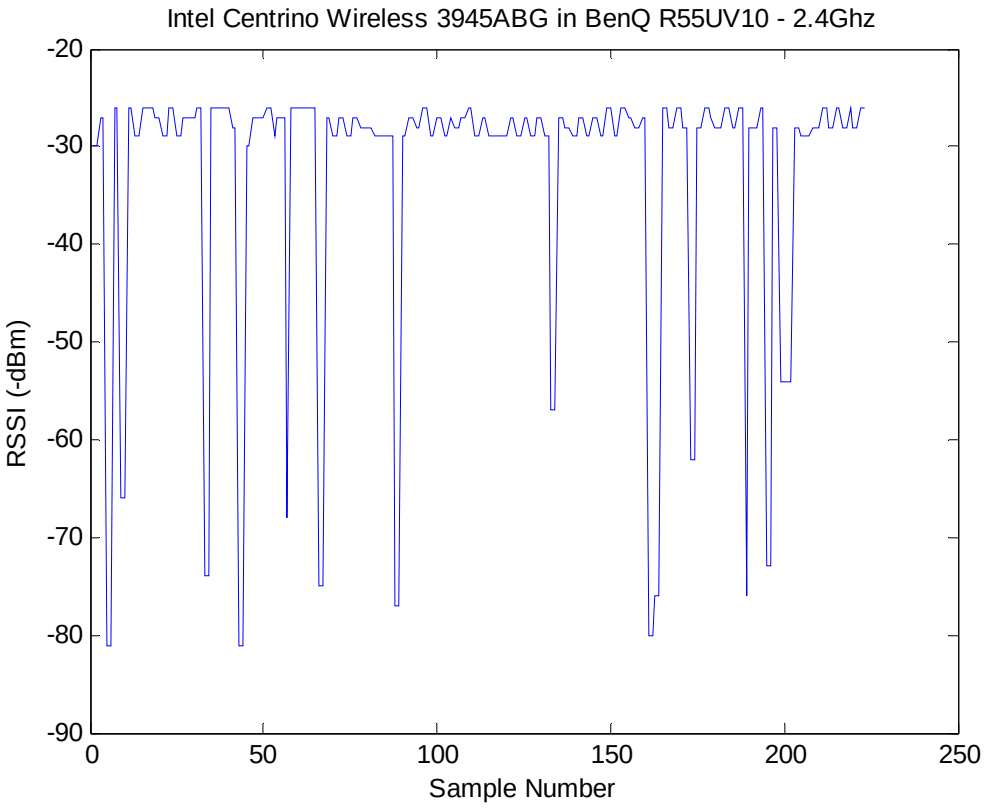

D-Link DWL-122G
-10
-15
-20
-25
-30
-35
-40
-45
-50
RSSI (-dBm)
0
50
100
150
200
250
Sample Number

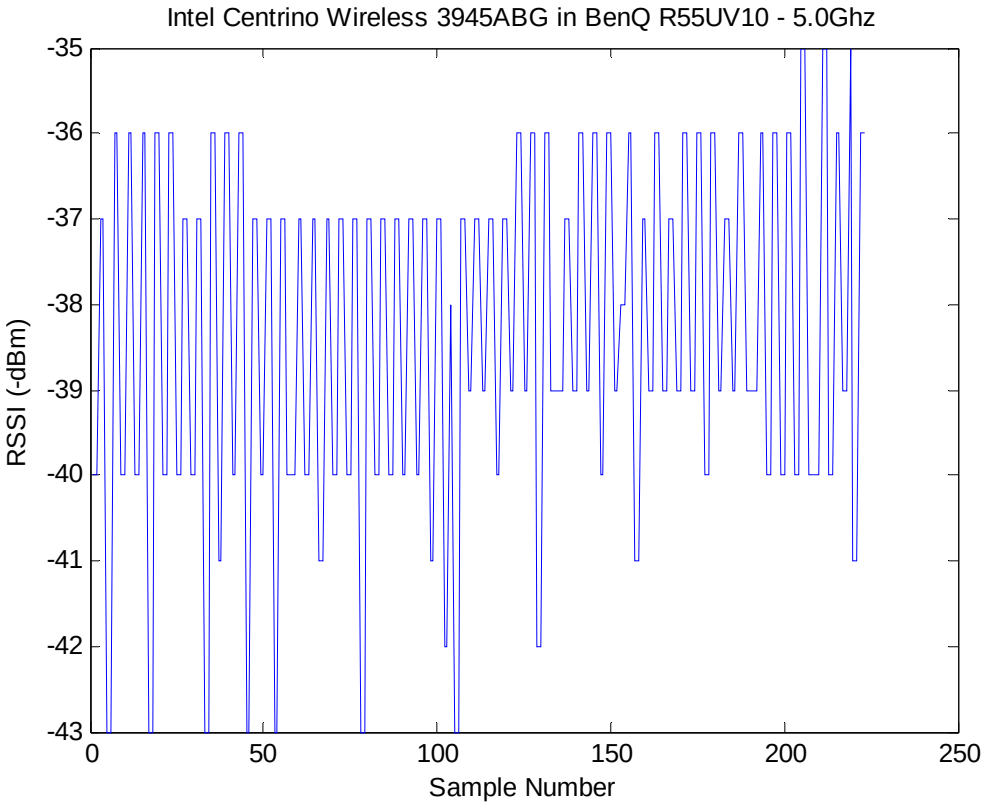

Atheros 5006UG in Asus EeePC701
RSSI (-dBm)
-26
-28
-30
-32
-34
-36
-38
-40
-42
0
50
100
150
200
250
Sample Number

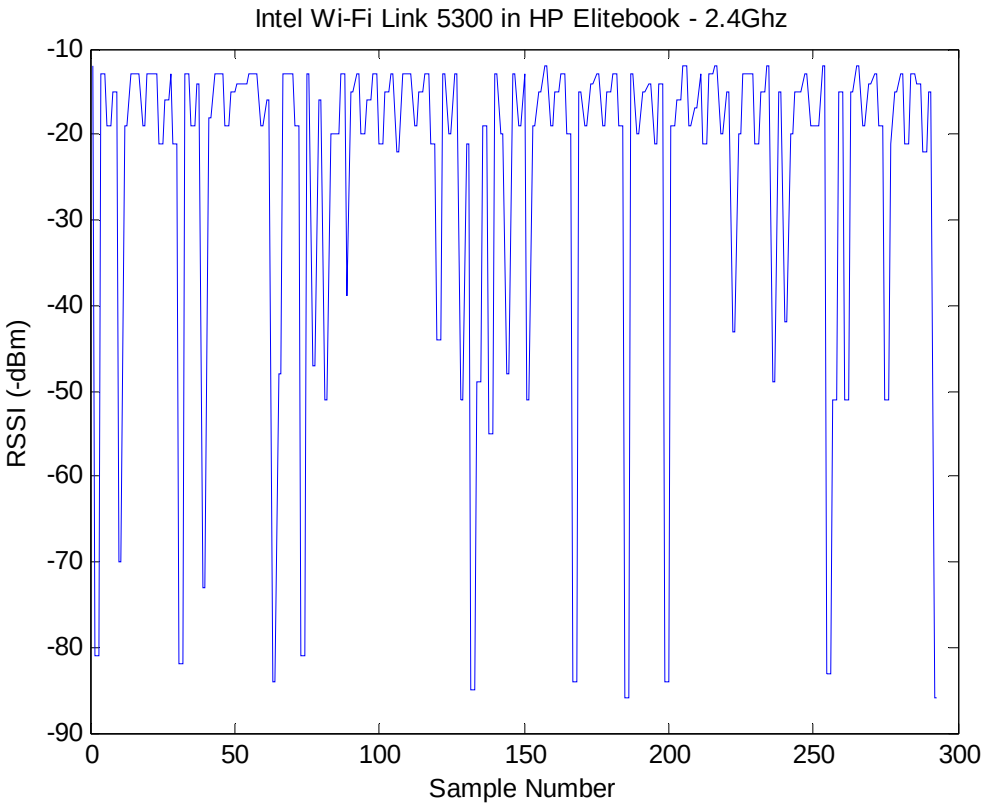

Broadcom BCM4312 in MSI Wind U100 - 2.4Ghz
RSSI (-dBm)
50
40
30
20
10
0
-10
-20
-30
-40
-50
0
50
100
150
200
250
300
Sample Number

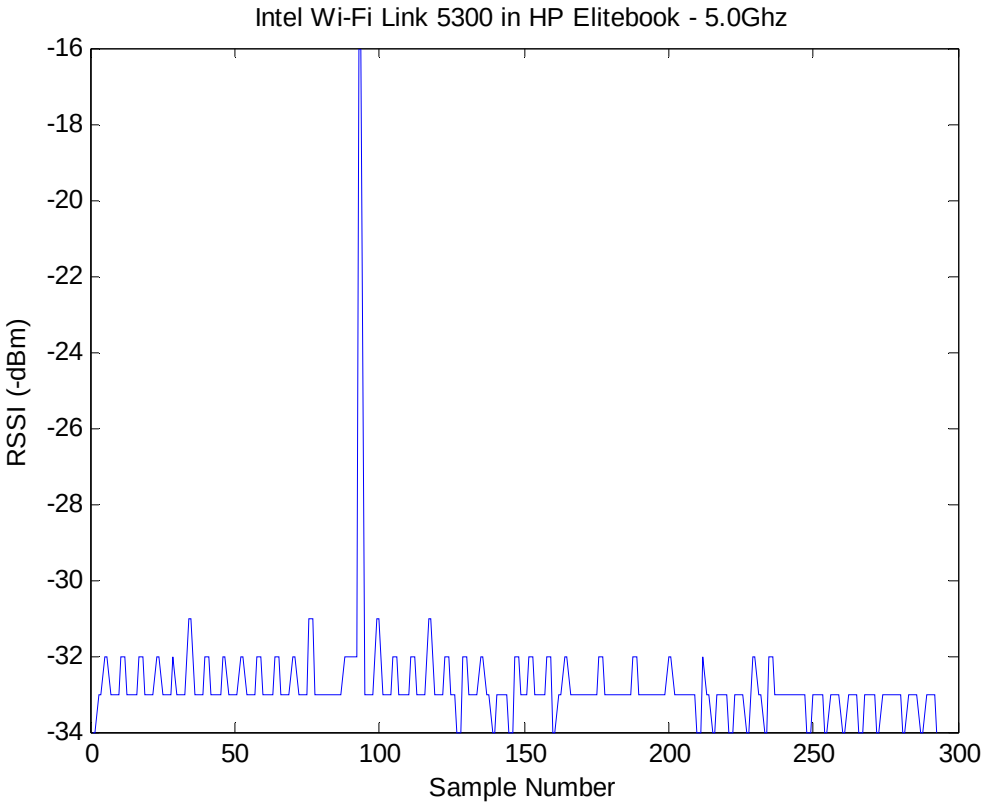

Broadcom BCM4312 in MSI Wind U100 - 5.0Ghz
RSSI (-dBm)
-32
-32.1
-32.2
-32.3
-32.4
-32.5
-32.6
-32.7
-32.8
-32.9
-33
0
50
100
150
200
250
300
Sample Number

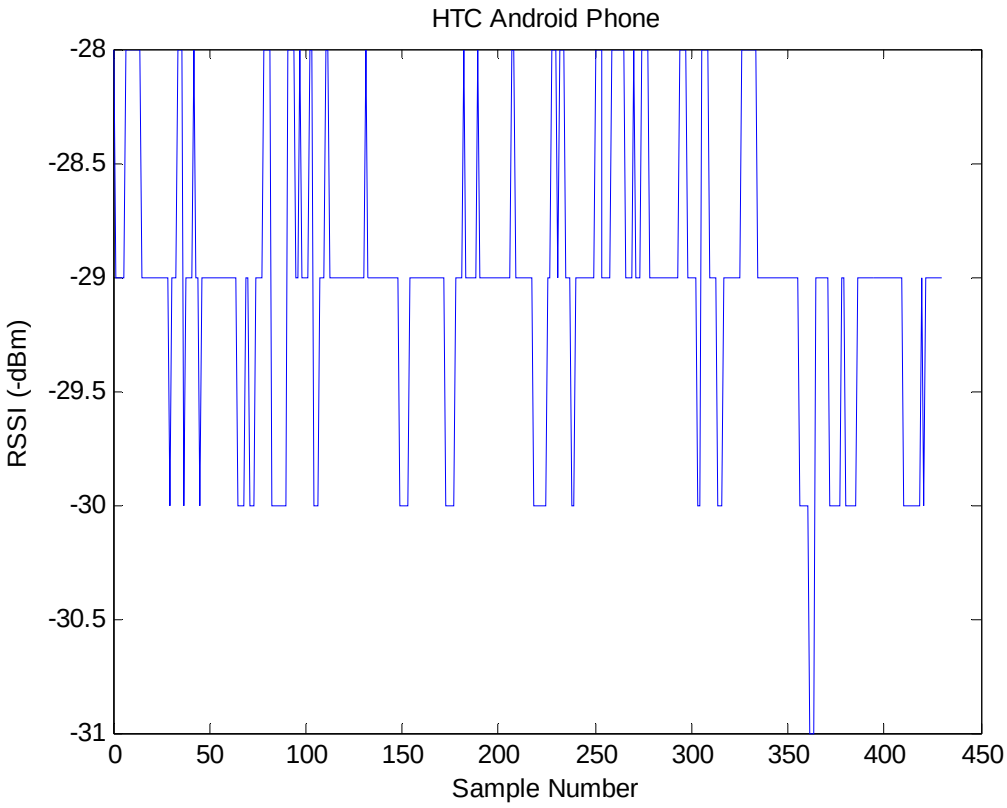

Intel Centrino Wireless 2200BG in HP Pavillion dv4000
RSSI (-dBm)
-20
-25
-30
-35
-40
-45
-50
-55
-60
-65
-70
0
20
40
60
80
100
120
140
160
180
200
Sample Number

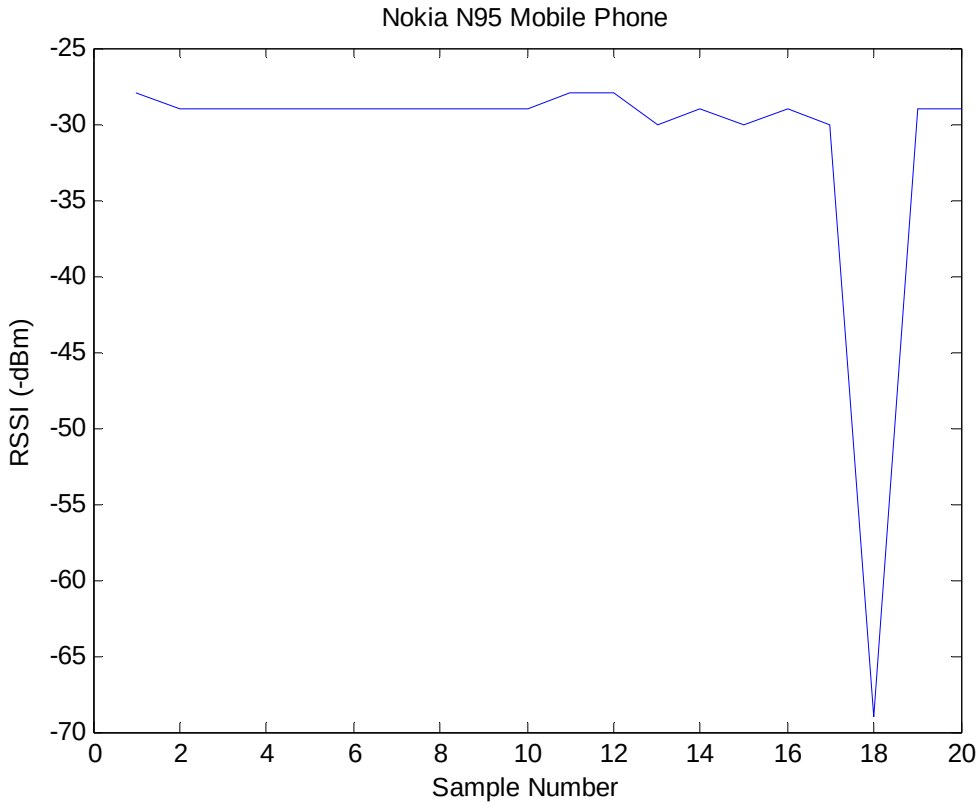

Intel Centrino Wireless 3945ABG in BenQ R55UV10 - 2.4Ghz
RSSI (-dBm)
-20
-30
-40
-50
-60
-70
-80
-90
0
50
100
150
200
250
Sample Number


Intel Centrino Wireless 3945ABG in BenQ R55UV10 - 5.0Ghz
RSSI (-dBm)
-35
-36
-37
-38
-39
-40
-41
-42
-43
0
50
100
150
200
250
Sample Number


Intel Wi-Fi Link 5300 in HP Elitebook - 2.4Ghz
RSSI (-dBm)
-10
-20
-30
-40
-50
-60
-70
-80
-90
0
50
100
150
200
250
300
Sample Number


Intel Wi-Fi Link 5300 in HP Elitebook - 5.0Ghz
RSSI (-dBm)
-16
-18
-20
-22
-24
-26
-28
-30
-32
-34
0
50
100
150
200
250
300
Sample Number


HTC Android Phone
RSSI (-dBm)
-28
-28.5
-29
-29.5
-30
-30.5
-31
0
50
100
150
200
250
300
350
400
450
Sample Number


Nokia N95 Mobile Phone
RSSI (-dBm)
-25
-30
-35
-40
-45
-50
-55
-60
-65
-70
0
2
4
6
8
10
12
14
16
18
20
Sample Number