ORIGINAL ARTICLE

Channel-Selective Information Regulation for Low-SNR Automatic Modulation Recognition

Isuru Fernando[1], Madushan Karunaratne[2], and Sajith Abeysekara[3]

[1]University of Vavuniya, Park Road, Vavuniya 43000, Sri Lanka
[2]University of Ruhuna, Wellamadama University Road, Matara 81000, Sri Lanka
[3]Gampaha Wickramarachchi University of Indigenous Medicine, Yakkala Road, Yakkala 11870, Sri Lanka

ABSTRACT

Wireless receivers operating in dense and contested spectral environments must often infer modulation type from short, corrupted, and weak observations. In such settings, the classification problem is shaped not only by additive noise but also by oscillator mismatch, fading, interference, finite observation windows, and the uneven distribution of informative signal structure across time and feature channels. The practical consequence is that many deep classifiers fail not because they lack nominal capacity, but because they allocate representational effort inefficiently: they preserve redundant activations, over-amplify nuisance responses, and allow fragile class evidence to be buried beneath feature congestion. This paper develops a technical research framework for low-SNR automatic modulation recognition centered on the concept of information regulation in hierarchical feature systems. The analysis treats deep modulation classifiers as nonlinear operators that compress, route, and reweight class evidence under noise and derives a representation-theoretic view of why channel-selective mechanisms become especially important when the latent signal manifold is weakly expressed. A mathematical formulation is given for the received waveform, multistage feature mapping, channel-gating dynamics, class margin evolution, and error concentration under adverse conditions. The paper then examines how channel-selective compression interacts with residual transport, multiscale memory, and global context summaries to improve latent separability without uncontrolled growth in computation. Rather than presenting architecture as an isolated engineering recipe, the study interprets low-SNR recognition through the geometry of preserved versus overloaded feature subspaces, the stability of decision boundaries under nuisance perturbation, and the statistical efficiency of selective feature amplification. The resulting treatment offers a principled account of why robust modulation recognition depends on controlling information flow as much as on increasing network depth or width.

1 | Introduction

Automatic modulation recognition is one of the most technically layered problems in modern signal intelligence and adaptive communication systems because it lies at the intersection of statistical inference, nonlinear representation learning, and physical channel uncertainty [1]. At a superficial level, the task appears to be a multiclass classification problem in which a receiver observes a finite sequence of in-phase and quadrature samples and must assign that observation to one among several modulation families. In operational settings, however, the task is considerably harder. The waveform reaching the classifier is shaped by random phase rotation, residual carrier offset, time misalignment, pulse-shaping overlap, nonlinear hardware response, propagation-induced fading, cochannel distortion, and additive noise whose impact depends not only on its energy but also on the local geometry of the observation window. The true challenge is therefore not only to discriminate classes, but to discover which fragments of the observation remain informative after multiple layers of corruption have already mixed signal structure with nuisance variability.

Traditional approaches addressed this problem through explicitly designed descriptors, such as higher-order statistics, spectral moments, cyclic quantities, or constellation-derived features. These methods remain conceptually valuable because they expose physically meaningful invariants. Yet they depend strongly on assumptions about synchronization, observation length, and channel regularity. Under low signal-to-noise ratio, the reliability of such handcrafted quantities degrades quickly because the transforms from which they are derived cease to separate class structure cleanly. Deep learning changed the formulation by replacing manually prescribed descriptors with end-to-end optimization directly on raw or lightly processed signal representations. Instead of deciding beforehand which signal property should matter, the network learns a hierarchy of local and global features that are optimized for discrimination. This shift increased empirical performance on benchmark datasets and opened the possibility of adapting the classifier to diverse waveform families without redesigning the front end for every scenario [2]. Despite these advantages, end-to-end convolutional or hybrid architectures encounter a persistent difficulty in the low-SNR regime. When the underlying signal is weak, a deeper model does not automatically become a better one. Greater depth may enlarge receptive fields and abstraction, but it also multiplies the opportunities for fragile discriminative traces to be diluted or overwritten. Greater width may increase expressive capacity, yet it also increases the number of channels that can absorb nuisance-driven activation. Under these conditions, poor performance is often not best explained as insufficient parameter count. Rather, it emerges because the feature hierarchy becomes overloaded. Some channels encode useful class content, others encode repetitive transforms of the same content, and still others respond mainly to noise or distortion. Once such channels are propagated forward without regulation, later layers are forced to classify on a representation whose informational density is low even if its dimensionality is high.

This perspective suggests that the central issue in low-SNR modulation recognition is not only feature extraction but feature governance. A robust network must decide which channels deserve amplification, which channels should be attenuated, and how shallow local evidence should remain accessible after deeper contextual integration. In a noisy and finite observation regime, the classifier must preserve weak but consistent cues while suppressing activations that are energetic yet uninformative. This is different from ordinary denoising [3]. The network is not simply filtering out additive noise in the classical signal-processing sense. It is performing task-oriented reallocation of representation capacity so that the latent space becomes more aligned with class-discriminative structure than with nuisance variability.

A useful empirical clue for this viewpoint comes from a recent low-SNR modulation study. Zou et al. (2022) [4] reported that adding a squeeze–excitation structure led to only a slight change in global accuracy but a marked improvement in average accuracy over the -20 dB to 0 dB interval, and they interpreted that gain as relief of feature-map information overload through channel reweighting. That observation is conceptually important because it points to a phenomenon more specific than generic model improvement. If a modification barely affects overall accuracy yet materially improves the low-SNR regime, then the modification is not merely adding broad capacity. It is altering how the representation behaves under adversity. It is improving the ratio between useful and useless information precisely where the classifier is most vulnerable to confusion.

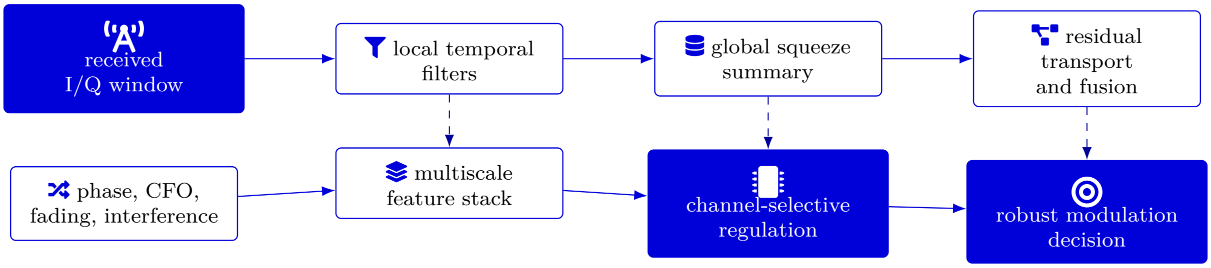

Figure 1: System-level view of channel-selective information regulation for low-SNR automatic modulation recognition. A short corrupted I/Q observation is expanded into multiscale feature channels, compressed through a global channel gate, and transported through residual fusion before classification. The figure emphasizes that the key design objective is not only feature extraction, but the disciplined preservation of weak class evidence under phase offset, fading, interference, and noise.

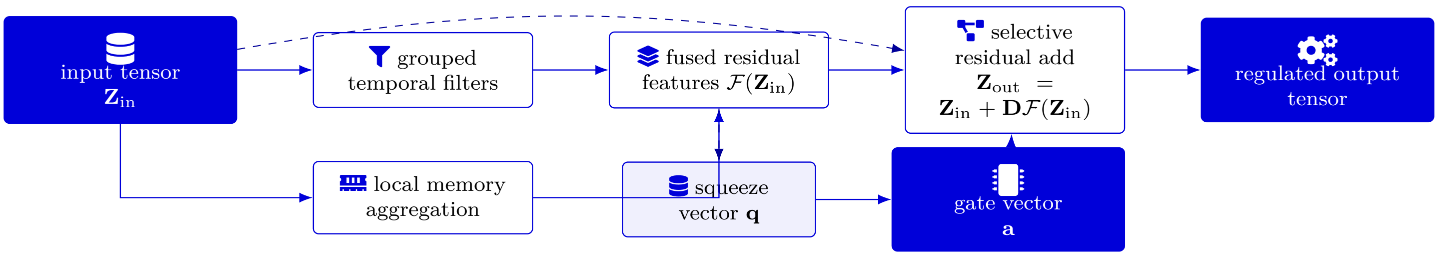

Figure 3: Residual channel-selective compression block. Multibranch local extraction first forms a residual feature tensor, which is then summarized by a squeeze vector and transformed into channel weights. The gate modulates only the residual perturbation before addition to the identity path, allowing baseline information transport to remain intact while limiting the injection of nuisance-heavy channels. This arrangement supports both stable optimization and sample-adaptive control of representational clutter.

The central hypothesis of this paper is that a deep classifier operating on complex-baseband samples can be understood as a sequence of compressive transport maps [5]. Each layer receives a tensor that contains a mixture of class content, nuisance variation, and redundant representation of earlier features. Without explicit or implicit control, this tensor can accumulate congestion. Congestion here means that the number of active channels grows or remains high while the proportion of channels contributing stable discriminative value declines. In such a state, downstream layers experience high-dimensional but low-quality evidence, and decision margins become fragile under small perturbations. Channel-selective modules, global context summarization, and structured skip transport can counter this by converting indiscriminate expansion into selective compression. This article therefore addresses several linked questions. First, how should the received signal be modeled when the observation is short, two-channel, and corrupted by multiple nuisance transformations? Second, how does information overload arise in deep feature systems for modulation recognition, and how can it be described geometrically and statistically? Third, how do channel-wise gating and residual transport affect the evolution of class manifolds, margins, and gradient flow? Fourth, what kinds of experimental behavior should be interpreted as evidence that information regulation, rather than raw scale, is driving improvement? The answers require combining ideas from communication theory, linear algebra, nonlinear dynamical systems, and modern deep representation analysis.

A useful starting principle is that modulation classes are not point targets in signal space. They are distributions over parameterized waveform families.

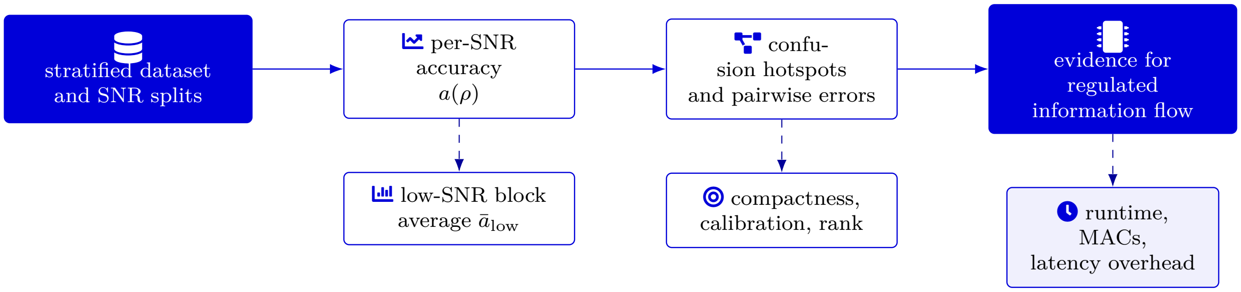

Figure 4: Experimental interpretation framework for evaluating whether performance gains arise from information regulation rather than from raw model scale. The measurement chain begins with SNR-stratified evaluation, then resolves low-SNR accuracy concentration, targeted confusions, latent compactness, effective rank, and calibration, before relating these outcomes to computational overhead. Improvements that localize in difficult SNR bands while preserving efficiency provide the strongest evidence that the architecture is managing feature congestion instead of merely increasing nominal capacity.

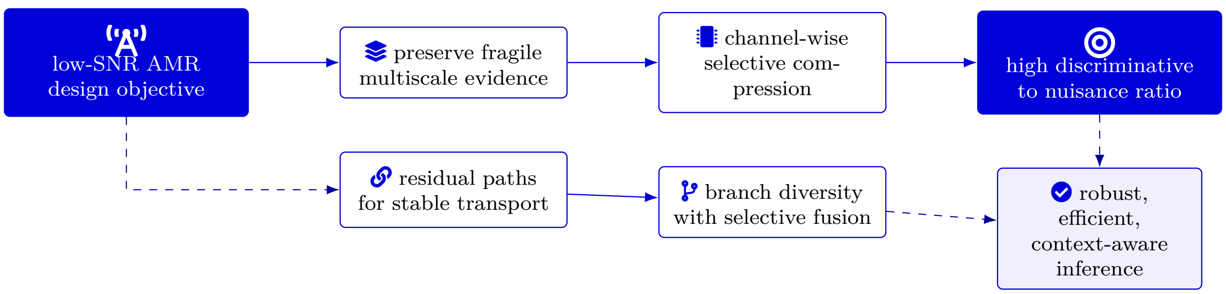

Figure 5: Architectural implications distilled from the information-regulation viewpoint. Robust modulation recognition in adverse environments requires multiscale evidence retention, residual transport, sample-adaptive channel compression, and selective fusion of diverse feature branches. The intended outcome is not maximum channel count, but a latent system in which the ratio of discriminative content to nuisance clutter remains high as the observation becomes weaker.

Even a single class generates a manifold of observations under variation in symbols, phase, timing, amplitude, and channel response. Noise thickens that manifold further. Under low SNR, the manifolds for several classes overlap strongly in the raw observation space [6]. The job of the feature extractor is to produce a new coordinate system in which overlap decreases and stable class-dependent directions become accessible to a linear or weakly nonlinear classifier. The problem is that many channels in a standard network may encode directions that are neither stable nor class-aligned. They may reflect accidental correlations produced by the training set, transitory noise patterns, or repetitive transformations of the same evidence. When such channels accumulate, the coordinate system becomes cluttered. Selective information regulation acts as a structural prior favoring channels whose global behavior is coherent with the sample-level evidence. From this point of view, channel selection is not an accessory added after feature extraction. It is part of feature extraction itself. The representation delivered to the classifier should not be interpreted as the raw output of a stack of convolutions but as the equilibrium result of extraction, transport, fusion, compression, and reweighting. This equilibrium determines how strongly the class means separate, how large the within-class covariance remains, how the gradients allocated to hard examples behave, and how sensitive the posterior becomes to nuisance perturbation. It also determines computational efficiency because a model that routes information selectively can often outperform a denser model whose extra channels mostly carry redundancy.

The paper proceeds in a staged manner. First, the signal and feature-transport model is formalized so that later arguments about overload and gating rest on a precise foundation [7]. Second, a theory of information congestion in multichannel hierarchies is introduced, together with matrix representations that make the phenomenon measurable in principle. Third,

Symbol	Description	Space
$r[t]$	Received complex-baseband sample at time index t	$\mathbb{C}$
$\mathbf{X}$	Two-channel real-valued representation of the observation window	$\mathbb{R}^{2\times T}$
y	Modulation class label	$\{1,\ldots,C\}$
T	Number of time samples in the observation window	$\mathbb{N}$
C	Number of modulation families	$\mathbb{N}$
$\nu[t]$	Additive disturbance sample at time t	$\mathbb{C}$
$\mathbf{Z}^{(\ell)}$	Hidden feature tensor at layer ℓ	$\mathbb{R}^{d_\ell\times T_\ell}$

Table 1: Key notation for the signal and observation model.

Parameter	Physical meaning	Effect on waveform
a	Amplitude scaling of the received complex baseband signal	Global gain and effective SNR
Δf	Residual carrier frequency offset	Rotating constellation and time-varying phase
ϕ	Static phase offset	Global constellation rotation across the window
$h[t;\boldsymbol{\eta}]$	Effective channel impulse response parameterized by $\boldsymbol{\eta}$	Fading, dispersion, and pulse overlap distortion
ϵ	Timing mismatch relative to symbol boundaries	Loss of ideal symbol alignment and intersymbol interference
Cochannel terms	Interfering emitters sharing time–frequency resources	Additional structured distortion beyond thermal noise

Table 2: Channel and nuisance parameters shaping the received modulation waveform.

channel-selective compression is analyzed as a nonlinear operator that reshapes covariance, spectral energy allocation, and decision margins. Fourth, the implications for optimization dynamics and generalization are examined, with particular emphasis on the uneven gradient utility of low- versus high-SNR samples. Fifth, an experimental interpretation framework is described for identifying when improvements are due to better information regulation rather than larger capacity alone. The paper concludes by distilling architectural principles for robust modulation recognition in adverse environments. The overall goal is to offer a technically coherent explanation for why low-SNR performance depends on managing information flow, not merely on stacking more layers or increasing the nominal richness of the model.

2 | Signal and Observation Model

Let the received signal be represented as a complex-baseband discrete-time process $r[t]$, observed for $t=0,1,\ldots,T-1$. For computation, the complex signal is decomposed into real channels,

$$\mathbf{X}=\begin{bmatrix}\Re r[0] & \Re r[1] & \cdots & \Re r[T-1]\\ \Im r[0] & \Im r[1] & \cdots & \Im r[T-1]\end{bmatrix}\in\mathbb{R}^{2\times T}.$$

The class label $y\in\{1,\ldots,C\}$ identifies the underlying modulation family. The received sequence is not assumed to be a clean draw from a class-conditional template. Instead, it is generated through a composition of signal formation, channel transformation, oscillator mismatch, and additive

Component	Description	Low-SNR behavior
$\mathbf{S}^{(\ell)}$	Class-stable signal content in the hidden tensor	Should become more linearly separable with depth
$\mathbf{U}^{(\ell)}$	Nuisance-sensitive or unstable content	Variance inflates under noise and channel mismatch
$\mathbf{R}^{(\ell)}$	Redundant content re-encoding earlier features	Grows with depth if not explicitly regulated
$\mathbf{Z}^{(\ell)}$	Full hidden representation $\mathbf{S}^{(\ell)} + \mathbf{U}^{(\ell)} + \mathbf{R}^{(\ell)}$	Can become congested when unstable and redundant parts dominate

Table 3: Decomposition of hidden representations into stable, unstable, and redundant components.

Metric	Definition	Purpose
$\Gamma^{(\ell)}$	Generalized margin based on class means and covariances at layer ℓ	Captures separability of class manifolds in latent space
r_{eff}	$r_{\text{eff}} = \dfrac{(\operatorname{tr}\mathbf{K})^2}{\operatorname{tr}(\mathbf{K}^2)}$ for channel covariance $\mathbf{K}$	Measures effective rank of the channel covariance
Ξ	$\Xi = \dfrac{\operatorname{tr}(\mathbf{K}_B)}{\operatorname{tr}(\mathbf{K}_W) + \epsilon}$	Quantifies discriminative efficiency of the representation
κ_c	$\kappa_c = \dfrac{1}{n_c} \sum_{i:y_i=c} \lVert \mathbf{z}_i - \boldsymbol{\mu}_c \rVert_2^2$	Within-class compactness for class c
$\delta_{c,c'}$	$\delta_{c,c'} = \lVert \boldsymbol{\mu}_c - \boldsymbol{\mu}_{c'} \rVert_2$	Between-class mean separation

Table 4: Latent-space metrics used to characterize congestion and class separability.

disturbance [8]. A broad parametric model is

$$r[t] = a\, e^{j(2\pi\Delta f t+\phi)} \sum_k s_y[k]\, p(t - k\tau - \epsilon) \\ * h[t;\boldsymbol{\eta}] + \nu[t], \tag{1}$$

where a denotes amplitude scaling, Δf is residual carrier offset, ϕ is phase offset, $s_y[k]$ are symbols drawn from the class-dependent modulation rule, $p(\cdot)$ is a pulse shape, τ is symbol spacing, ϵ is timing mismatch, $h[t;\boldsymbol{\eta}]$ is an effective channel impulse response parameterized by nuisance variables $\boldsymbol{\eta}$, and $\nu[t]$ is additive noise. The convolution with $h[t;\boldsymbol{\eta}]$ is not merely a mild blur. It may distort the temporal relation among symbol transitions, modify envelope shape, and interact with pulse overlap in ways that change the apparent local statistics of the observation window.

The low-SNR regime is commonly summarized by a scalar ratio between signal power and noise power, yet for representation learning that scalar description is incomplete. The classifier does not observe signal and noise separately. It observes a short realization whose task-relevant coordinates may have very different effective SNRs. A local phase transition may remain visible while the amplitude envelope is buried. Conversely, envelope variation may survive while phase information becomes untrustworthy. Therefore, the relevant object is not a single global SNR but an anisotropic distribution of informative energy across feature directions. One direction in input space may carry substantial class evidence, while another orthogonal direction is dominated by nuisance activity. The role of the network is to discover and amplify the former while compressing the latter.

The observation window length T imposes a fundamental identifiability constraint. In many radio

Stage	Operation	Expression	Purpose
Squeeze	Temporal averaging per channel	$q_j = \frac{1}{T}\sum_t U_{j,t}$	Summarize global channel behavior across the window
Bottleneck	Nonlinear mixing	$\mathbf{h} = \delta(\mathbf{W}_1\mathbf{q})$	Learn low-dimensional context descriptor
Gate	Channel weighting	$\mathbf{a} = \sigma(\mathbf{W}_2\mathbf{h})$	Produce sample-adaptive channel importances
Recalibration	Feature scaling	$\mathbf{U}' = \mathrm{diag}(\mathbf{a})\mathbf{U}$	Suppress nuisance-heavy or redundant channels
Covariance update	Channel-wise contraction	$\mathbf{K}' = \mathrm{diag}(\mathbf{a})\mathbf{K}\,\mathrm{diag}(\mathbf{a})$	Reduce nuisance variance while preserving useful directions

Table 5: Channel-selective gating pipeline applied to an intermediate feature tensor.

Quantity	Expression	Interpretation
Margin $m_{y,c}$	$m_{y,c}(\mathbf{z}) = (\mathbf{w}_y - \mathbf{w}_c)^\top \mathbf{z} + b_y - b_c$	Signed distance between class y and competing class c
Mean margin	$\mu_m^{(y,c)} = \mathbb{E}[m_{y,c}(\mathbf{z}) \mid y]$	Average separation for class pair (y, c)
Margin variance	$\sigma_m^{2\,(y,c)} = \mathrm{Var}[m_{y,c}(\mathbf{z}) \mid y]$	Sensitivity of the decision to nuisance perturbations
Reliability ratio	$\rho_{y,c} = \dfrac{\mu_m^{(y,c)}}{\sigma_m^{(y,c)} + \epsilon}$	Margin stability indicator under low SNR
Fisher ratio	$F_{c,c'} = \dfrac{\lVert \boldsymbol{\mu}_c - \boldsymbol{\mu}_{c'} \rVert_2^2}{\mathrm{tr}(\boldsymbol{\Sigma}_c) + \mathrm{tr}(\boldsymbol{\Sigma}_{c'}) + \epsilon}$	Discriminability between classes c and c'

Table 6: Geometric quantities describing decision margins and class manifold structure.

datasets, T is chosen so that the input captures a modest number of symbol intervals, enough to reveal local dynamics but not enough to average out all nuisance effects [9]. Short windows make the problem more realistic for responsive receivers but increase ambiguity. Two classes with distinct asymptotic spectral behavior may still look similar within a finite noisy window. In such circumstances, the classifier cannot rely only on global summary statistics. It must make use of small local regularities, their repetition pattern, and their compatibility with broader context. This multiscale structure is central to the later argument about information overload, because channels at different depths often encode different scales of evidence, and not all of that evidence should be retained with equal strength.

A deep feature extractor defines a sequence of maps

$$\mathbf{Z}^{(0)} = \mathbf{X}, \qquad \mathbf{Z}^{(\ell+1)} = \Phi_\ell\Big(\mathbf{Z}^{(\ell)}\Big),$$

for layers $\ell = 0, 1, \ldots, L-1$. Each hidden tensor $\mathbf{Z}^{(\ell)} \in \mathbb{R}^{d_\ell \times T_\ell}$ contains d_ℓ channels and temporal extent T_ℓ. In most practical networks, the maps Φ_ℓ are made of convolution, normalization, nonlinearity, and sometimes residual addition or pooling. A final pooling and classification head produces logits

$$\mathbf{g}(\mathbf{X}) \in \mathbb{R}^C,$$

and posterior probabilities

$$p_c(\mathbf{X}) = \frac{\exp(g_c(\mathbf{X}))}{\sum_{j=1}^{C} \exp(g_j(\mathbf{X}))}.$$

The conventional training objective is the empirical

Aspect	Symbol / expression	Level	Role
Sample loss	$\ell(\Theta; \mathbf{X}, y) = -\log p_y(\mathbf{X})$	Sample	Drives supervised learning for each observation
Gradient noise	$\mathbf{G} = \mathbb{E}[\nabla\ell\nabla\ell^\top] - \mathbb{E}[\nabla\ell]\mathbb{E}[\nabla\ell]^\top$	Dataset	Captures stochastic optimization variability
Gate entropy	$E(\mathbf{a}) = -\sum_j \tilde{a}_j \log \tilde{a}_j$	Layer	Measures how concentrated channel usage is
Low-SNR accuracy	$\bar{a}_{\text{low}}$ over a SNR band	SNR band	Summarizes performance in difficult regimes
Effective rank	r_{eff} of channel covariance	Layer	Indicates degree of feature congestion

Table 7: Optimization and statistical quantities relevant for selective information regulation.

Measure	Notation	Domain	Emphasis
Per-SNR accuracy	$a(\rho) = \mathbb{P}(\hat{y} = y \mid \text{SNR} = \rho)$	SNR level	Resolution of the performance curve versus SNR
Low-SNR average	$\bar{a}_{\text{low}} = \frac{1}{\lvert\mathcal{R}\rvert} \sum_{\rho\in\mathcal{R}} a(\rho)$	SNR band	Aggregate behavior in a difficult interval (e.g., $[-20, 0]$ dB)
Confusion matrix	$C_{i,j}(\rho)$	Class pair, SNR	Structure of misclassifications across classes
Posterior entropy	$H(\mathbf{p}) = -\sum_c p_c \log p_c$	Sample	Calibration and ambiguity assessment per example
Calibration error	ECE by SNR stratum	SNR band	Agreement between predicted and empirical confidence

Table 8: Evaluation measures highlighting where channel-selective regulation improves performance.

cross-entropy [10]

$$\mathcal{L}(\Theta) = -\frac{1}{N}\sum_{i=1}^{N} \log p_{y_i}(\mathbf{X}_i),$$

where Θ collects all trainable parameters.
This formulation is standard, but to understand low-SNR behavior we need a more structured view of the hidden tensors. Let a hidden representation at a fixed layer be decomposed as

$$\mathbf{Z}^{(\ell)} = \mathbf{S}^{(\ell)} + \mathbf{U}^{(\ell)} + \mathbf{R}^{(\ell)},$$

where $\mathbf{S}^{(\ell)}$ denotes class-stable signal content, $\mathbf{U}^{(\ell)}$ denotes nuisance-sensitive or unstable content, and $\mathbf{R}^{(\ell)}$ denotes redundant content that largely re-encodes information already available elsewhere in the hierarchy. This is not an observable decomposition, but it is analytically useful. The model should evolve $\mathbf{S}^{(\ell)}$ toward greater linear separability, suppress $\mathbf{U}^{(\ell)}$, and control the growth of $\mathbf{R}^{(\ell)}$. Under ordinary feedforward accumulation, however, there is no guarantee that such control emerges. A layer can increase activation energy or even apparent complexity while adding mostly unstable or redundant directions. The geometry of the class manifolds can be described through class means and covariances in hidden space. Let $\mathbf{z}^{(\ell)}$ denote a pooled vector form of $\mathbf{Z}^{(\ell)}$. Define

$$\boldsymbol{\mu}_c^{(\ell)} = \mathbb{E}\left[\mathbf{z}^{(\ell)} \mid y = c\right], \qquad \boldsymbol{\Sigma}_c^{(\ell)} = \mathrm{Cov}\left[[11]\mathbf{z}^{(\ell)} \mid y = c\right].$$

If a hidden layer is favorable for discrimination, then the class means should separate while the within-class covariance should contract along nuisance-heavy

Principle	Mechanism	Expected effect
Conditional depth	Residual blocks with gated perturbations	Deeper models that avoid uncontrolled feature accumulation
Embedded channel control	Gates placed within feature hierarchy	Early suppression of nuisance-heavy channels
Selective fusion	Skip connections combined with gating	Preservation of fragile shallow cues without congestion
Branched diversity	Grouped or parallel convolutions with regulation	Rich feature sets that remain sample-efficient
Conditional invariance	Context-dependent channel weighting	Robustness to nuisance while preserving class cues
Efficiency focus	Wide architectures with selective activation	Improved low-SNR robustness at modest cost

Table 9: Architectural principles motivated by channel-selective information regulation.

directions. This is captured loosely by a generalized margin score

$$\Gamma^{(\ell)} = \frac{1}{C(C-1)} \sum_{c \neq c'} (\boldsymbol{\mu}_c^{(\ell)} - \boldsymbol{\mu}_{c'}^{(\ell)})^\top \times \left(\frac{\boldsymbol{\Sigma}_c^{(\ell)} + \boldsymbol{\Sigma}_{c'}^{(\ell)}}{2} \right)^{-1} (\boldsymbol{\mu}_c^{(\ell)} - \boldsymbol{\mu}_{c'}^{(\ell)}). \tag{2}$$

Low-SNR observations tend to reduce $\Gamma^{(\ell)}$ in early layers because added disturbance inflates covariance and blurs mean separation. A well-designed network must then reconstruct an increase in $\Gamma^{(\ell)}$ deeper in the hierarchy. Yet if the deeper representation becomes overloaded by unstable channels, $\Gamma^{(\ell)}$ may fail to recover despite substantial model capacity.

The complex nature of the input also induces algebraic structure that a network should ideally respect. A phase rotation by angle φ acts on the two real channels through

$$\mathbf{x}'_t = \mathbf{R}(\varphi)\mathbf{x}_t, \qquad \mathbf{R}(\varphi) = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}.$$

Thus, physically equivalent observations may trace curved or rotated trajectories in the input space. A feature extractor need not be strictly invariant to all such transformations, since some phase structure remains class informative, but it must learn coordinate systems in which nuisance rotations do not dominate the classification decision. The difficulty is that local convolutions alone can only approximate this behavior implicitly. Different channels may each respond to a different rotated projection, which again raises the possibility of feature proliferation without selective governance [12]. A channel-weighting mechanism can be understood as deciding which among these rotated or transformed projections are coherent with the sample-level evidence.

Temporal structure creates an analogous issue. Some class information depends on local transitions occurring over only a few samples. Other information depends on consistency across longer portions of the sequence. A hidden tensor must therefore retain access to several temporal scales simultaneously. If too much downsampling or abstraction occurs early, weak local evidence disappears. If too many local channels are propagated without compression, the later representation becomes congested. This tradeoff motivates architectures that combine residual transport, limited cross-depth access, and selective channel scaling. Even when the architecture is not described that way in implementation terms, this is its functional role from the viewpoint of information flow.

To formalize channel importance, consider a pooled feature vector $\mathbf{u} \in \mathbb{R}^d$ produced from some hidden tensor. The logit for class c may be written as

$$g_c(\mathbf{u}) = \mathbf{w}_c^\top \mathbf{u} + b_c.$$

For a particular sample, the contribution of channel j to the margin between true class y and competing class c is

$$m_{y,c}^{(j)} = (w_{y,j} - w_{c,j})u_j.$$

Channels with large and stable positive contributions across nuisance conditions are useful; channels whose contributions are volatile or near-zero are inefficient or harmful [13]. However, this scalar view still ignores interactions among channels. In practice, groups of channels may jointly encode the same signal attribute. The network must therefore regulate not only individual channels but the effective subspaces they span. This motivates the later use of covariance and spectral measures to describe overload.

Another key observation is that the empirical risk experienced during training is unevenly informative across examples. High-SNR examples often generate gradients pointing in broadly consistent directions because class-discriminative patterns are strong. Low-SNR examples generate gradients that are smaller, noisier, and more sensitive to accidental channel activations. Without selective mechanisms, training tends to favor channels helpful for easier examples while leaving difficult ones underrepresented. Therefore, the signal model and the learning model cannot be separated. The shape of the data distribution directly affects which channels survive training.

In summary, the received waveform should be viewed as a structured random object whose class identity is embedded in an anisotropic and nuisance-corrupted observation space. A deep classifier maps this object into successive hidden tensors that may accumulate stable evidence, unstable evidence, and redundancy at different rates [14]. The central problem for low-SNR recognition is to control this accumulation. The next section introduces a formal account of information overload in multichannel feature systems and explains why a channel-selective compression mechanism can be interpreted as a remedy for congestion rather than as a simple attention heuristic.

3 | Information Overload and Channel-Selective Compression

The expression information overload is often used informally to suggest that a model receives more data than it can handle. In the context of low-SNR modulation recognition, the phenomenon is more precise. Overload occurs when the dimensional growth or persistence of hidden channels exceeds the growth of class-relevant information. The network then carries forward a feature tensor whose energy and apparent complexity are high, but whose proportion of discriminative content is low. This is not identical to overfitting, though the two are related. A model can be overloaded even before it memorizes the training set, simply because too many channels respond to nuisance or redundancy relative to the amount of stable class structure available in a short noisy observation.

To formalize the idea, let $\mathbf{Z} \in \mathbb{R}^{d\times T}$ be a hidden tensor. Define its channel covariance matrix

$$\mathbf{K} = \frac{1}{T}\left(\mathbf{Z} - \bar{\mathbf{Z}}\right)\left(\mathbf{Z} - \bar{\mathbf{Z}}\right)^{\top},$$

where $\bar{\mathbf{Z}}$ repeats the channel means across time. The eigenvalues of $\mathbf{K}$ indicate how activation energy is distributed among channel directions. If a few eigenvalues dominate while many others remain small but noisy, the representation may contain a narrow effective signal subspace contaminated by many low-utility directions. The raw dimensionality d is then misleading because the effective rank relevant to class discrimination is much smaller. Define the participation ratio [15]

$$r_{\text{eff}} = \frac{(\operatorname{tr}\mathbf{K})^2}{\operatorname{tr}\left(\mathbf{K}^2\right)}.$$

This is a soft rank measure. When r_{eff} is much smaller than d, the representation uses only a limited number of strong directions. That fact alone is not problematic. The problem arises when the remaining directions are not zero but are populated by unstable or redundant responses that later layers must still process. Such directions burden the classifier without improving separability.

A more task-oriented measure compares between-class and within-class covariance. Let $\mathbf{K}_W$ denote within-class covariance of pooled features and $\mathbf{K}_B$ denote between-class covariance. Then a discriminative efficiency score may be defined as

$$\Xi = \frac{\operatorname{tr}(\mathbf{K}_B)}{\operatorname{tr}(\mathbf{K}_W) + \epsilon},$$

for small $\epsilon > 0$. A channel-rich tensor can have large total covariance yet poor Ξ if most energy lies in within-class variation. In low-SNR settings, nuisance-driven channels often raise $\operatorname{tr}(\mathbf{K}_W)$ more than $\operatorname{tr}(\mathbf{K}_B)$, thereby reducing discriminative efficiency. Overload can thus be interpreted as a misallocation of covariance budget. The network spends representational variance on directions that do not improve class separation.

One source of overload is branch proliferation without selective integration. Modern convolutional designs often employ grouped or parallel transforms because they increase representational diversity [16]. This is beneficial, but each branch may capture overlapping or unstable versions of the same local phenomenon. If all branch outputs are propagated equally, the later tensor contains duplicated evidence and branch-specific nuisance artifacts. A second source is the repeated preservation of shallow features through skip transport without sufficient filtering. While residual transport improves gradient flow and prevents catastrophic loss of detail, it can also allow weakly useful activations to persist far beyond the layer where they matter. A third source is scale mismatch. Features extracted at different temporal scales may be concatenated or added, but the resulting tensor may not be normalized around task relevance. Some channels can dominate solely because they have larger raw activation norms, not because they are more informative.

Channel-selective compression addresses these problems by converting undifferentiated channel persistence into data-dependent scaling. Let $\mathbf{U} \in \mathbb{R}^{d \times T}$ be an intermediate tensor. A global summary vector is formed by temporal squeezing,

$$q_j = \frac{1}{T} \sum_{t=1}^{T} U_{j,t}, \qquad \mathbf{q} \in \mathbb{R}^d.$$

The summary is passed through a gating map [17]

$$\mathbf{a} = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{q})), \tag{3}$$

where δ is a hidden nonlinearity, σ is a squashing function such as sigmoid, and $\mathbf{a} \in (0,1)^d$ gives channel weights. The recalibrated tensor is

$$\mathbf{U}' = \operatorname{diag}(\mathbf{a})\mathbf{U}.$$

This operator has three simultaneous effects. It compresses channels that are globally weak or inconsistent, it preserves or amplifies channels aligned with the sample summary, and it makes the later representation depend on a lower-dimensional global descriptor rather than on raw local activity alone.

From a covariance perspective, the transformation changes the channel covariance to

$$\mathbf{K}' = \operatorname{diag}(\mathbf{a})\mathbf{K} \operatorname{diag}(\mathbf{a}).$$

If nuisance-heavy channels systematically receive low weights, then their contribution to both variance and cross-covariance contracts. This reduces the burden on later layers, which no longer need to disentangle strong nuisance directions from weak class directions. In principle, the gate can also enhance harmful channels if poorly trained. But with suitable optimization and enough examples, the gating network tends to align channel energy with loss reduction because its gradients are driven by downstream classification success. The result is a feedback process in which channels survive in proportion to their usefulness.

Compression is not a synonym for information loss. The key distinction is between raw channel count and effective information quality. Suppose a class-discriminative signal occupies a low-dimensional subspace $\mathcal{S}$ while nuisance occupies $\mathcal{N}$. If the hidden tensor expands both indiscriminately, the downstream classifier sees a mixed space $\mathcal{S} \oplus \mathcal{N}$. Channel-selective compression aims to reduce the projection onto $\mathcal{N}$ while preserving or modestly enhancing the projection onto $\mathcal{S}$. In this sense, compression can increase usable information even while decreasing raw activation entropy [18]. The network becomes less descriptive of every fluctuation and more descriptive of the fluctuations that matter.

To express this more formally, let the pooled hidden vector satisfy

$$\mathbf{u} = \mathbf{P}_S \mathbf{u} + \mathbf{P}_N \mathbf{u},$$

where $\mathbf{P}_S$ and $\mathbf{P}_N$ project onto class-relevant and nuisance subspaces, respectively. After channel gating,

$$\mathbf{u}' = \mathbf{D}\mathbf{u}, \qquad \mathbf{D} = \operatorname{diag}(\mathbf{a}).$$

The relevant gain is then

$$\Lambda = \frac{\|\mathbf{P}_S \mathbf{u}'\|_2^2}{\|\mathbf{P}_N \mathbf{u}'\|_2^2 + \epsilon} = \frac{\|\mathbf{P}_S \mathbf{D}\mathbf{u}\|_2^2}{\|\mathbf{P}_N \mathbf{D}\mathbf{u}\|_2^2 + \epsilon}. \tag{4}$$

A good gate increases Λ even when the total norm $\|\mathbf{u}'\|_2$ decreases. This is exactly why low-SNR gains can appear without large changes in global accuracy. High-SNR examples may already have high Λ, so gating changes little. Low-SNR examples may begin with low Λ, and even a moderate increase can significantly improve classification.

The dependence of the gate on global context is especially important. Consider two local channels that show similar activation magnitude at a given time location, but one is consistent with a constant-envelope digital signal and the other with a distorted analog mode. Local magnitude alone cannot distinguish them. The global squeeze vector $\mathbf{q}$ summarizes the broader configuration of the sample and thereby makes the gate conditional on the full observation. Hence channel weighting is not a static saliency measure; it is a sample-adaptive prior over which channels are plausible carriers of class evidence in the current context [19].

There is also a temporal interpretation. The squeeze operation averages across time, which at first seems risky because averaging might erase transient details. However, the gate does not replace the local tensor; it only modulates it. The local detail remains in $\mathbf{U}$. The global summary simply determines the strength with which each channel should be trusted. When the sample is noisy, a time-average of channel behavior can be more reliable than any single local activation. Thus, the gate leverages one of the few robust statistics still available under low SNR: the global consistency of a channel's response over the observation window.

The compression mechanism interacts naturally with residual transport. In a residual block,

$$\mathbf{Z}_{\text{out}} = \mathbf{Z}_{\text{in}} + \mathcal{F}(\mathbf{Z}_{\text{in}}).$$

If channel scaling is applied to $\mathcal{F}(\mathbf{Z}_{\text{in}})$, then the residual perturbation itself becomes selective:

$$\mathbf{Z}_{\text{out}} = \mathbf{Z}_{\text{in}} + \mathbf{D}\,\mathcal{F}(\mathbf{Z}_{\text{in}}).$$

This is an elegant arrangement because the identity pathway preserves the baseline information transport, while the nonlinear perturbation is admitted only in channels judged useful for the current sample. Such a block resists overload in two ways: it keeps the stable transport route intact and constrains the volume of new feature content injected at each stage.

A related benefit concerns classifier conditioning. Suppose the final classifier operates on pooled features $\mathbf{z}$. Without channel regulation, several nuisance-heavy channels may contribute small but erratic terms to the logits [20]. Their aggregate effect can destabilize the margin. With sample-dependent scaling, the effective class weights become $\mathbf{D}\mathbf{w}_c$, and the classifier adapts its emphasis to the current observation. This reduces the risk that a nuisance channel with a large static classifier weight dominates the decision under unusual noise realizations.

It is useful to distinguish channel-selective compression from generic sparsity. Sparsity encourages many activations to be zero or small in a model-wide sense. Channel-selective compression is more structured and more adaptive. The same channel may be amplified on one sample and attenuated on another. The objective is not to keep the representation uniformly sparse, but to keep it aligned with current task relevance. This adaptivity is especially important in modulation recognition because different classes rely on different physical attributes. A model that uses the same small subset of channels for every sample would likely underperform across diverse modulation families.

One may ask whether the gate itself could become overloaded, since it must map a high-dimensional summary into channel weights. In practice, the squeeze vector is already a compressed statistic, and the gating multilayer operator is typically low-dimensional relative to the main network. This bottleneck is advantageous [21]. It forces the gate to infer broad patterns of channel utility rather than memorizing fine local details. In effect, the gate becomes a coarse supervisor over the feature tensor, not a second full classifier.

The central theoretical claim of this section is therefore that information overload in low-SNR modulation recognition is a structured imbalance between feature dimensionality and feature utility. Channel-selective compression corrects that imbalance by reweighting hidden channels according to global evidence, thereby increasing discriminative efficiency, reducing nuisance covariance, and stabilizing the propagation of useful information. The next section examines how these operations alter the geometry of class manifolds and the evolution of decision margins in latent space.

4 | Feature Geometry and Decision Margins

The success of a modulation classifier can be interpreted geometrically. Each class induces a cloud or manifold of points in latent space, and classification is correct when the decision boundary passes between these manifolds with adequate margin. Low-SNR conditions complicate this picture because additive and multiplicative nuisances inflate each class cloud and introduce anisotropy. Some directions in latent space become highly variable while others remain relatively stable. The question is therefore not simply how far apart the class means are, but how separation compares with the orientation and spread of the within-class covariance ellipsoids. Channel-selective regulation influences this geometry by contracting unstable directions and enhancing directions whose sample-wise consistency aligns with class identity.

Let $\mathbf{z} \in \mathbb{R}^d$ denote a pooled latent vector after some stage of the network. The class posterior is determined by logits [22]

$$g_c(\mathbf{z}) = \mathbf{w}_c^\top \mathbf{z} + b_c.$$

For a sample from class y, the margin against class c is

$$m_{y,c}(\mathbf{z}) = (\mathbf{w}_y - \mathbf{w}_c)^\top \mathbf{z} + (b_y - b_c).$$

A correct decision requires $m_{y,c}(\mathbf{z}) > 0$ for all $c \neq y$. If $\mathbf{z}$ is decomposed into stable, unstable, and redundant parts,

$$\mathbf{z} = \mathbf{z}_S + \mathbf{z}_U + \mathbf{z}_R,$$

then the margin becomes

$$\begin{aligned} m_{y,c}(\mathbf{z}) = {} & (\mathbf{w}_y - \mathbf{w}_c)^\top \mathbf{z}_S \\ & + (\mathbf{w}_y - \mathbf{w}_c)^\top \mathbf{z}_U \\ & + (\mathbf{w}_y - \mathbf{w}_c)^\top \mathbf{z}_R + (b_y - b_c). \end{aligned} \tag{5}$$

The first term is the desired contribution. The second term fluctuates under nuisance perturbations. The third term is often near-zero in expectation yet can still affect variance and conditioning. Channel-selective compression acts primarily by reducing the second and third contributions relative to the first. Even if it does not enlarge the mean stable contribution dramatically, it can increase reliability by lowering margin variance. This can be quantified through conditional margin moments. Let

$$\mu_m^{(y,c)} = \mathbb{E}[m_{y,c}(\mathbf{z}) \mid y], \qquad \sigma_m^{2\,(y,c)} = \mathrm{Var}[m_{y,c}(\mathbf{z}) \mid y].$$

A robust classifier should increase the ratio

$$\rho_{y,c} = \frac{\mu_m^{(y,c)}}{\sigma_m^{(y,c)} + \epsilon},$$

since a larger ratio indicates not just larger average margin but more reliable separation [23]. In low-SNR data, many models fail because $\sigma_m^{2\,(y,c)}$ becomes large. Channel regulation is particularly effective in such cases because it can reduce variance even when the mean margin changes only modestly. This explains why low-SNR improvements can be pronounced without equally large gains at high SNR.

Class manifolds are also shaped by cross-depth transport. Shallow features often encode fine local detail such as short phase transitions, envelope perturbations, or pulse overlap signatures. Deep features encode broader contextual consistency. If only deep features survive, the manifold can become smoother but also more overlapped for classes that differ mainly in delicate local attributes. If only shallow features survive, the manifold can become fragmented and overly sensitive. A practical architecture therefore combines selective transport of shallow information with deeper contextual abstraction. Geometrically, this means that the manifold is represented in a coordinate system containing both coarse and fine axes, with channel gates deciding which axes deserve prominence for the current sample.

Suppose two representations $\mathbf{z}_s$ and $\mathbf{z}_d$ are derived from shallow and deep stages. A fused vector is

$$\mathbf{z}_f = \mathbf{A}\mathbf{z}_s + \mathbf{B}\mathbf{z}_d.$$

After gating,

$$\mathbf{z}' = \mathbf{D}(\mathbf{z}_f)\mathbf{z}_f.$$

The manifold geometry now depends on both the fusion matrices and the diagonal operator $\mathbf{D}$. Even if the shallow and deep manifolds individually overlap for certain class pairs, their combination may produce a more separable configuration because the gate suppresses the coordinates in which overlap remains largest [24]. Thus, fusion and gating are not merely additive conveniences. Their interaction creates a sample-adaptive latent chart.

The covariance structure after gating is particularly revealing. Let the pooled latent covariance for class c be $\mathbf{\Sigma}_c$. After a deterministic diagonal scaling $\mathbf{D}$,

$$\mathbf{\Sigma}'_c = \mathbf{D}\mathbf{\Sigma}_c\mathbf{D}.$$

If $\mathbf{D}$ is sample-dependent, then the actual class covariance is a mixture over gates, but the same intuition applies: unstable channels shrink. This shrinkage is not isotropic. It is concentrated on selected axes. Consequently, the covariance ellipsoid becomes more compact in nuisance-dominated directions while preserving spread in class-relevant directions where some variability may still be needed to distinguish among related modes. This anisotropic contraction is preferable to uniform norm regularization, which may erase useful variation together with harmful variation.

One can interpret the gate as introducing a Riemannian-like metric on latent space. Without gating, distance between latent vectors is measured in the Euclidean geometry inherited from the feature coordinates. With a diagonal positive scaling $\mathbf{D}$, the local metric becomes

$$\mathbf{M} = \mathbf{D}^\top\mathbf{D}.$$

Then the squared distance between nearby vectors $\mathbf{z}_1$ and $\mathbf{z}_2$ is

$$d^2_{\mathbf{M}}(\mathbf{z}_1, \mathbf{z}_2) = (\mathbf{z}_1 - \mathbf{z}_2)^\top\mathbf{M}(\mathbf{z}_1 - \mathbf{z}_2).$$

Coordinates corresponding to small gate values contribute less to distance; those with large gate values contribute more. In this sense, channel-selective regulation changes the effective geometry in which the classifier compares examples [25]. It does not just rescale activations. It deforms the latent space so that motion along some directions becomes inexpensive and motion along other directions becomes expensive. A well-trained gate makes nuisance motion cheap and class-relevant motion expensive, which benefits margin preservation.

This geometric view clarifies why low-SNR recognition is sensitive to feature congestion. In an overloaded representation, many nuisance axes remain expensive because they carry moderate activation and nontrivial classifier weights. Small perturbations along these axes can then cause large changes in distance or margin. The latent geometry becomes rough. Selective gating smooths that geometry by flattening low-utility directions. The manifold of each class becomes less jagged and more coherent under nuisance transformations.

A complementary viewpoint comes from singular value analysis of class-centered feature matrices. Let $\mathbf{Z}_c$ be the matrix whose rows are pooled latent vectors from class c, centered by subtracting the class mean. If the singular values are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d$, then a highly overloaded class representation often exhibits a long noisy tail of small singular values. These correspond to directions along which the class cloud is thin but unstable [26]. They do not help separate the class from others yet can interfere with optimization and classification. Channel-selective compression tends to steepen the spectral decay by damping channels contributing mostly to that tail. The representation becomes effectively lower-dimensional without losing discriminative core structure.

Decision geometry also depends on interclass asymmetry. Some classes are intrinsically more fragile because their useful features are weak under low SNR. For such classes, overload is especially damaging because nuisance channels can dominate the few class-specific channels that remain. Let the pairwise Fisher-like ratio between classes c and c' be

$$F_{c,c'} = \frac{\|\boldsymbol{\mu}_c - \boldsymbol{\mu}_{c'}\|_2^2}{\mathrm{tr}(\boldsymbol{\Sigma}_c) + \mathrm{tr}(\boldsymbol{\Sigma}_{c'}) + \epsilon}. \tag{6}$$

Selective gating may increase $F_{c,c'}$ primarily by reducing the denominator. This is critical because some difficult class pairs may have only limited room for mean separation due to genuine signal similarity under short windows. In those cases, the practical way to improve classification is to contract nuisance variance rather than to expect much larger between-class displacement.

There is also an adversarial interpretation. Consider a small perturbation δ in input space aligned with nuisance directions. The change in the margin is approximately [27]

$$m_{y,c}(\mathbf{X} + \delta) - m_{y,c}(\mathbf{X}) \approx \nabla_{\mathbf{X}} m_{y,c}(\mathbf{X})^\top \delta.$$

If the feature hierarchy contains many nuisance-sensitive channels, the gradient norm with respect to such perturbations is large. Channel regulation can reduce this by attenuating channels whose activation patterns contribute disproportionately to nuisance sensitivity. The classifier thereby becomes less fragile not only to random noise but to structured mismatch such as phase drift or mild spectral distortion.

A final geometric point concerns calibration. Posterior probabilities are often unreliable in low-SNR regimes because the latent space may contain scattered regions where the classifier is forced to extrapolate from unstable combinations of channels. If gating reduces the occupation of these scattered regions, then the posterior entropy for recoverable samples can decrease while the entropy for truly ambiguous samples remains appropriately high. In other words, the latent geometry becomes more faithful to actual uncertainty. The classifier is not simply becoming more confident; it is becoming more selective about when confidence is justified.

The feature-geometric analysis therefore supports a broad conclusion. Channel-selective regulation and controlled transport of multiscale features reshape the latent space by contracting nuisance directions, preserving fine but useful structure, and stabilizing class margins. These effects matter most where the observation is weak and ambiguous [28]. The next section turns from geometry to dynamics and studies how these architectural mechanisms change optimization behavior and statistical efficiency during training.

5 | Optimization Dynamics and Statistical Efficiency

Training a deep modulation classifier is a stochastic optimization problem in which not all samples contribute equally informative gradients. High-SNR examples tend to produce confident predictions early, and the corresponding gradients often reinforce features that are broadly useful across the dataset. Low-SNR examples, by contrast, may generate uncertain posteriors, unstable feature activations, and gradients that vary substantially from one mini-batch to another. If the architecture does not preserve weak discriminative cues and does not regulate nuisance-heavy channels, the optimization process naturally gravitates toward features favored by the easier examples. This produces a model that appears successful globally yet remains fragile in the regime where robustness is most needed.

Let the sample-wise loss be

$$\ell(\Theta; \mathbf{X}, y) = -\log p_y(\mathbf{X}).$$

The gradient covariance governing stochastic optimization noise is

$$\mathbf{G} = \mathbb{E}\left[\nabla_\Theta \ell \nabla_\Theta \ell^\top\right] - \mathbb{E}[\nabla_\Theta \ell]\,\mathbb{E}[\nabla_\Theta \ell]^\top. \tag{7}$$

When low-SNR samples induce highly variable gradients, the dominant eigendirections of $\mathbf{G}$ may correspond to nuisance fluctuations rather than coherent class structure. Optimization then wastes updates on unstable parameter directions.
Channel-selective compression reduces this effect because it suppresses feature channels that repeatedly fail to contribute stable loss reduction. In other words, gating changes the mapping from sample noise to gradient noise.
To see this, consider a simplified final-stage model in which the pooled feature vector $\mathbf{u}$ is gated before classification:

$$\mathbf{z} = \mathbf{D}(\mathbf{u})\mathbf{u}, \qquad [29] g_c = \mathbf{w}_c^\top \mathbf{z} + b_c.$$

The gradient of the loss with respect to $\mathbf{u}$ contains two terms:

$$\frac{\partial \ell}{\partial \mathbf{u}} = \mathbf{D}^\top \frac{\partial \ell}{\partial \mathbf{z}} + \left(\frac{\partial \mathbf{D}}{\partial \mathbf{u}}\mathbf{u}\right)^\top \frac{\partial \ell}{\partial \mathbf{z}}. \tag{8}$$

The first term rescales the backpropagated error channel-wise. The second term accounts for the dependence of the gate on the feature summary. Together they implement a gradient filter. Channels with small gate values receive weaker direct gradients, and channels whose global context does not justify emphasis also receive reduced gate-induced correction. This does not freeze them permanently, because the gate is sample-adaptive. But it does reduce the influence of persistently low-utility channels on optimization.
Residual transport contributes a different kind of optimization stability. In a residual block,

$$\mathbf{Z}_{\ell+1} = \mathbf{Z}_\ell + \mathbf{D}_\ell \mathcal{F}_\ell(\mathbf{Z}_\ell),$$

the local Jacobian is

$$\mathbf{J}_\ell = \mathbf{I} + \frac{\partial}{\partial \mathbf{Z}_\ell}\left(\mathbf{D}_\ell \mathcal{F}_\ell(\mathbf{Z}_\ell)\right).$$

The identity component keeps the singular values of $\mathbf{J}_\ell$ from drifting too far from unity if the residual perturbation is reasonably controlled. Channel gating further moderates the perturbation term by reducing the scale of unhelpful channels [30]. This improves gradient transport across depth. Such improvement is especially relevant for low-SNR cues stored in earlier layers. Without reliable gradient flow, the network may fail to learn filters that preserve these cues because their contribution to the final loss is weak and delayed.

Statistical efficiency can be discussed through a bias-variance lens. Let the estimator of the posterior function be $\widehat{f}$. A model with insufficient capacity has high bias, while an excessively flexible model without appropriate inductive structure has high variance. In low-SNR modulation recognition, variance often manifests not only as sensitivity to training data but also as sensitivity to nuisance realizations. A dense network with many unconstrained channels may fit the training set well while relying on channel configurations that are unstable under slight distortions. Channel-selective bottlenecks act as a form of adaptive regularization. They reduce variance in feature space by discouraging reliance on channels that do not repeatedly demonstrate task value across the training distribution.
A formal illustration can be obtained by considering a simplified linearized feature model

$$\mathbf{u} = \mathbf{A}\mathbf{x} + \boldsymbol{\epsilon},$$

where $\mathbf{x}$ is a latent class signal, $\mathbf{A}$ is a learned feature map, and $\boldsymbol{\epsilon}$ represents nuisance-induced activation noise. If the classifier uses $\mathbf{u}$ directly, the effective noise covariance is $\boldsymbol{\Sigma}_\epsilon$. Under diagonal gating $\mathbf{D}$, the classifier uses

$$\mathbf{u}' = \mathbf{D}\mathbf{u} = \mathbf{D}\mathbf{A}\mathbf{x} + \mathbf{D}\boldsymbol{\epsilon},$$

and the noise covariance becomes

$$\boldsymbol{\Sigma}'_\epsilon = \mathbf{D}\boldsymbol{\Sigma}_\epsilon\mathbf{D}.$$

Even without changing the signal map $\mathbf{A}$, the gate lowers estimator variance along selected directions. Since classification error in noisy regimes is often variance-dominated, this can yield outsized benefits [31].
Another important issue is curriculum imbalance. In typical training, examples are sampled roughly uniformly from the dataset, yet the optimization does not experience them uniformly. Easy examples are fit sooner and can dominate the learned representation. Hard examples often begin to shape the model only after the easy decision boundary is already formed. If the architecture at that stage lacks preserved low-level detail or retains too much nuisance congestion, the gradients from hard examples may not have enough leverage to correct the representation. Feature transport and gating change that leverage. Transport keeps weak cues available. Gating reduces the background clutter against which those cues must compete. As a result, gradients from hard examples can act on a cleaner and more responsive latent substrate.

This suggests a useful decomposition of training progress across SNR strata. Let $\mathcal{L}_\rho(\Theta)$ denote the expected loss at SNR level ρ. Then the total loss is [32]

$$\mathcal{L}(\Theta) = \sum_\rho \pi_\rho \mathcal{L}_\rho(\Theta),$$

where π_ρ is the empirical prevalence of that stratum. During training, the derivative of $\mathcal{L}_\rho$ with respect to time or iteration may behave differently across strata. For many architectures, $\mathcal{L}_\rho$ drops rapidly for high ρ and slowly or erratically for low ρ. A network with selective information regulation may show a different pattern: after an initial period of broad feature formation, the low-SNR losses continue to decrease more steadily because the representation becomes progressively less overloaded. This interpretation is consistent with the idea that architecture shapes not only the endpoint of optimization but the order in which the problem becomes learnable.

Generalization under distribution shift can also be framed dynamically. Suppose the test environment introduces a nuisance transformation $\mathcal{T}$ not fully represented in training, such as slightly different phase noise statistics or amplitude scaling. A classifier that relies on overloaded channels will often exhibit a large derivative with respect to $\mathcal{T}$, because many channels respond in loosely constrained ways. A classifier trained with channel-selective regulation tends to rely on a smaller set of more globally coherent channels, making the derivative with respect to $\mathcal{T}$ smaller on average. This is not a guarantee of invariance, but it is a better starting point for graceful degradation.

There is also a computational-statistical tradeoff. Increasing width or branch count can reduce bias by providing more representational choices, but beyond a point it inflates optimization noise and the risk of nuisance fitting. Selective gating improves the tradeoff because it allows the network to instantiate many channels while using only a context-appropriate subset strongly for any given sample. In effect, the model can be globally rich but locally disciplined. This differs from permanently pruning channels, which may remove capacity needed for some classes. Local discipline is preferable because the relevant channels depend on the modulation family and on the realized distortion.

A final optimization consideration concerns calibration of the gate itself [33]. The gate is trained through the same classification objective as the rest of the model, which means it can in principle overreact and suppress channels prematurely. However, residual transport mitigates that risk. Since the identity path preserves baseline information flow, aggressive suppression of the residual perturbation does not fully erase the signal. Instead, the model can learn gradually which channels deserve amplification. This makes the combined system less brittle than a purely multiplicative architecture lacking skip transport.

The dynamic picture that emerges is therefore one in which selective information regulation improves learning not by simplifying the task externally, but by improving the internal economics of optimization. Useful gradients are transmitted more effectively, nuisance-induced gradient variance is reduced, weak cues from hard examples remain actionable, and the model acquires a better bias-variance balance under finite data. These effects help explain why channel reweighting may have limited visible impact on easy examples yet substantial impact where the signal is weak and confusion is high.

6 | Experimental Framework and Interpretive Analysis

A rigorous empirical evaluation of low-SNR modulation recognition should do more than compare a few top-line accuracy numbers. Because the central phenomenon is selective improvement in difficult regimes, the experimental framework must resolve where in the task distribution the architecture changes behavior. This requires stratification by SNR, analysis of class-pair confusions, examination of latent feature compactness, and explicit attention to efficiency. The value of channel-selective information regulation is not fully visible if evaluation collapses all these dimensions into a single aggregate metric [34].

The standard benchmark format for this problem uses complex-baseband observations represented as two real channels with fixed sample length. A dataset spanning multiple digital and analog classes across a wide SNR range is desirable because it forces the model to handle both discrete transition patterns and broader waveform statistics. The benchmark should also include realistic nuisance effects such as frequency shift, fading, and additive white noise so that robustness is tested under nontrivial distortions rather than idealized corruption alone. A train-test split should preserve the distribution of SNR levels and classes or explicitly stratify them. Otherwise, reported gains may reflect accidental sampling differences rather than genuine robustness.

Within such a framework, the first indispensable metric is per-SNR accuracy:

$$a(\rho) = \mathbb{P}(\widehat{y} = y \mid \mathrm{SNR} = \rho).$$

The entire function $a(\rho)$ matters, not merely its average. High-SNR performance can saturate for many models, so improvements there often reveal little about architecture quality. The key region is the transition band where the task moves from near-chance performance toward reliable discrimination. If channel-selective regulation is effective, the curve should rise earlier or more steeply in this band. A model that improves mainly in already easy SNR regions is likely benefiting from generic capacity rather than from better information control [35].

A second metric is low-SNR block average,

$$\bar{a}_{\text{low}} = \frac{1}{|\mathcal{R}|} \sum_{\rho \in \mathcal{R}} a(\rho),$$

where $\mathcal{R}$ may be chosen as a difficult interval such as $[-20, 0]$ dB. This metric directly reflects the regime in which information overload is most damaging. It also aligns with the observation that a gating mechanism can leave global accuracy nearly unchanged while substantially improving $\bar{a}_{\text{low}}$. Such behavior is informative rather than anomalous. It means the model is solving the part of the problem where representation quality matters most.

Confusion matrices are equally important because low-SNR errors are not uniformly distributed across classes. Let

$$C_{i,j}(\rho) = \mathbb{P}(\hat{y} = j \mid y = i, \text{SNR} = \rho).$$

Improvement should be interpreted in terms of movement of off-diagonal mass. If the architecture reduces confusion between classes that are known to overlap strongly under noise while preserving performance on easier classes, then the gain is structurally meaningful. By contrast, a model might improve average accuracy through diffuse marginal changes that do not resolve persistent failure modes. Channel-selective regulation predicts more targeted movement. Since it acts by suppressing nuisance-heavy channels and preserving sample-consistent ones, its benefits should appear disproportionately in class pairs whose distinctions depend on weak but stable cues.

A third perspective uses posterior entropy. For a test sample, [36]

$$H(\mathbf{p}) = -\sum_{c=1}^{C} p_c \log p_c.$$

At low SNR, entropy naturally rises, but the useful question is whether entropy is high for the right reasons. If a sample is truly ambiguous under the observation window, high entropy is appropriate. If the sample is recoverable but the model is overloaded, entropy will also be high, but for the wrong reason. An effective channel-selective system should reduce entropy selectively on recoverable low-SNR samples while leaving intrinsically ambiguous ones uncertain. Calibration plots and expected calibration error measured separately by SNR strata can expose this distinction.

The experimental design should include ablations that separate the contributions of feature transport, grouped or residual extraction, and channel gating. A baseline deep convolutional model provides a reference point. Adding residual or grouped transport without gating reveals the effect of improved feature extraction and optimization stability. Adding channel gating on top of that reveals whether selective compression provides an independent benefit. A final ablation may weaken or remove the global squeeze summary while keeping multiplicative scaling, testing whether global context rather than local activation magnitude is the main source of utility. Interpretation of these ablations should focus on low-SNR accuracy, confusion reduction, and feature compactness, not only on aggregate test accuracy [37].

Latent-space diagnostics are particularly useful for supporting the theoretical claims of earlier sections. One diagnostic is the class-wise effective rank of pooled features, measured through the participation ratio or singular value decay. If channel gating is mitigating overload, then the latent representations should exhibit fewer low-utility tail directions while maintaining or improving class separability. Another diagnostic is a class-conditional compactness score,

$$\kappa_c = \frac{1}{n_c} \sum_{i: y_i = c} \|\mathbf{z}_i - \boldsymbol{\mu}_c\|_2^2.$$

Selective gating is expected to lower κ_c for classes heavily affected by nuisance congestion. A complementary between-class measure

$$\delta_{c,c'} = \|\boldsymbol{\mu}_c - \boldsymbol{\mu}_{c'}\|_2$$

can then be used to compute compactness-to-separation ratios. If κ_c decreases without proportional collapse of $\delta_{c,c'}$, the geometry has improved in the desired direction.

Feature-map statistics within the network can also reveal whether overload is being controlled. For example, one may monitor the distribution of channel variances before and after gating. A useful model often shows that many channels remain available at the architectural level, but only a subset are strongly active for any given sample. This sample-adaptive specialization differs from static pruning [38].

Similarly, one may inspect the entropy of the gate vector

$$E(\mathbf{a}) = -\sum_{j=1}^{d} \tilde{a}_j \log \tilde{a}_j, \qquad \tilde{a}_j = \frac{a_j}{\sum_k a_k}.$$

Low entropy suggests concentrated channel emphasis; high entropy suggests diffuse usage. Neither extreme is universally best. The desirable pattern is class- and SNR-dependent: difficult low-SNR samples may require more selective gating than clearer ones. Such variation would support the hypothesis that the gate adapts to information quality rather than imposing a fixed sparsity pattern.

Runtime and complexity should be reported alongside accuracy because selective information regulation is valuable partly because it can increase usable capacity without proportionally increasing computational cost. If a model with channel gating and structured transport achieves strong low-SNR gains with only modest overhead relative to a plain baseline, then the gains are likely due to better representation organization rather than brute-force scale. Metrics such as parameter count, multiply-accumulate operations, and latency per sample help establish this point. This is especially important for embedded or near-real-time receivers where computational budget is constrained.

Interpretation should also consider the shape of error reduction over training epochs [39]. A model that controls overload may not show its full advantage immediately. Early in training, all models may learn coarse features sufficient for high-SNR discrimination. The difference may emerge later, when hard low-SNR examples begin to influence the decision boundary. Tracking per-SNR validation curves can reveal whether the selective model continues to improve in difficult regimes after the baseline saturates. Such temporal behavior is consistent with the optimization picture developed earlier and helps distinguish genuine low-SNR robustness from incidental variance.

Robustness checks under mild distribution shift provide further evidence. Small changes in amplitude normalization, phase rotation statistics, or timing jitter can be introduced at test time. A network whose improvement relies mainly on overloaded capacity may degrade sharply because its extra channels encode fragile correlations. A network whose improvement derives from selective regulation should degrade more gracefully, since its effective decision coordinates are already biased toward sample-consistent channels. Even limited robustness experiments can therefore strengthen the interpretive case.

Finally, empirical claims should be matched to operationally meaningful scenarios. A receiver concerned with spectrum awareness may value accurate rejection of a few specific confusions more than a modest increase in average accuracy [40]. A network that sharply reduces confusion among low-SNR analog and constant-envelope digital classes may be more useful than one that slightly improves already easy quadrature classes. Therefore, the evaluation framework should permit task-specific weighting of errors and not assume that all misclassifications have equal practical cost.

The broader interpretive lesson is that experiments should reveal whether the architecture changes the internal distribution of representational effort. If low-SNR accuracy improves, targeted confusions decline, feature compactness increases, posterior calibration stabilizes, and computational overhead remains controlled, then the evidence supports the theory of channel-selective information regulation. If gains appear only as diffuse global improvements with no low-SNR concentration, then the architecture may simply be larger. The distinction matters because it determines whether the model is learning to manage information or merely to absorb more of it.

7 | Architectural Implications for Robust Modulation Recognition

The theoretical and empirical arguments developed so far suggest a set of architectural implications that extend beyond any single network instance. The first implication is that useful depth in modulation recognition is conditional depth. Additional layers are beneficial when they transport, reshape, and selectively amplify information; they are less beneficial when they merely accumulate channels and nonlinearities. This means that architecture design should prioritize mechanisms that preserve fragile shallow evidence, expose it to broader context, and then regulate its influence. Depth without governance risks converting faint structure into feature congestion.

The second implication is that channel-wise control should be regarded as a core component of the feature extractor rather than an optional refinement appended near the output [41]. In many signal tasks, the representation quality depends on how energy is allocated across channels at intermediate stages. Once nuisance-heavy channels dominate those stages, later layers may not be able to recover the lost signal-to-nuisance ratio. A compact squeeze-and-gate module is attractive precisely because it intervenes before the overload compounds. It can be inserted at multiple depths, especially after grouped or residual transformations where feature diversity is high and some mechanism is needed to adjudicate among channels.

A third implication concerns fusion. Shallow and deep features should not be fused indiscriminately. The value of fusion lies in complementarity, not in mere accumulation. Shallow features carry local temporal detail and phase-amplitude microstructure; deeper features carry context and emergent semantic regularities. Fusion is helpful when these two views address different failure modes. It becomes harmful when it reintroduces redundant or noisy channels without regulation. Therefore, fusion should typically be paired with some channel-selective mechanism that can decide whether the preserved shallow information is currently useful. Architecturally, this means that skip transport and gating should be designed together [42].

The fourth implication is that grouped transformations and branch diversity are most effective when paired with a selection mechanism. Parallel branches allow the network to represent multiple interpretations of the same local signal pattern. One branch may respond to smooth envelope structure, another to abrupt transitions, another to oscillatory texture. This diversity is valuable in low-SNR settings where no single descriptor is consistently reliable. However, branch diversity also increases the risk of redundant or nuisance-specialized channels. Channel regulation makes branch diversity economically sustainable by ensuring that not all branches are expressed equally on every sample.

The fifth implication concerns the granularity of invariance. In modulation recognition, one should not seek blanket invariance to all physical transformations. Some phase, frequency, and amplitude behavior is class-informative. A better objective is conditional invariance: the representation should be insensitive to nuisance variation only insofar as that variation is not needed to preserve class distinctions. Channel-selective mechanisms naturally support conditional invariance because they act on the current sample context. They do not force the same invariance pattern for all classes [43]. Instead, they let the model preserve channels needed for one class while suppressing analogous channels for another if the evidence warrants it.

A sixth implication is computational. Selective models can achieve a favorable balance between global capacity and local efficiency. A network can be architecturally rich, with many branches and channels, while remaining sample-efficient because only a subset of that richness is strongly expressed at a time. This is valuable in wireless applications where different modulation families and environmental conditions may call for different latent features. It also means that model comparison should not focus solely on static parameter count. Two models with similar parameter counts may differ substantially in their effective complexity per sample depending on how selectively they activate channels.

The seventh implication is methodological. Researchers should analyze hidden representations explicitly rather than relying only on final accuracy. If the central problem is information overload, then feature covariance, effective rank, gate entropy, and class-conditional compactness become natural diagnostic tools. Such diagnostics can reveal whether an architecture is behaving as intended. A model might show only modest test-set gains yet still display much cleaner latent organization, suggesting that further training or data augmentation could unlock larger improvements [44]. Conversely, a model might show a temporary accuracy gain while its latent features remain highly congested, indicating fragile generalization.

The eighth implication is that low-SNR robustness should be treated as a distributional property, not as a single point estimate. A model that performs well at one or two difficult SNR levels but erratically elsewhere may still be unstable. Selective information regulation should ideally produce smooth improvements across a low-SNR band because it addresses a structural issue in representation quality. Architecture search and hyperparameter tuning should therefore target the shape of the SNR-performance curve and the stability of class-specific confusions, not merely the best score at one threshold.

A ninth implication concerns future integration with physics-informed front ends. Channel-selective gating and residual transport need not be viewed as alternatives to domain-aware preprocessing. They can complement transforms such as time-frequency maps, cyclostationary cues, or synchronization-aware preprocessing. In such hybrid systems, selective gating could operate over channels that correspond to different physical descriptors, effectively learning when each descriptor family is useful. The central principle remains the same: the network should regulate information according to task relevance under the current sample context.

A final implication is conceptual. The design of robust modulation classifiers should shift from the language of larger and deeper networks toward the language of controlled information flow [45]. The decisive question is not how many channels the network can create, but how well it can maintain a high ratio of discriminative to nuisance content as the observation becomes weaker. Channel-selective compression, multiscale transport, and structured residual perturbation together form one coherent answer to that question. They provide a path toward models that do not merely survive the low-SNR regime, but organize their internal evidence in a way that is appropriate for it.

8 | Conclusion

Low-SNR automatic modulation recognition is most productively understood as a problem of representational regulation under uncertainty. The received waveform contains class structure, nuisance distortion, and finite-window ambiguity in an inseparable mixture, and deep classifiers succeed only when they transform that mixture into a latent space where class-relevant directions are preserved and nuisance-dominated directions are attenuated. The analysis in this paper has argued that information overload is a central obstacle in that process. Hidden tensors can become rich in channels yet poor in usable evidence, particularly when weak signal traces are carried forward together with unstable or redundant activations. Channel-selective compression provides a principled response by using global context to reweight hidden channels, reducing nuisance covariance and improving the reliability of class margins without requiring indiscriminate growth in model size.
The broader framework developed here also shows that selective channel weighting is most effective when embedded within a larger architecture of controlled information transport. Residual pathways preserve baseline signal flow, multiscale fusion retains fragile shallow evidence, and gating governs which channels should influence later decisions for each sample. From the viewpoints of feature geometry, optimization dynamics, and experimental interpretation, these mechanisms all support the same objective: to improve the ratio between stable class information and overloaded feature clutter in the regime where the observation is weakest. This perspective provides a foundation for future work on robust modulation recognition and suggests that the most valuable architectural advances will come not from scale alone, but from more deliberate control of how information is preserved, compressed, and used [46].

References

[1] H. Hasegawa, H. Kanai, N. Hoshimiya, and Y. Koiwa, "Reduction of influence of decrease in signal-to-noise ratio in measurement of change in thickness of arterial wall due to heartbeat," *Japanese Journal of Applied Physics*, vol. 39, pp. 3257–3261, 5 2000.

[2] D. Joshi, N. Sharma, and J. B. Singh, "Spectrum sensing for cognitive radio using hybrid matched filter single cycle cyclostationary feature detector," *International Journal of Information Engineering and Electronic Business*, vol. 7, pp. 13–19, 9 2015.

[3] Y. Zhang, Q. Zhang, and S. Wu, "Entropy-based robust spectrum sensing in cognitive radio," *IET Communications*, vol. 4, pp. 428–436, 3 2010.

[4] B. Zou, H. Yan, F. Wang, Y. Zhou, and X. Zeng, "Research on signal modulation classification under low snr based on resnext network," *Electronics*, vol. 11, no. 17, p. 2662, 2022.

[5] D.-H. Seo, J.-W. Choi, and Y.-H. Kim, "Impulsive sound localization using time-domain beamformer," *Journal of the Acoustical Society of America*, vol. 137, pp. 2232–2232, 4 2015.

[6] M. J. Burfeindt, J. D. Shea, B. D. V. Veen, and S. C. Hagness, "Beamforming-enhanced inverse scattering for microwave breast imaging.," *IEEE transactions on antennas and propagation*, vol. 62, pp. 5126–5132, 7 2014.

[7] M. J. Teschner, B. A. Seybold, B. J. Malone, J. Hüning, and C. E. Schreiner, "Effects of signal-to-noise ratio on auditory cortical frequency processing.," *The Journal of neuroscience : the official journal of the Society for Neuroscience*, vol. 36, pp. 2743–2756, 3 2016.

[8] S. Mohammadi, K. Tabelow, L. Ruthotto, T. Feiweier, J. Polzehl, and N. Weiskopf, "High-resolution diffusion kurtosis imaging at 3t enabled by advanced post-processing.," *Frontiers in neuroscience*, vol. 8, pp. 427–427, 1 2015.

[9] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, "Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery," *IEEE Transactions on Network Science and Engineering*, vol. 1, pp. 10–22, 1 2014.

[10] A. Ozen, “A novel variable step size adjustment method based on channel output autocorrelation for the lms training algorithm,” *International Journal of Communication Systems*, vol. 24, pp. 938–949, 1 2011.

[11] S. S. Chauhan and S. Kumar, “Channel capacity and outage probability analysis of multiuser diversity mimo mrc systems with antenna selection in the presence of delayed feedback,” *Telecommunication Systems*, vol. 64, pp. 559–567, 6 2016.

[12] . , . , and . , “Translational motion compensation for ballistic targets based on delayed conjugated multiplication,” *Journal of Radars*, vol. 3, pp. 505–510, 10 2014.

[13] J. Dong, F. Chen, D. Zhou, T. Liu, Z. Yu, and Y. Wang, “Phase unwrapping with graph cuts optimization and dual decomposition acceleration for 3d high-resolution mri data.,” *Magnetic resonance in medicine*, vol. 77, pp. 1353–1358, 3 2016.

[14] S.-H. Shin, T. Hashimoto, and S. Hatano, “Automatic detection system for cough sounds as a symptom of abnormal health condition,” *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 13, pp. 486–493, 4 2008.

[15] E. Vassallo and M. Visintin, “Carrier phase synchronization for gmsk signals,” *International Journal of Satellite Communications*, vol. 20, pp. 391–415, 11 2002.

[16] B. Ahuja and G. Kaur, “Design of an improved spectrum sensing technique using dynamic double thresholds for cognitive radio networks,” *Wireless Personal Communications*, vol. 97, pp. 821–844, 6 2017.

[17] D. Anderson, “Nanoseismic monitoring fills the gap between microseismic networks and passive seismic,” *First Break*, vol. 26, 6 2008.

[18] K. Balakrishnan, G. Anwar, M. G. Chapman, T. Nguyen, A. Kesavaraju, and L. L. Sohn, “Node-pore sensing: a robust, high-dynamic range method for detecting biological species,” *Lab on a chip*, vol. 13, pp. 1302–1307, 4 2013.

[19] K. G. Schilling, V. Nath, J. A. Blaber, P. Parvathaneni, A. W. Anderson, and B. A. Landman, “Empirical consideration of the effects of acquisition parameters and analysis model on clinically feasible q-ball imaging.,” *Magnetic resonance imaging*, vol. 40, pp. 62–74, 4 2017.

[20] S. Poudel, H. Wijanto, and F. Y. Suratman, “Spectrum sensing in cognitive radio using combined sequential energy detector and cyclostationary feature detector,” *Journal of Measurements, Electronics, Communications, and Systems*, vol. 2, pp. 31–, 12 2016.

[21] J. Ghosh, S. D. Beck, and C.-C. Chu, “Evidence combination techniques for robust classification of short-duration oceanic signals,” *SPIE Proceedings*, vol. 1706, pp. 266–276, 8 1992.

[22] Z. Wen, S. B. Reeder, A. R. Pineda, and N. J. Pelc, “Noise considerations of three-point water-fat separation imaging methods,” *Medical physics*, vol. 35, pp. 3597–3606, 7 2008.

[23] W. Tang, M. Z. Shakir, M. Imran, R. Tafazolli, K. A. Qaraqe, and J. Wang, “Spectral and energy efficient cognitive radio-aided heterogeneous cellular network with uplink power adaptation,” *Wireless Communications and Mobile Computing*, vol. 16, pp. 2144–2162, 5 2016.

[24] M.-A. Chao, X.-Y. Shih, and A.-Y. Wu, “Matrix merging scheme and efficient decoding techniques for reconfigurable qc-ldpc decoders,” *Journal of Signal Processing Systems*, vol. 68, pp. 183–202, 6 2011.

[25] E. Verteletskaya and K. Sakhnov, “Voice activity detection for speech enhancement applications,” *Acta Polytechnica*, vol. 50, 1 2010.

[26] A. M. Jorgensen, D. Mozurkewich, J. T. Armstrong, H. R. Schmitt, T. A. Pauls, and R. B. Hindsley, “Improved coherent integration through fringe model fitting,” *The Astronomical Journal*, vol. 134, pp. 1544–1550, 8 2007.

[27] X. Jiang, Y. Lai, and M. Xia, “Noise robust target identification based on the wave-coefficients-2dimension case,” *Inverse Problems*, vol. 32, pp. 125006–, 11 2016.

[28] X. Wang, J.-K. Wang, Z.-G. Liu, and Q. Zhao, “Signal recognition algorithm based on random forests for spectrum sensing in cognitive network,” *Journal of Information and Computational Science*, vol. 11, pp. 2551–2558, 5 2014.

[29] Y. Honjo, H. Hasegawa, and H. Kanai, “Optimization of correlation kernel size for

accurate estimation of myocardial contraction and relaxation," *Japanese Journal of Applied Physics*, vol. 51, pp. 7GF06–07GF06, 7 2012.

[30] X.-Z. Ren and L.-N. Chen, "Four-dimensional sar imaging algorithm using bayesian compressive sensing," *Journal of Electromagnetic Waves and Applications*, vol. 28, pp. 1661–1676, 7 2014.

[31] Z. Zhang, S. W. Cheung, and T. I. Yuk, "Upper bound for block-error rate of s-t codes," *Electronics Letters*, vol. 43, pp. 116–117, 1 2007.

[32] D. Nandi, D. Pati, and K. S. Rao, "Implicit excitation source features for robust language identification," *International Journal of Speech Technology*, vol. 18, pp. 459–477, 6 2015.

[33] J.-Y. Choi, S. Park, and C.-S. Park, "A gps positioning and receiver autonomous integrity monitoring algorithm considering sa fade away," *Journal of Control, Automation and Systems Engineering*, vol. 8, pp. 425–433, 5 2002.

[34] S. M. Anzar, "An efficient pso optimized integration weight estimation using d-prime statistics for a multibiometric system," *International Journal on Bioinformatics & Biosciences*, vol. 2, pp. 31–42, 9 2012.

[35] C. Birnie, K. Chambers, D. Angus, and A. L. Stork, "Effect of noise on microseismic event detection and imaging using icova statistical noise modelling method," *SEG Technical Program Expanded Abstracts 2016*, pp. 2622–2626, 9 2016.

[36] E. T. Petersen, K. Mouridsen, and X. Golay, "The quasar reproducibility study, part ii: Results from a multi-center arterial spin labeling test-retest study.," *NeuroImage*, vol. 49, pp. 104–113, 8 2009.

[37] N. Yanasak, J. D. Allison, and T. hu, "An empirical characterization of the quality of dti data and the efficacy of dyadic sorting," *Magnetic resonance imaging*, vol. 26, pp. 122–132, 7 2007.

[38] R. Giryes and M. Elad, "Sparsity-based poisson denoising with dictionary learning," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 23, pp. 5057–5069, 10 2014.

[39] Z. Ahmed, I. E. Naqa, and I. R. Levesque, "Th-cd-207-03: A constrained linear reference region model for dce-mri," *Medical Physics*, vol. 42, pp. 3735–3735, 6 2015.

[40] R. R. Cordero, G. Seckmeyer, S. Riechelmann, A. Damiani, and F. Labbe, "Monte carlo-based uncertainty analysis of uv array spectroradiometers," *Metrologia*, vol. 49, pp. 745–755, 10 2012.

[41] L. S. Mok, C. K. Ng, and N. K. Noordin, "Linear multi-hop amplify-and-forward (af) cooperative relay channels," *Proceedings of the Asia-Pacific Advanced Network*, vol. 40, pp. 64–69, 10 2015.

[42] H. A. Khan, T. Shin, M. H. Lee, and J.-G. Chung, "Signal constellations of quasi-orthogonal space—time codes for mimo systems," *Wireless Personal Communications*, vol. 85, pp. 2003–2019, 7 2015.

[43] D. Nath, M. V. Ratnam, V. V. M. J. Rao, B. V. K. Murthy, and S. V. B. Rao, "Gravity wave characteristics observed over a tropical station using high-resolution gps radiosonde soundings," *Journal of Geophysical Research: Atmospheres*, vol. 114, 3 2009.

[44] K. Sidaros, K. Olofsson, M. J. Miranda, and O. B. Paulson, "Arterial spin labeling in the presence of severe motion," *Journal of Cerebral Blood Flow & Metabolism*, vol. 25, pp. S382–S382, 8 2005.

[45] J. Rajan, K. Kannan, and M. R. Kaimal, "An improved hybrid model for molecular image denoising," *Journal of Mathematical Imaging and Vision*, vol. 31, pp. 73–79, 1 2008.

[46] N. Kunieda, T. Shimamura, and J. Suzuki, "Characteristics of pitch extraction by aclos (autocorrelation of log spectrum)," *The Journal of the Acoustical Society of America*, vol. 100, pp. 2602–2603, 10 1996.