

split attention probability for all attention heads for post-norm

