Virtual Homonuclear Decoupling in Direct Detection NMR Experiments using Deep Neural Networks

Gogulan Karunanithy[1], Harold W Mackenzie[1], and D Flemming Hansen[1,*]

1) Department of Structural and Molecular Biology, Division of Biosciences, University College London, London, UK WC1E 6BT

To whom correspondence should be addressed:

D. Flemming Hansen;   E-mail: d.hansen@ucl.ac.uk

Abstract

Nuclear magnetic resonance (NMR) experiments are frequently complicated by the presence of homonuclear scalar couplings. For the growing body of biomolecular ^{13}C-detected NMR methods, one-bond ^{13}C-^{13}C couplings significantly reduce sensitivity and resolution. The solution to this problem has typically been to record in-phase and anti-phase (IPAP) or spin state selective excitation (S^{3}E) spectra and take linear combinations to yield singlet-resolved resonances. This however, results in a doubling of the effective phase cycle and requires additional delays and pulses to create the necessary magnetisation. Here, we propose an alternative method of virtual decoupling using deep neural networks. This methodology requires only the in-phase spectra, halving the experimental time and, by decoupling signals, gives a significant boost in resolution while concomitantly doubling sensitivity relative to the in-phase spectrum. We successfully apply this methodology to virtually decouple in-phase CON (^{13}CO-^{15}N) protein NMR spectra and ^{13}C-^{13}C correlation spectra of protein side chains.

Introduction

Traditionally, solution-state biomolecular NMR has focussed on ^{1}H detection-based methods owing to its high gyromagnetic ratio and therefore sensitivity. However, over the last two decades, methods referred to as direct detection, involving the detection of ^{13}C or ^{15}N magnetisation have increasingly come to the fore[1–3]. Despite their lower intrinsic sensitivity, the absence of solvent signals, higher chemical shift dispersion of signals and favourable relaxation properties have made these direct detection methods very useful tools and offered new opportunities for elucidating the structures and behaviour of biomolecules.

One key advantage of employing direct detection is that spectra are not affected by the deleterious effects of labile ^{1}H exchange with the solvent[4]. As a result of solvent exchange, ^{1}H resonances in commonly employed ^{1}H-^{15}N backbone amide correlation experiments will sometimes be broadened beyond detection. This issue is particularly acute for intrinsically disordered proteins (IDPs), where the high degree of solvent exposure throughout the protein can lead to a significant loss of signal, particularly at physiologically relevant pH and temperatures[5,6]. Conversely, by focussing on non-labile ^{13}C and ^{15}N nuclei, approaches such as the CON experiment[7] and others can provide important insight into biomolecular behaviour, even in the presence of significant ^{1}H solvent exchange. Moreover, with respect to IDPs, where chemical shift dispersion in the ^{1}H dimension is low, using direct-detect based methods can also confer a significant resolution advantage[8–10].

A further advantage of ^{13}C- and ^{15}N-detected NMR experiments is that although low gyromagnetic ratio nuclei have reduced intrinsic sensitivity this also corresponds to lower intrinsic relaxation rates[11]. We recently harnessed this property to record high resolution ^{13}C-^{13}C correlation spectra of the side chains of uniformly deuterated proteins[12]. The slow relaxation of the ^{13}C nuclei in these deuterated systems means that spectra can be recorded on even relatively large proteins, providing a uniquely detailed insight into the structure and dynamics of protein side chains. It has also been shown that even on protonated samples, the additional steps in a pulse sequence required to return the magnetisation back to protons and subsequent signal loss due to solvent exchange, can result in {^{1}H-excite, ^{13}C-detect} sequences having higher overall sensitivity in certain cases[13,14].

Despite these clear advantages, an important issue that affects ^{13}C direct-detection experiments is the presence of large one bond ^{13}C-^{13}C scalar couplings, typically 30-55 Hz in size, which evolve during acquisition[15]. This scalar coupling doubles the number of peaks in

the spectrum, significantly reducing their resolution and complicating interpretation. As the couplings are homonuclear, decoupling schemes can cause a range of deleterious effects in the spectra including sidebands, Bloch-Siegert shifts, and severe relaxation losses[16,17]. Consequently, the preferred method for eliminating these one-bond couplings involves virtual decoupling.

The two most common methods for virtually decoupling one-bond scalar couplings in direct-detection methods are in-phase anti-phase[18] (IPAP) and spin-state selective[19] (S^{3}E) detection. In both cases, two spectra are required, doubling the effective phase cycle of the experiment. For IPAP, virtual decoupling of the spectra is obtained by recording spectra with in-phase and anti-phase magnetisation detected, respectively. Taking the sum and difference of these two spectra yields two spectra each containing a peak at one component of the doublet. The peaks are subsequently centred by shifting them by plus or minus half the coupling constant and then added together to give the final result (Fig 1). S^{3}E follows a similar scheme except that in the detected magnetisation the two components of the doublets are 90° out of phase and in the second spectrum the sign of one of the peaks is reversed. Subsequently, after taking linear combinations of the two spectra, one of the spectra requires an additional 90° phase correction before the spectra are summed.

Both the IPAP and S^{3}E methods result in a doubling of the time required to perform the experiment and require additional time in the pulse sequences to create the requisite magnetisation, causing additional signal loss due to relaxation. The S^{3}E methodology is slightly less onerous in terms of time (taking time $0.25/J$ rather than $0.5/J$ required for IPAP), but is also less robust, when there are slight variations in the scalar couplings in the spectrum. An alternative method of virtual decoupling that avoids these disadvantages is to use a signal processing algorithm that produces the decoupled spectrum from one recorded coupled spectrum. Several methods for virtual decoupling by deconvolution have been developed, most notably in the context of ^{13}C direct detection is deconvolution with maximum entropy reconstruction[20–22]. However, despite the potential benefits associated with decoupling via signal processing, these methods lack widespread adoption and, in most cases, IPAP or S^{3}E virtual decoupling is still employed.

Recently, we demonstrated that a deep neural network (DNN) based on dilated convolutional layers, FID-Net, could be trained to perform a variety of transformations on time domain NMR data including reconstructing non-uniformly sampled (NUS) spectra and

virtually decoupling $^{13}C_\alpha$-$^{13}C_\beta$ couplings in triple-resonance HNCA spectra[23]. Building on this idea, we show here that FID-Net can be trained to decouple directly detected spectra using only the in-phase part of the experiment. The result is that high quality decoupled (singlet) spectra can be obtained in half the time of an IPAP or S^3E based scheme, since only a single plane is required. Furthermore, in the applications shown, the DNN effectively decouples spectra while also suppressing noise. Thus, overall, our method gives a large increase in resolution and sensitivity (over a factor of two) compared to a directly detected in-phase plane alone in half the time of full IPAP detection.
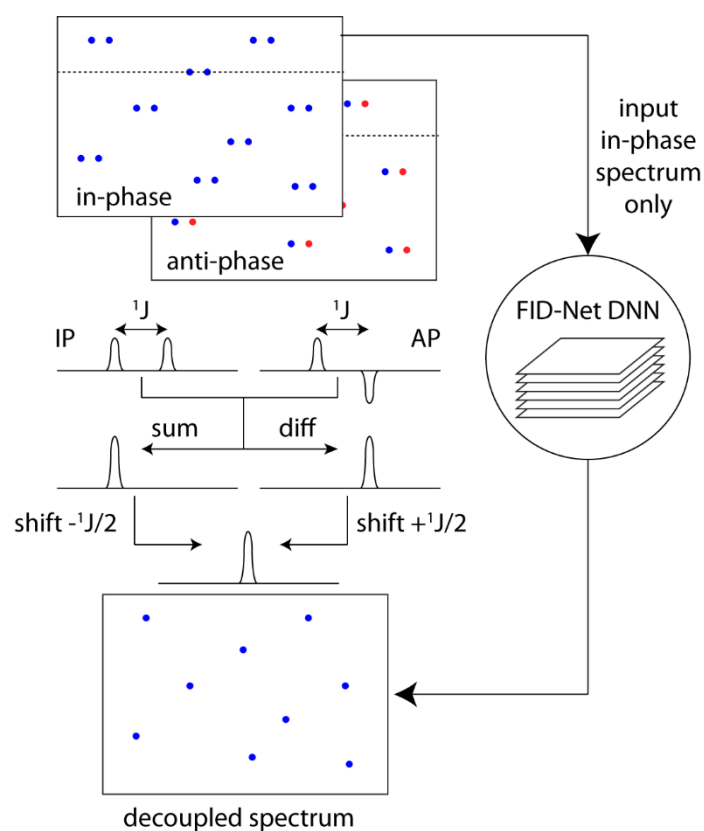

Fig 1 Schematic comparison of IPAP and FID-Net based virtual decoupling. IPAP decoupling requires in-phase and anti-phase spectra to be collected from which linear combinations are taken and peaks are centred, giving the final result. Conversely FID-Net decouples the spectrum on the basis of the in-phase spectrum alone.

Results and Discussion

We have recently described the FID-Net DNN architecture in detail[23] and limit ourselves here to a brief outline of its features. When using a DNN to model time domain NMR data (FIDs; free induction decays) a key challenge to overcome is that information on the resonating nuclei is not localised to a specific part of an FID but rather contained over its entire length. Consequently, in order to analyse and manipulate this kind of data it is necessary for a DNN to pick up on both short- and long-range patterns within the data. In FID-Net, this is achieved using stacked convolutional layers with different dilation rates[24] (Fig S1). A dilation refers to a skipped point in a convolution and by stacking convolutions with increasing dilation it is possible to create what is approximately a single convolutional layer with a very large kernel size but with much higher efficiency.

Here, the FID-Net based DNNs were trained to perform virtual decoupling of ^{13}C directly detected experiments using only the in-phase component of an IPAP acquisition. In all cases here the anti-phase component of magnetisation is also collected allowing us to compare the DNN result with the traditional full IPAP virtually decoupled result, however in general only the in-phase component is required as input for the DNN. Using the DNN method of virtual decoupling on the basis of a single spectrum halves the experiment time as only one component of the magnetisation is required and removes the need for additional delays in the sequence to create the required magnetisation and balance relaxation effects, potentially increasing signal at the start of the acquisition time.

Firstly, we test the FID-Net on ^{13}C-detected experiments developed in our own lab to characterise protein side chains[12]. These experiments are performed on uniformly [^{2}H,^{13}C] isotopically labelled proteins and rely on the existence of 'terminal' ^{13}C resonances (^{13}C$_t$) in amino acid side chains that are attached to only one other 'penultimate' ^{13}C nucleus (^{13}C$_p$), which, in turn, has a significantly different chemical shift distribution. Using selective pulses, 2D ^{13}C$_t$-^{13}C$_p$ correlation maps are obtained, providing an effective means of probing the structure and dynamics of side chains, even in large proteins. The scalar coupling between ^{13}C$_t$ and ^{13}C$_p$ (~35 Hz) evolves during the direct acquisition period resulting in a doublet. In the original experiment this was resolved using the IPAP method.

In Figure 2 the performance of DNN-decoupling using FID-Net is compared with the traditional decoupling IPAP method for ^{13}C-^{13}C correlation maps of the 18 kDa protein T4 Lysozyme (T4L). Panels (A) and (B) of the figure show the results of decoupling the valine
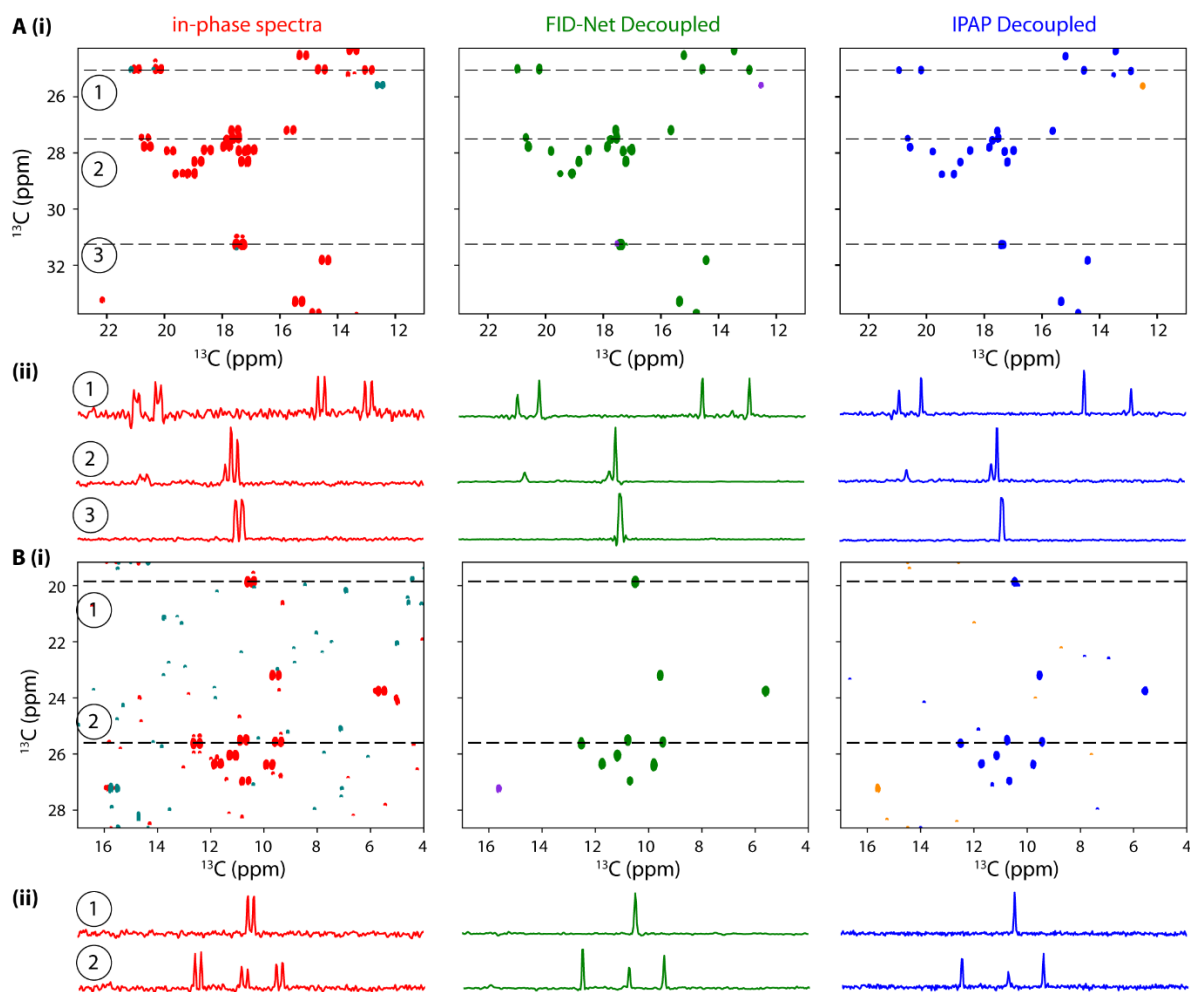
Fig 2 Comparison of in-phase (red), FID-Net decoupled (green) and IPAP decoupled (blue) $^{13}C_t$-$^{13}C_p$ spectra for the **(A)** $^{13}C_{\gamma 1,2}$-$^{13}C_\beta$ valine and **(B)** $^{13}C_{\delta 1}$-$^{13}C_{\gamma 1}$ isoleucine spectra of [^{2}H,^{13}C] labelled T4L recorded at 14.1 T (600 MHz) and at a temperature of 298 K. For all spectra, the alternative colour indicates negative contour levels. The FID-Net DNN carries out the decoupling procedure using only the in-phase spectra while the IPAP procedure uses both the in-phase and anti-phase (not shown) spectra. 1D slices from the indicated regions of each spectra are shown in **(ii)**. For 2D spectra, FID-Net decoupled and IPAP decoupled spectra are shown at the same contour levels relative to the maximum signal, facilitating comparisons of signal-to-noise between the spectra.

$^{13}C_{\gamma1,2}$, $^{13}C_\beta$ and isoleucine $^{13}C_{\delta1}$-$^{13}C_{\gamma1}$ spectra respectively. The spectra clearly demonstrate that the FID-Net decoupling provides comparable or better decoupling than the IPAP method in half the recording time. The 1D slices taken from the spectra show that the decoupling fidelity in the FID-Net spectra is excellent, both in terms of peak positions and intensities, and the procedure works well even when many peaks are present in a slice and with low intensity peaks.

An important issue to address is that of thermal noise in the spectrum. Although recording both an in-phase and an anti-phase spectrum doubles the time requirement of a direct-detect experiment, taking linear combinations for the two components gives a $\sqrt{2}$ enhancement in the signal-to-noise for each component. A further $\sqrt{2}$ enhancement in signal-to-noise is then obtained when the two components of the doublet are centred and summed to give the final spectra giving a total factor of 2 improvement of signal-to-noise of the singlet compared to each of the components of the in-phase spectrum alone. As shown in Figure 2, the signal-to-noise ratio in the FID-Net decoupled spectra is substantially higher than for the in-phase spectra alone and also comparable to or greater than the IPAP decoupled spectrum. The reason for this is that, in addition to decoupling the spectra, the DNN is also trained to minimise the noise in the spectrum. The consequence is that for these spectra the DNN approach of decoupling offers comparable or higher quality spectra in half the time.

While benchmarks on the $^{13}C_t$-$^{13}C_p$ spectra show the potential of a DNN-based approach for virtual decoupling, each of these spectra has relatively few peaks that are well dispersed. To assess the robustness of the DNN-based decoupling method we switched our attention to the CON class of experiments, that provide ^{13}CO-^{15}N correlations maps[7]. This experiment can be considered the ^{13}C direct detect analogue of ^1H-^{15}N HSQC spectra and they have become a very useful in the arsenal of NMR spectroscopists studying IDPs[8]. For these systems, the large degree of solvent exposure can render many peaks broadened beyond detection in classical ^{15}N-^1H HSQC spectra and moreover, ^{13}CO chemical shifts have significantly larger chemical shift dispersion than ^1H making interpretation and assignment of the spectra easier. Furthermore, the flexibility of IDPs leads to favourable relaxation properties, which reduces the penalty associated with the lower intrinsic sensitivity of ^{13}C detection. It should be noted, however, that the sensitivity of these experiments can in some cases be significantly boosted by employing sequences that initially excite ^1H magnetisation[13]. CON spectra are also able to provide information on proline residues that are invisible in HSQC spectra.
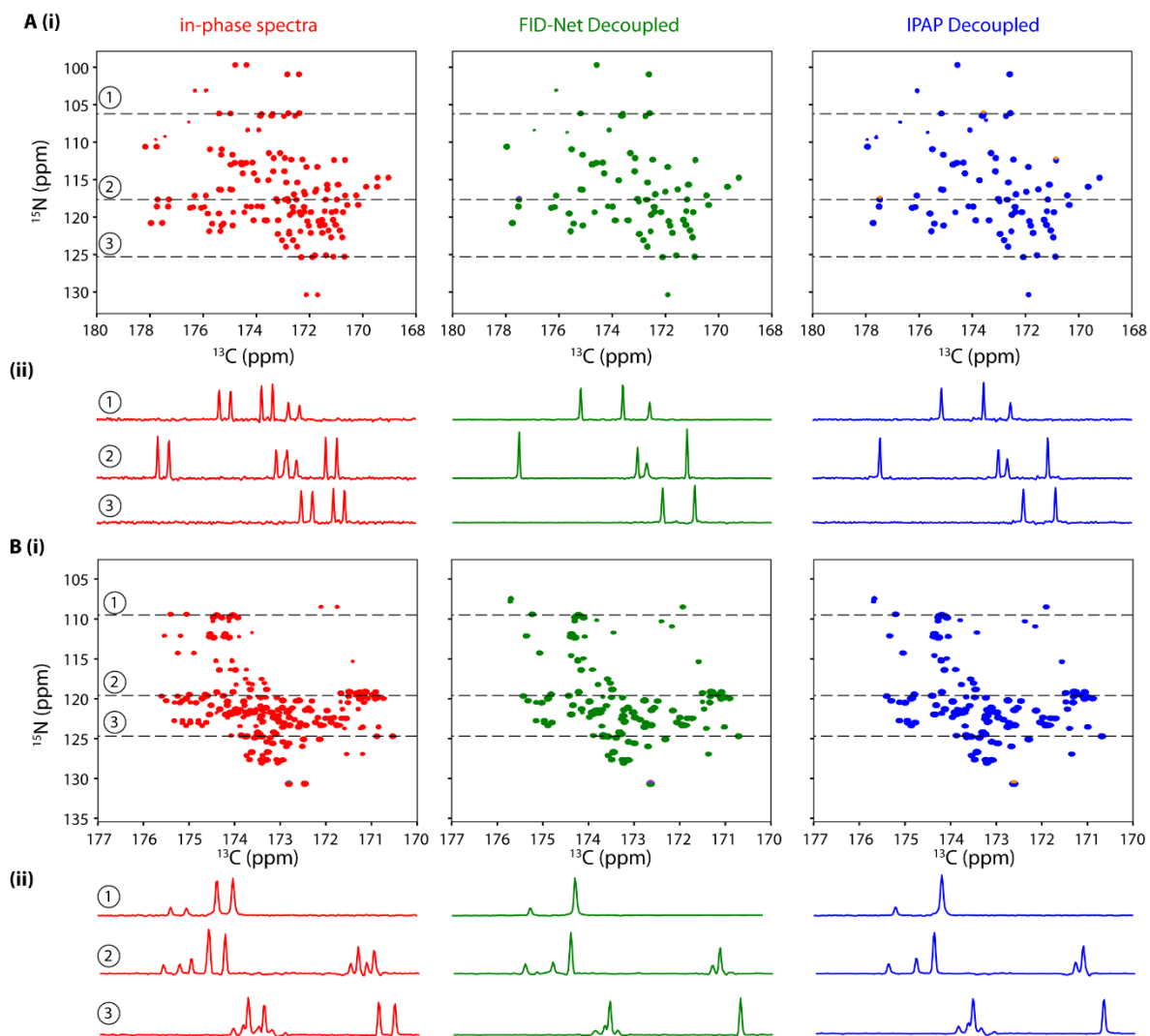
8

Fig 3 Comparison of in-phase (red), FID-Net decoupled (green) and IPAP decoupled (blue) CON spectra for **(A)** ubiquitin and **(B)** α-synuclein. **(ii)** 1D slices for each of the spectra from the indicated regions. Full acquisition details for the spectra are provided in the supporting information.

For decoupling CON spectra using FID-Net a new network was trained, using synthetic data, where the coupling constant was between 48-56 Hz. Unlike for the $^{13}C_t$-$^{13}C_p$ spectra where only one residue can be interrogated at a time, severely restricting the total number of signals in the spectrum, a CON experiment simultaneously excites all backbone residues. Therefore, the spectrum will contain many more cross-peaks and a much greater probability of overlap between resonances, making the task of virtual homonuclear decoupling more demanding.

In Figure 3 we compare the results of decoupling using the FID-Net method and IPAP for CON spectra of the proteins ubiquitin and α-synuclein. In both cases we see that the DNN-approach is able to decouple the spectra effectively and shows very favourable performance compared to the IPAP approach, even in the presence of substantially more cross-peaks. Notably, even when the components of different doublets are overlapping the DNN is able to correctly decouple the peaks. As shown in Figures S2 and S3 all expected cross-peaks are recovered in the FID-Net decoupled spectra for both proteins. The ability to virtually decouple CON spectra containing large numbers of peaks further demonstrates the robustness of the DNN decoupling approach and its viability as an alternative to IPAP or S^{3}E methods. As previously mentioned, the FID-Net DNN architecture is trained to not only decouple spectra but also minimise the noise. An important aspect to address here is therefore the point at which the methodology breaks down. During training of the DNN, random normal noise with standard deviation between 0.001 and 0.15 times the maximum signal is added, meaning that the network should be robust to low signal to noise. To test this, varying amounts of random noise was added to the most challenging spectrum, the CON spectrum of α-synuclein, and the decoupling was assessed for both DNN-based decoupling and IPAP decoupling. The results of this assessment are shown in Figure 4. It is seen that the DNN-based method remains robust even in the presence in large amounts of noise. Specifically, peaks that remain clearly distinguishable above the noise remain well decoupled and the underlying noise in the spectra is kept minimal – particularly in comparison to IPAP approach. It is clear, however, that when a cross-peak is close in intensity to the noise it can be suppressed in the DNN-approach, resulting in it missing from the spectrum or having a distorted peak shape and/or intensity. Overall, we find that the DNN-approach has a preference for peaks with very low signal-to-noise to not be present in the decoupled spectrum as opposed to generating artefacts. In Figure S4 we systematically test the performance of the network on decoupling a synthetic spectrum,
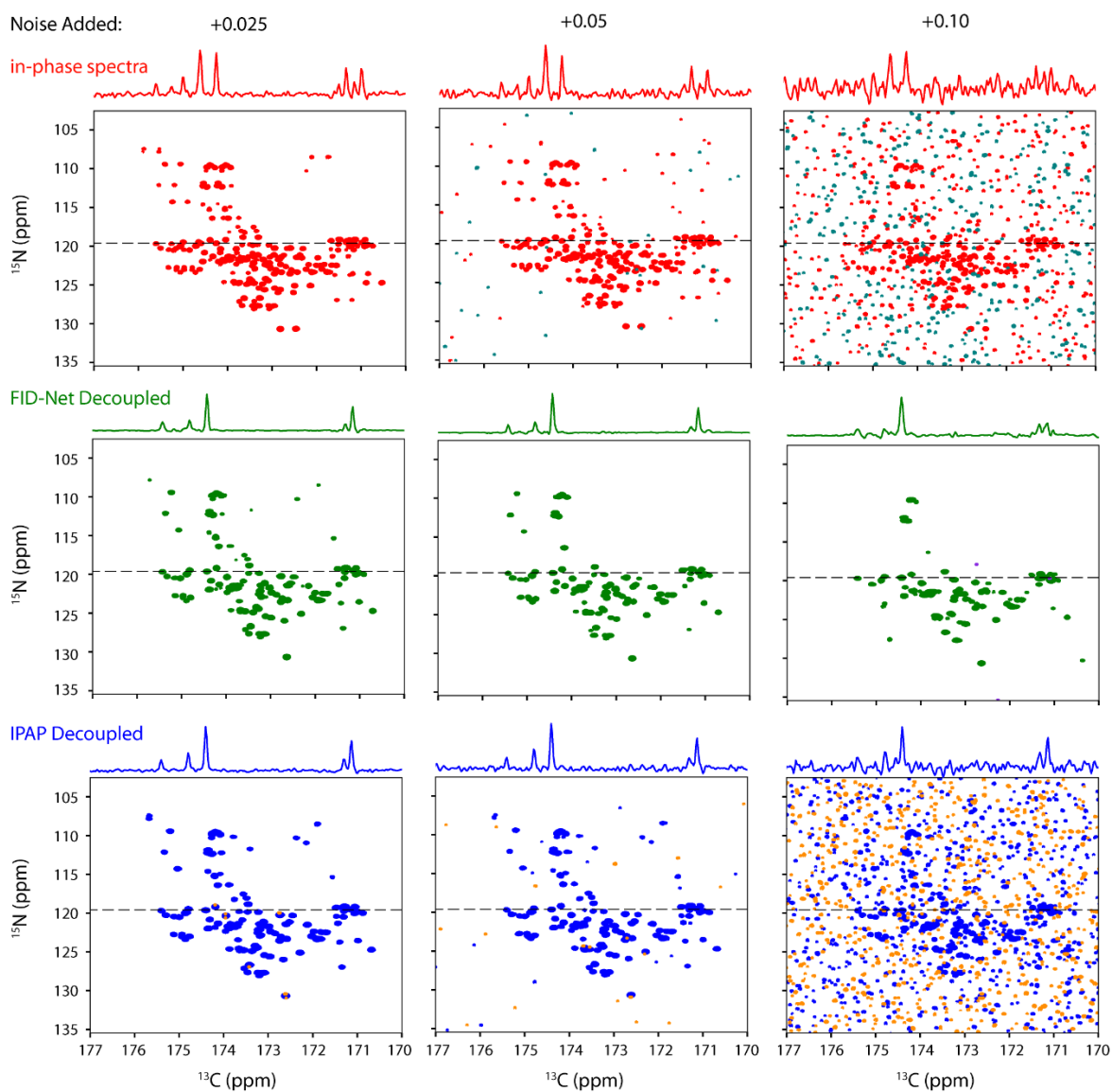
Fig 4 Comparing the outcome of decoupling following the addition of random normal noise with standard deviations of 0.025, 0.05 and 0.10 times the maximum signal. The FID-Net and IPAP decoupled spectra are plotted at the same contour level to allow for comparisons.

demonstrating that peaks above the noise threshold are effectively decoupled whilst those that cannot be reliably differentiated from noise are suppressed. It should be borne in mind that here we compare FID-Net that has only been recorded for half the time compared to the IPAP decoupled scheme. Clearly, if the in-phase spectrum were recorded for double the time this would improve the signal to noise in the in-phase spectrum and therefore the performance of the FID-Net decoupling. However, for samples where signal-to-noise is very low, the IPAP method is likely preferable as it provides a boost in signal to noise by a factor of two compared to the in-phase spectrum alone while also removing couplings.

A key guiding principle in the design of our FID-Net architecture is that a single network should be generalisable to a range of situations – a network that is able to do this should have good robustness. All the results shown above are from only two DNNs: one trained to decouple $^{13}C_t$-$^{13}C_p$ spectra and one trained to decouple CON spectra. This versatility means that there is no requirement for the end user to perform any further training and the methodology can therefore easily be implemented.

Conclusion

A perennial challenge facing many ^{13}C-direct detect NMR experiments is the presence of homonuclear scalar couplings that evolve during acquisition. While virtual decoupling via IPAP or S^{3}E methods can solve this issue, this is at the cost of doubling the duration of the experiment. Here we have demonstrated how the FID-Net architecture can be trained to perform the decoupling on the basis of the in-phase spectra alone with comparable or better results than a full IPAP experiment. This methodology can be simply implemented as part of the spectral processing procedure, offering substantial sensitivity gains for direct detect experiments. The methodology presented here adds to the growing applications of deep learning and artificial intelligence for analysing NMR data and paves the way for forthcoming developments to transform the way in which NMR spectroscopy data are processed and analysed.

Code Availability

Code (python) for virtual decoupling CON and ^{13}C$_t$-^{13}C$_p$ in-phase spectra using FID-Net (including pre-trained networks and examples) is available on GitHub: https://github.com/gogulan-k/FID-Net.

References

[1] W. Bermel, I. C. Felli, R. Kümmerle, R. Pierattelli, *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **2008**, *32*, 183–200.

[2] Z. Serber, C. Richter, D. Moskau, J. M. Böhlen, T. Gerfin, D. Marek, M. Häberli, L. Baselgia, F. Laukien, A. S. Stern, et al., *J. Am. Chem. Soc.* **2000**, *122*, 3554–3555.

[3] K. Takeuchi, H. Arthanari, M. Imai, G. Wagner, I. Shimada, *J. Biomol. NMR* **2016**, *64*, 143–151.

[4] M. Bastidas, E. B. Gibbs, D. Sahu, S. A. Showalter, *Inc. Concepts Magn Reson Part A* **2015**, *44*, 54–66.

[5] E. B. Gibbs, R. W. Kriwacki, *Methods* **2018**, *138–139*, 39–46.

[6] S. Gil, T. Hošek, Z. Solyom, R. Kümmerle, B. Brutscher, R. Pierattelli, I. C. Felli, *Angew. Chemie Int. Ed.* **2013**, *52*, 11808–11812.

[7] W. Bermel, I. Bertini, I. C. Felli, R. Kümmerle, R. Pierattelli, *J. Magn. Reson.* **2006**, *178*, 56–64.

[8] E. C. Cook, G. A. Usher, S. A. Showalter, *Methods Enzymol.* **2018**, *611*, 81–100.

[9] I. C. Felli, R. Pierattelli, *J. Magn. Reson.* **2014**, *241*, 115–125.

[10] K. Takeuchi, H. Arthanari, I. Shimada, G. Wagner, *J. Biomol. NMR* **2015**, *63*, 323–331.

[11] S. Chhabra, P. Fischer, K. Takeuchi, A. Dubey, J. J. Ziarek, A. Boeszoermenyi, D. Mathieu, W. Bermel, N. E. Davey, G. Wagner, et al., *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E1710–E1719.

[12] R. B. Pritchard, D. F. Hansen, *Nat. Commun.* **2019**, *10*.

[13] W. Bermel, I. Bertini, I. C. Felli, R. Pierattelli, *J. Am. Chem. Soc.* **2009**, *131*, 15339–15345.

[14] G. Karunanithy, J. Reinstein, D. F. Hansen, *J. Phys. Chem. Lett.* **2020**, *11*, 5649–5654.

[15] W. Bermel, I. Bertini, I. C. Felli, M. Piccioli, R. Pierattelli, *Prog. Nucl. Magn. Reson. Spectrosc.* **2006**, *48*, 25–45.

[16] F. Bloch, A. Siegert, *Phys. Rev.* **1940**, *57*, 522–527.

[17] J. Ying, F. Li, J. H. Lee, A. Bax, *J. Biomol. NMR* **2014**, *60*, 15–21.

[18] M. Ottiger, F. Delaglio, A. Bax, *J. Magn. Reson.* **1998**, *131*, 373–378.

[19] M. D. Sørensen, A. Meissner, O. W. Sørensen, *J. Biomol. NMR* **1997**, *10*, 181–186.

[20] N. Shimba, A. S. Stern, C. S. Craik, J. C. Hoch, V. Dötsch, *J. Am. Chem. Soc.* **2003**, *125*, 2382–2383.

[21] Z. Serber, C. Richter, V. Dötsch, *ChemBioChem* **2001**, *2*, 247–251.

[22] M. A. Delsuc, G. C. Levy, *J. Magn. Reson.* **1988**, *76*, 306–315.

[23] G. Karunanithy, D. F. Hansen, *Chemrxiv* **2020**

[24] A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, K. Kavukcuoglu, *arXiv* **2016**, 1–15.

[25] M. W. Maciejewski, A. D. Schuyler, M. R. Gryk, I. I. Moraru, P. R. Romero, E. L. Ulrich, H. R. Eghbalnia, M. Livny, F. Delaglio, J. C. Hoch, *Biophys. J.* **2017**, *112*, 1529–1534.